Supplementary Information for

The capsid lattice engages a bipartite NUP153 motif to mediate nuclear entry of HIV-1 cores

Qi Shen,[a,b,c] Sushila Kumari,[d] Chaoyi Xu,[e] Sooin Jang,[f,g] Jiong Shi,[h] Ryan C. Burdick,[d] Lev Levintov,[e] Qiancheng Xiong,[b,c] Chunxiang Wu,[a] Swapnil C. Devarkar,[a] Taoran Tian,[c] Therese N. Tripler,[a] Yingxia Hu,[a] Shuai Yuan,[a] Joshua Temple,[a] Qingzhou Feng,[b,c] C. Patrick Lusk,[b] Christopher Aiken,[h] Alan N. Engelman,[f,g] Juan R. Perilla,[e] Vinay K. Pathak,[d] Chenxiang Lin,[b,c,j] and Yong Xiong [a,1]

[a]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, 06511; [b]Department of Cell Biology, Yale School of Medicine, New Haven, CT, 06520; [c]Nanobiology Institute, Yale University, West Haven, CT, 06516; [d]Center for Cancer Research, National Cancer Institute at Frederick, Frederick, MD, 21702; [e]Department of Chemistry and Biochemistry, University of Delaware, Newark, DE, 19716; [f]Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, MA, 02215; [g]Department of Medicine, Harvard Medical School, Boston, MA, 02115; [h]Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN, 37232; [j]Department of Biomedical Engineering, Yale University, New Haven, CT, 06511.

Corresponding author: Yong Xiong
Email: yong.xiong@yale.edu

This PDF file includes:

Supplemental Materials and Methods

Cloning and expression

$NUP153_{CTD}$ from *homo sapiens* (amino acids 896-1475), as well as yeast nucleoporin NSP1 (amino acids 2-603) from *S. cerevisiae*, was cloned as 10× His-MBP-SUMO-NUP153 (or NSP1)-SNAP constructs (Fig. S1) into a pET-28a-derived vector (Novagen) and expressed in BL21-Gold (DE3). Cells were grown in terrific broth (TB) at 37 °C with shaking until they reached an OD_{600} of 1.0. Protein expression was then induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 3 h at 20 °C before collection by centrifugation. Cell pellets were frozen at −80 °C until use.

All CA constructs were cloned into pET-11a (EMD Millipore). CA-Foldon was generated by directly fusing Foldon to the C-terminus of CA. CA proteins were overexpressed in BL21-Gold (DE3) cells at 25 °C for 12 h by induction with 0.5 mM IPTG at 0.8 OD600. Cell pellets were collected by centrifugation and frozen at −80 °C until use.

Protein purification

NUP153/NSP1-expressing cell pellets were resuspended in lysis buffer (1× PBS, 0.5 mM TCEP, 0.1 mM PMSF, 1× Roche cOmplete protease inhibitors, pH 7.4), and lysed in a homogenizer. Whole-cell lysates were then spun at 35,000 rpm for 45 min in a Type 45 Ti rotor (Beckman Coulter). Subsequently, the supernatant was decanted and filtered through a 0.45 µm cellulose acetate membrane. The resulting filtered lysate was spiked with 25 mM imidazole, then applied to a 5 ml HisTrap column (GE) on an ÄKTA system (GE) at a flow rate of 1 ml/min. The column was rinsed with a wash buffer (1× PBS, 25 mM imidazole) and eluted on a buffer gradient (1×PBS, 500 mM imidazole). Concentrations of purified protein were determined by Nanodrop (Thermo Fisher). Samples were flash-frozen in liquid nitrogen and stored at −80 °C until use.

Untagged CA proteins were purified by 25% w/v ammonium sulfate precipitation, dialysis into a low-salt buffer (25 mM HEPES, 0.1 mM TCEP, pH 7), and cation exchange chromatography. CA-foldon fusions were purified by 35% w/v ammonium sulfate precipitation and anion exchange chromatography. All CA constructs were dialyzed into CA storage buffer before freezing or further experiments (50 mM Tris-HCl, 75 mM NaCl, 40 mM b-mercaptoethanol (BME), pH 8). All purification steps were validated using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

CA assembly

<u>Tri-hexamer$_{center}$, hexamer and pentamer</u>

Tri-hexamer$_{center}$ was assembled using a 1:1 molar ratio of the appropriate CA proteins (1), resulting in 10-40 mg/ml of total proteins. Mixtures were dialyzed overnight (using Thermo Slide-a-lyzer dialysis cassettes) in a 50 mM Tris-HCl, 1 M NaCl, pH 8 buffer. Mixtures were dialyzed for a second

night in a 50 mM Tris-HCl, pH 8 buffer. Tri-hexamer$_{center}$ assemblies eluted from an anion exchange column at approximately 250 mM NaCl. Each assembly was purified using a Superdex 200 PG or Superdex 200 GL size-exclusion chromatography column (GE Healthcare) in 50 mM Tris-HCl, 300 mM NaCl, pH 8 buffer.

Hexamer (A14C/E45C/W184A/M185A) and pentamer (N21C/A22C/W184A/M185A) assemblies were dialyzed (using Thermo Slide-a-lyzer dialysis cassettes) for 48 h in a 50 mM Tris-HCl, pH 8 buffer without redox reagents. Subsequently, the crosslinked hexamers or pentamers were purified using a Superdex 200 GL column. All assemblies were concentrated to 20-50 mg/ml and frozen at −80 °C for storage.

Crosslinked CA tubes

A14C/E45C disulfide crosslinked CA tubes were assembled by dialyzing (in Slide-A-Lyzer dialysis cassettes) CA A14C/E45C into a 50 mM Tris-HCl, 1 M NaCl, pH 8 buffer at 15 mg/ml for one night, followed by dialysis into 50 mM Tris-HCl, pH 8 for another night. The crosslinked tubes were stored in a 50 mM Tris-HCl, 50 mM NaCl, pH 8 buffer and could remain stable at 4 °C for months.

Wild-type CA tubes

Wild-type CA tubes (100 μM CA) were assembled by dialyzing into a 50 mM Tris-HCl, 1 M NaCl, pH 8 buffer at 5 mg/ml for three to four nights. Assemblies were stored in a 50 mM Tris-HCl, 1 M NaCl, pH 8 buffer.

Inositol hexakisphosphate (IP6) stabilized wild-type CA tube

IP6 was added at 6.67 mM to pre-assembled WT CA tubes (80-150 μM CA) in 50 mM Tris-HCl, 1 M NaCl, pH 8 buffer for 30 min. The mixture was then diluted 6.67-fold in 50 mM Tris-HCl, pH 8 buffer to obtain IP6-stabilized WT CA tubes in a 50 mM Tris-HCl, 150 mM NaCl, 1 mM IP6, pH 8 buffer. The concentration of the IP6-stabilized CA tubes was adjusted by pelleting and resuspension in the desired volume of 1× PBS, 1 mM IP6 buffer before use.

CA tube co-pelleting assays

A14C/E45C disulfide crosslinked CA tubes were dialyzed into 25 mM Tris-HCl, pH 8 buffer. NUPs were incubated in 21 μl reactions with CA tubes for 30 min at room temperature, then spun at 20,000 × g for 10 min at 4 °C. Total, soluble, and pellet fractions were collected and analyzed using SDS-PAGE. NUP153$_{CTD}$ and all mutant constructs were incubated at 3 μM with 100 μM CA in a 25 mM Tris-HCl, 75/150/300 mM NaCl, pH 8 buffer. NUP153 variants (3 μM) co-pelleted with the IP6 stabilized WT CA tube (100 μM CA) in 1× PBS buffer containing 1 mM IP6.

For co-pelleting assays with cell lysates, human embryonic kidney (HEK293T) cells (ATCC CRL-3216; 3.5 × 10^6 cells/10-cm dish) were seeded and transfected the next day with WT or mutant mRuby-NUP153 expression plasmids (10 μg/10-cm dish). Cells were harvested 24 h post-

transfection with TrypLE Select (Gibco; Cat#12563), washed once with 1× PBS, and lysed with 1 mL lysis buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% TX-100) containing cOmplete protease inhibitor cocktail (Roche; Cat.#05892491001) on ice for 10 min. The supernatant was collected after centrifugation at 16,000 x g for 10 min at 4°C. Next, we incubated A14C/E45C disulfide crosslinked CA tubes (3 µl of 200 µM stock; 4.9 µM final concentration) with cell extract (120 µl) at room temperature for 20 min and then spun at 20,000 × g for 10 min at 4 °C. Total, soluble, and pellet fractions were collected and analyzed using SDS-PAGE and western blot using anti-NUP153 antibody (Abcam; Cat#ab848872; 1:1,000 dilution) followed by goat anti-rabbit secondary antibody (IRDye-800CW, 1:10,000 dilution) and anti-HIV-1 p24 Gag Monoclonal (#24-3) (NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH, from Dr. Michael Malim; ARP6458; 1:10,000 dilution), which was followed by goat anti-mouse secondary antibody (IRDye-680RD, 1:10,000 dilution). Western blots were imaged and quantified using the Odyssey infrared imaging system (LiCor).

Size-exclusion chromatography co-elution assays

Assembled CA hexamer, tri-hexamer$_{center}$, and pentamer were mixed with NUP153$^{1411-1475}$ for 30 min to 1 h on ice in the same buffer used for CA tube co-pelleting assays (unless specified). Binding reactions were carried out in a volume of 100 µl and contained 12 µM NUPs and 72 µM monomeric concentration of CA. All binding tests were performed on GE Superdex 200 10/300 GL column in a 25 mM Tris-HCl, 75 mM NaCl, pH 8 buffer with a flow rate of 0.5 ml/min. The 280 nm absorbance was recorded.

Size-exclusion chromatography for assembled N21C/A22C CA spheres, NUP153$^{1411-1475}$, and mixtures were performed on GE Superose 6 10/300 GL column in 25 mM Tris-HCl buffer pH 8, 50/1000 mM NaCl. The 280 nm absorbance was recorded.

Isothermal titration calorimetry

Binding reactions were performed in a TA Instruments NanoITC machine at 25 °C using 25 mM Tris-HCl, 75 mM NaCl, pH 8 buffer. Tri-hexamer$_{center}$ was stable in these conditions after overnight dialysis and during experiments. Synthesized NUP153 peptide (Genscript Co.) was resuspended in the same buffer and injected as the titrant to the tri-hexamer$_{center}$ in the cell. Data were analyzed using the NanoAnalyze (TA Instruments) software, and curves were fitted with an independent one-site binding model.

Size exclusion chromatography linked to multi-angle light scattering (SEC-MALS)

Multiangle laser light-scattering experiments were performed at room temperature in a 50 mM Tris-HCl, 150 mM NaCl, pH 8 buffer. Light-scattering data were measured using a Dawn Heleos-II spectrometer (Wyatt Technology) coupled to an Opti-lab T-rEX (Wyatt Technologies) interferometric refractometer. Samples (100 µl) at 1 mg/ml were injected and run over a Superdex 200 GL (GE Healthcare) column at a 0.5 ml/min flow rate. Data on light scattering (690 nm laser),

280 nm absorbance, and refractive index were measured. Before sample runs, the system was calibrated and normalized using the isotropic protein standard, monomeric bovine serum albumin. The dn/dc value (changes in solution refractive index with respect to protein concentration) was relatively constant for proteins (2), and set to 0.185 for all experiments and analysis. Data were processed in ASTRA as previously described (3).

Molecular dynamics (MD) simulation

Atomic model for MD simulation

The 3-dimension structure of the C-terminal 75 residues of NUP153$^{1401-1475}$ was built in Modeller (4). A total of 5,000 structures were generated and scored by the discreet optimized protein energy (DOPE) method (5). The candidate model of NUP153$^{1451-1475}$, which was from the structure of NUP153$^{1401-1475}$ with the lowest DOPE score, was subjected to a 100-ns long equilibration in explicit water in NAMD2.13 (6). The equilibrated NUP153$^{1451-1475}$ peptide was initially placed 1.5 nm above the tri-hexamer interface in the HIV-1 CA tri-hexamer$_{center}$ model, which was built based on the structure of the full-length native HIV-1 CA monomer (PDB:4XFX). The combined structure (Fig. S6A bottom) was then solvated with TIP3P water (7) and neutralized by NaCl at 150 mM concentration. The resulting model had a total number of 168,500 atoms (Fig. S6A).

MD simulation

Before production run, the NUP153$^{1451-1475}$-CA tri-hexamer$_{center}$ complex model was subjected to a minimization and thermalization step. Subsequently, the whole system was equilibrated for over 100 ns at 310 K. During the equilibration, the Cα atoms of the peripheral helices of CA monomer and the N-terminal residue in NUP153$^{1451-1475}$ were restrained. The minimization, thermalization and equilibration steps were completed in NAMD2.13 (6). This equilibrated model was then run for a total simulation time of 8.7 μs on a special purpose computer Anton2 (8) in the Pittsburgh supercomputing center. To mimic the CA hexamer lattice environment in the simulation, the Cα atoms of three sets of CA α-helices on the peripheries of the tri-hexamer model were applied with a harmonic restraint of 1 Kcal/mol Å^2 in x, y and z directions (Fig. S6A bottom). Also, the Cα atoms of NUP153 residue 1451 was applied with a harmonic restraint of 1 Kcal/mol Å^2 in z direction. CHARMM 36m (9) force field was used for all MD simulations. During the simulation, the temperature (310 K) and pressure (1 atm) was maintained by employing the Multigrator integrator (10) and the simulation time step was set to 2.5 fs/step, with short-range forces evaluated at every time step, and long-range electrostatics evaluated at every second time step. Short-range non-bonded interactions were cut off at 17 Å; long range electrostatics were calculated using the k-Gaussian Split Ewald method (11).

NUP153$^{1451-1475}$ was found to be mostly unstructured at the CA tri-hexamer interface (Fig. S6B). The secondary structure calculation of the NUP153$^{1451-1475}$ from the simulation indicated that it does

not have an ordered structure when present in the tri-hexamer region. Because of its disordered structure, the time-lagged independent component analysis (TICA) (12) was employed to better understand the conformation of the unstructured $NUP153_{CTD}$, using internal degrees of freedom and Cartesian coordinates of the $NUP153^{1451-1475}$ obtained in the Anton2 simulation. $NUP153^{1451-1475}$ conformation distribution was mapped onto the first two IC space (Fig. S7A). The distributed conformation suggested that the $NUP153^{1451-1475}$ peptide was highly dynamic and there were several metastable states, indicated by the high probability areas (Fig. S7A).

Markov state model analysis and Model validation

Markov state models (MSMs) have been widely applied to study the conformational dynamics of biomolecules (13-15). To identify the important metastable states of $NUP153^{1451-1475}$ in the MD simulation, Markov state models were constructed and validated in the PyEMMA 2.5.7 package (16). First, the coordinates of $NUP153^{1451-1475}$ from the Anton2 simulation were directly transformed into two features, the Cartesian coordinates of $NUP153^{1451-1475}$ backbone atoms and 4,853 distance pairs between the backbone atoms of $NUP153^{1451-1475}$. Subsequently, time-lagged independent component analysis (TICA) (12) was performed to decompose these features onto 100 slow independent components (ICs). The projected data set was then discretized using the k-means method (17) and resulted in fifty microstates. The transition probability matrix between these microstates was then computed at a selected lag time of 1.2 ns (Fig. S7B). Finally, an eight-state MSM using the PCCA+ algorithm (18) was constructed (Fig. S7C-E) and the mean first passage times (MFPTs) between different states were estimated (Fig. S7F). The eight-state MSMs built were validated with the Chapman-Kolmogorov test (19, 20).

TRIM-NUP153 mediated restriction assays

Plasmid construction

TRIM-NUP153 proteins were expressed from the bicistronic expression vector pIRES2-eGFP (21). To make TRIM-HA-$NUP153^{896-1475}$, TRIM and NUP153 sequences were PCR-amplified from pLPCX-Trim-HA-$NUP153_{CTD}$ (22) using primer pairs 5'-GATCTCGAGCTCAAGCTTCGAATTC/5'-GTAATCTGGAACATCGTATGGGTAGCCACCTCCAGATCCCCAGTAGCGTCGG3' and 5'-CATACGATGTTCCAGATTACGCTGGAGGTGGATCTCCGCGGAACTCAGCAGCCTCCTC/5'-GATCCCGGGCCTTTATTTCCTGCGTCTAACAG, respectively (SacII restriction site underlined). The linked TRIM-HA-NUP153 insert, amplified from these DNAs using primers 5'-GATCTCGAGCTCAAGCTTCGAATTC and 5'-GATCCCGGGCCTTTATTTCCTGCGTCTAACAG, was digested with XhoI and XmaI and ligated with XhoI/XmaI-digested pIRES2-eGFP. Trim-HA-$NUP153^{1401-1475}$ vector was created by amplifying $NUP153^{1401-1475}$ sequences using primers 5'-CAGCCGCGGACTACAAATTTCAACTTCACAAACAACAG and 5'-GATCCCGGGCCTTTATTTCCTGCGTCTAACAG, followed by digestion with SacII and XmaI and swapping this DNA for the analogous region within Trim-HA-$NUP153^{896-1475}$ via SacII and XmaI

restriction sites. To create Trim-HA-NUP153$^{1411-1475}$ lacking residues 1425-1464, double-stranded DNA representing NUP153 residues 1411-1424 and 1465-1475 was synthesized by IDT (5'CAGCCGCGGCCATCAGGAGTGTTCACATTTGGTGCAAATTCTAGCACACCTCGCAAGATA AAGACTGCTGTTAGACGCAGGAAATAAAGGCCCGGGATC3'), digested with SacII and XmaI, and swapped for the NUP153$^{896-1475}$ region as above. DNA fragments encoding Trim-HA-NUP153$^{1401-1475}$ with FTFG>4A, RRR>3A and FTFG>4A+RRR>3A mutations were synthesized by Integrated DNA Technologies and inserted into EcoRI/BamHI-digested pIRES2-eGFP using Gibson assembly as recommended by the manufacturer (NEB). TRIM-NUP coding regions of all plasmids were verified by DNA sequencing.

Cells, virus production, and infection

HEK293T cells were cultured in Dulbecco's modified Eagle medium supplemented to contain 10% fetal bovine serum (FBS), 100 IU/ml penicillin, and 0.1 mg/ml streptomycin. Single-round HIV-1 carrying the gene for firefly luciferase (HIV-Luc), produced by transfecting HEK293T cells in 10 cm dishes with 7.5 µg pNLX.Luc.R-.ΔAvrII and 1.5 µg pHCMV-G, which expresses the vesiculostomatitis virus G glycoprotein (23) using PolyJet (SignaGen Laboratories), was concentrated by ultracentrifugation as described (24). HIV-Luc yield was assessed by p24 ELISA (Advanced Bioscience Laboratories). Cells transfected with pIRES2-eGFP expression constructs were used in HIV infection assays as previously described (25). Briefly, HEK293T cells in 6-well plates were transfected with 2 µg pIRES2-eGFP or Trim-HA-NUP153 derivatives using Effectene (Qiagen). At 24 h post-transfection, GFP-positive cells were selected in basic sorting buffer (1 mM EDTA, 25 mM HEPES, pH 8.0, 1% FBS, Mg^{2+}/Ca^{2+}-free phosphate-buffered saline) by fluorescence-activated cell sorting. Approximately 3-5 x 10^5 sorted cells were lysed for immunoblotting while ~2 x 10^5 cells were infected in duplicate with HIV-Luc (0.25 pg p24 per cell) in the presence of 4 µg/ml polybrene (Sigma). At 48 h post-infection, cells were lysed and luciferase activity was determined as previously described (26). Luciferase values were normalized to the level of total protein in cell lysates as determined by the BCA assay (Pierce).

Western Blotting

Details of the procedure were as described (25). Total cellular proteins (10 µg) were separated through 3-8% polyacrylamide Tris-acetate or 4-12% polyacrylamide Bis-Tris gels (Invitrogen) and were subsequently transferred to polyvinylidene difluoride (PVDF) membranes. Fusion protein expression was detected with anti-hemagglutinin (HA) antibody sc7392 (Santa Cruz, 1:500 dilution) followed by horseradish peroxidase (HRP)-mouse IgG antibody (Dako, 1:10,000 dilution); actin was detected by HRP-conjugated β-actin antibody (Abcam, 1:10,000 dilution). Membranes developed using ECL prime reagent (Amersham Biosciences) were imaged with a ChemiDoc MP imager (Bio-Rad).

Assays of HIV-1 infection in NUP153-depleted cells

HeLa cells were seeded at 100,000 cells per well in 12-well dishes and cultured in complete medium for 24 h. Monolayers were washed once with 1× PBS and transfected with Dharmacon siRNA duplexes (20 nM concentrations) using Lipofectamine RNAimax according to the manufacturers' instructions. Two days later, cells were detached with trypsin and replated in 96-well dishes (10,000 cells per well). The next day, cultures were inoculated with dilutions of VSV-G-pseudotyped HIV-GFP reporter virus stocks produced by transfection of 293T cells in media containing 20 μg/ml DEAE-dextran. Two days later, cells were detached with trypsin, fixed with paraformaldehyde solution, and analyzed for GFP expression by flow cytometry after at least 24 h of fixation. Relative infectivity was calculated by normalization for CA antigen in the virus inocula, determined by p24 ELISA.

Immunoblot analysis of NUP153 depletion

Whole cell lysates of HeLa cells transfected with control or NUP153-targeting siRNAs were prepared with RIPA buffer three days after transfection. Protein concentrations were quantified by bicinchoninic acid (BCA) protein assay, and 15 μg of protein from each cell lysate were separated by denaturing electrophoresis on 4-20% polyacrylamide gradient gels (Genscript). Proteins were transferred electrophoretically to nitrocellulose membrane and NUP153 was detected by immunoblotting with a monoclonal antibody (Abcam ab171074, 1:1,000 dilution) followed by monoclonal antibody against and anti-GAPDH (Ambion AM4300, 1:4,000 dilution). Proteins were detected with a LiCor Odyssey imager after probing with IRdye-conjugated secondary antibodies. Molecular weight markers were All-Blue Precision Plus Protein standards from Bio-Rad.

Live-cell imaging of virus nuclear entry

Cell lines and reagents

HeLa (ATCC CCL-2) and HEK293T cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (Hyclone, Logan, UT) and penicillin-streptomycin (50 units/ml and 50 μg/ml, respectively; Lonza, Walkersville, MD). All cell lines were maintained in a humidified incubator at 37 °C with 5% CO_2. ON-TARGETplus siRNA duplexes were purchased from Dharmacon. Non-targeting control siRNAs were cat. D-001810-01-05 and D-001810-02-05. NUP153-targeting sequences were: 5'-CAAAUUCGUCUCAAGCAUUA-3' and 5'-GAUAGGAGU-GGGAUAGAUA-3'. Lipofectamine RNAimax transfection reagent was purchased from Invitrogen.

The lentiviral vector pLKO.1-Hygro (gift from Bob Weinberg; Addgene plasmid #24150) was modified to express shRNA targeting NUP153 sequence 5'-AGTGTTCAGTATGCTGTGTTTCT-3', generating pLKO.1-Hygro-NUP153-shRNA. A lentiviral bicistronic vector that expresses mRuby-NUP153 under the control of the human elongation factor 1 (EF1) promoter and puromycin resistance gene (Puro) from porcine teschovirus self-cleaving peptide (P2A) (pLVX-EF1-mRuby-NUP153-P2A-Puro) was created by replacing cleavage and polyadenylation specificity factor 6 (CPSF6) from pLVX-EF1-mRuby-CPSF6-P2A-Puro (27) with NUP153. NUP153 was PCR-

amplified from EGFP(C3)-NUP153 (kind gift from Birthe Fahrenkrog; Addgene Plasmid #64268). NUP153 mutants containing FTFG>4A, RRR>3A, or FTFG>4A+RRR>3A were created using gBlocks (Integrated DNA Technologies, Inc., Coralville, Iowa). The 3' end of NUP153 in pLVX-EF1-mRuby-NUP153-P2A-Puro was replaced with 540-bp long gBlocks containing FTFG>4A, RRR>3A, or FTFG>4A+RRR>3A mutations by replacing a 540-bp fragment using BamHI and AgeI restriction enzyme sites. Next, a 669-bp long gBlock containing a NUP153 fragment that encoded synonymous changes in the nucleotide sequence targeted by NUP153 shRNA (5'-AGTGCAGTGTGTGTTGCGTGAGT-3') was introduced into each vector using NsiI and BstyBI enzyme sites and generated three vectors that express shRNA-resistant WT and mutant mRuby-NUP153 proteins (pLVX-EF1-mRuby-NUP153^{shR}-P2A-Puro, pLVX-EF1-mRuby-NUP153[FTFG>4A]^{shR}-P2A-Puro, pLVX-EF1-mRuby-NUP153[RRR>3A]^{shR}-P2A-Puro, and pLVX-EF1-mRuby-NUP153 [FTFG>4A+RRR>3A]^{shR}-P2A-Puro).

Each lentiviral vector was co-transfected with pC-Help (28), 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 HEK293T cells were filtered and the virus particles were concentrated by ultracentrifugation (100,000 × g) for 1.5 h at 4 ºC through a 20% sucrose cushion (wt/vol) in 1X phosphate buffered saline (PBS). To generate the HeLa-based cell lines expressing WT or mutant mRuby-NUP153, 3 × 10^4 HeLa cells were transduced with the lentiviral vectors expressing wild-type NUP153, or NUP153 mutants FTFG>4A or RRR>3A and selected for puromycin resistance (0.25 µg/ml). After ~10 days, 3 × 10^4 HeLa cells stably expressing the mRuby-NUP153 WT or mutant proteins were transduced with the lentiviral vector pLKO.1-Hygro-NUP153-shRNA. The HeLa-based cell lines expressing shRNA-resistant mRuby-NUP153 and shRNA that targets endogenous NUP153 RNA for degradation were maintained in media containing puromycin (0.5 µg/ml) and hygromycin (200 µg/ml).

Endogenous NUP153 and mRuby-NUP153 were detected by sodium dodecyl sulphate–polyacrylamide gel electrophoresis and western blot analysis using an anti-NUP153 antibody (Abcam; Cat#ab171074, 1:1000 dilution) followed by goat anti-rabbit antibody (IRDye-800CW; LI-COR, Lincoln, NE, 1:10,000 dilution); α-tubulin was used as a loading control and detected using an anti-tubulin (Sigma-Aldrich; Cat#T9026,1:10,000 dilution) followed by goat anti-mouse antibody (IRDye-680RD; LI-COR, 1:10,000 dilution). Western blots were imaged and quantitated using the Odyssey infrared imaging system (LI-COR).

Production and infection of HIV-1 virions labeled with GFP content marker

Infectious virions labeled with the GFP content marker were prepared by co-transfection of 293T cells with HIV-1 based vectors NL4-3 Gag-iGFP ΔEnv [NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH, from Dr. Benjamin Chen (29); Cat#12455] and pHGFP (27) at a 1:2 plasmid ratio (30) and pHCMV-G as previously described. NL4-3 Gag-iGFP ΔEnv is a full-length non-

infectious HIV-1 clone that contains a GFP that is inserted between the MA and CA domains of Gag and is flanked by HIV-1 protease cleavage sites. During virion maturation, GFP is fully processed by HIV-1 protease; some GFP is trapped between the viral membrane and the mature viral core and some GFP is trapped within the mature viral core. pHGFP contains a *gfp* reporter gene in place of *nef* and does not express *env*. Supernatants from the transfected HEK293T cells were collected 24 h post-infection, 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 PBS. HeLa-based cell lines (4 × 10^4 cells/well) were seeded in ibiTreated µ-slides (Ibidi, Gräfelfing, Germany) one day prior to infection. Cells were infected with a normalized number of fluorescently labeled particles via spinoculation at 16°C, which permitted virion binding to cell membranes but prevented virion endocytosis. After centrifugation, cells were placed in a humidified incubator at 37 ºC with 5% CO_2. Cells were fixed with 4.0% (wt/vol) paraformaldehyde (PFA) 6 h post-infection, stained with 4',6-diamidino-2-phenylindole (DAPI; ThermoFisher Scientific Cat#D1306), and imaged by confocal microscopy (described below). To determine virus infectivity, flow cytometry (LSRFortessa; BD Biosciences, San Jose, CA) was used to determine the percentage of GFP^+ cells ~30 h post-infection.

Confocal microscopy and image analysis

Confocal microscopy and image analysis were performed as previously described (27). Briefly, a custom-written MATLAB (Mathworks) program was used to generate a mask of the nuclear envelope (NE) and the nucleus using the DAPI staining and to determine the percentage of total GFP+ particles (cytoplasmic + nuclear) that colocalized with the nucleus mask and the percentage of cytoplasmic GFP particles that colocalized with the NE mask.

CA tube stabilization assay

Purified MBP-NUP153$^{1411\text{-}1475}$ fusion protein or synthetic peptides (GenScript USA) were incubated with pre-assembled WT CA tubes (80 µM) in 50 mM Tris-HCl, 1 M NaCl, pH 8 buffer for 10 min, then diluted 6.67 or 13.3 fold to final 150 mM or 75 mM NaCl, 50 mM Tris-HCl, pH 8 buffer. Negative-stain transmission electron microscopy (TEM) was used to validate the formation of WT CA tubes.

Negative-stain TEM

Capsid assemblies were deposited onto glow-discharged 400 mesh formvar/carbon-coated copper grids (Electron Microscopy Sciences). Grids were then stained with 2% uranyl formate. Imaging was performed on a JEOL JEM-1400 Plus microscope operated at 80 kV with a bottom-mount 4k×3k CCD camera (Advanced Microscopy Technologies).

Statistical analysis

The data analysis was performed using the SPSS 26.0 software package (IBM, United States). All data were expressed in mean ± standard deviation (SD) if not specifically addressed. One-way ANOVA and Tukey's multiple comparisons test were applied to evaluate the differences in the co-pelleting assays. The Student's t-test with variances verified by Fisher's F-test was applied to evaluate differences in TRIM-fusion constructs HIV-1 restriction assay. The two-tailed Welch's t-test was applied to evaluate differences in the co-pelleting assay with cell lysate and GFP-labeled HIV-1 core nuclear entry and infection assay.

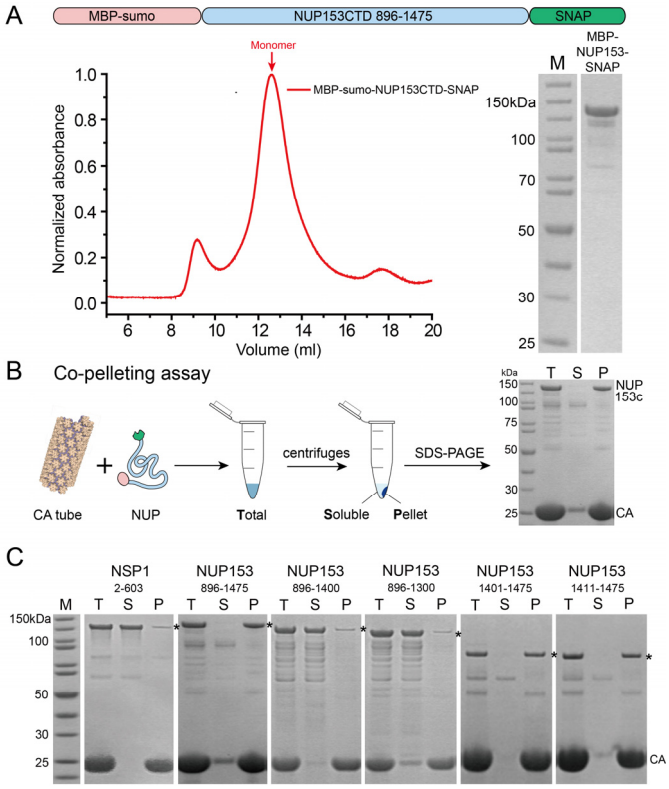

Fig. S1. NUP153$_{CTD}$ purification and CA tube co-pelleting assay (A) Top: Schematic of the MBP-sumo-NUP153-SNAP construct. Bottom-left: Size exclusion chromatography analysis shows that NUP153 is mainly monodispersed monomer in solution. Bottom-right: SDS-PAGE analysis of protein purity. (B) Schematic of the CA tube co-pelleting assay. Total (T), supernatant (S), and pellet (P) fractions of a NUP153$_{CTD}$ binding reaction are shown. (C) Co-pelleting assays using A14C/E45C disulfide crosslinked CA tubes in 25 mM Tris-HCl, 75 mM NaCl buffer with indicated NUP proteins, analyzed by SDS-PAGE; *, NUP. The gel with NUP153$^{896\text{-}1475}$ is the same as shown in panel B. Co-pelleting assays were performed three times, and a representative example is shown.

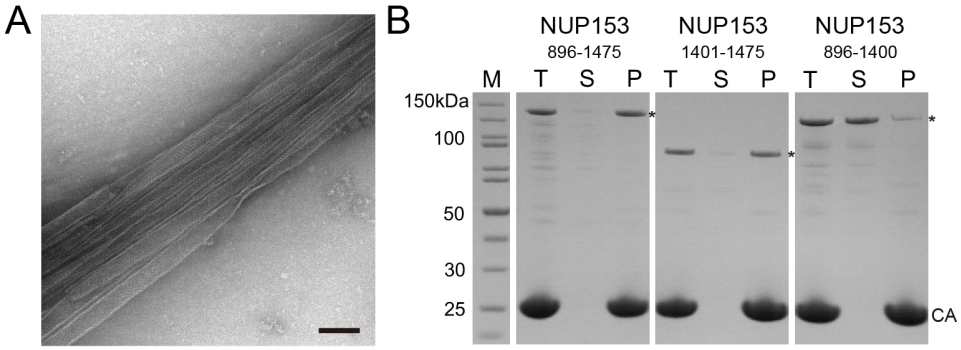

Fig. S2. Co-pelleting assays using IP6-stabilized wild-type (WT) CA tubes (100 μM CA) and NUP153 variants (3 μM). (A) Negative-stain TEM micrograph validating the formation of IP6 (1 mM) stabilized WT CA tubes under 150 mM NaCl. Scale bar: 100 nm. (B) Co-pelleting assays using WT CA tubes in 1× PBS buffer (150 mM NaCl) containing 1 mM IP6, analyzed by SDS-PAGE. The asterisks mark the NUP153 construct positions. Co-pelleting assays were performed three times, and a representative example is shown.

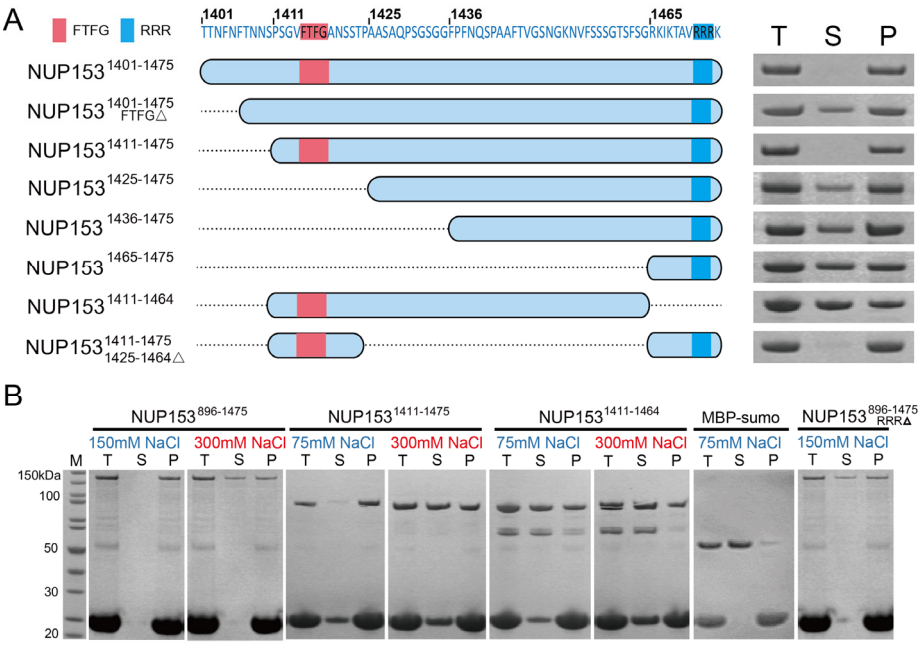

Fig. S3. Minimal NUP153 constructs for maximal CA tube binding. (A) Co-pelleting assay of NUP153 variants with CA tubes in a 25 mM Tris-HCl, 75 mM NaCl buffer. Middle: Schematics of various truncated NUP153$_{CTD}$ segments. Right: Total (T), Soluble (S), and Pellet (P) fractions of co-pelleting assays using A14C/E45C disulfide crosslinked CA tubes analyzed by SDS-PAGE. (B) Results of the co-pelleting assay under different salt conditions. Co-pelleting assays were performed three times, with representative examples shown.

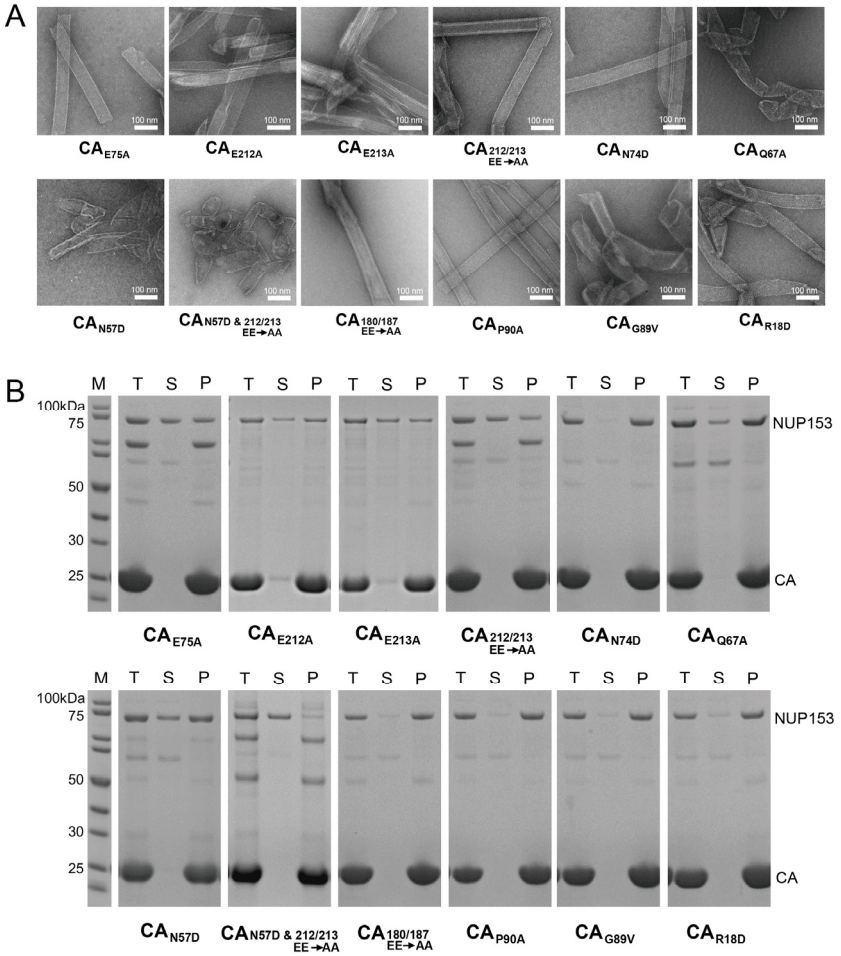

Fig. S4. Mapping the NUP153$^{1411\text{-}1475}$ binding site using CA mutations. (A) Negative-stain TEM validation of the formation of various mutant CA tubes. (B) Co-pelleting assays using mutant CA tubes in 25 mM Tris-HCl, 150 mM NaCl buffer, analyzed by SDS-PAGE. Co-pelleting assays were performed three times, with representative gel images shown.

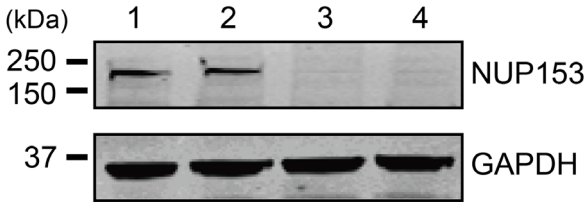

Fig. S5. Immunoblot analysis of NUP153 depletion. NUP153 was detected by immunoblotting with a monoclonal antibody (top), then the blot was reprobed with a monoclonal antibody against GAPDH (bottom). Lanes 1 and 2 contain lysates of cells transfected with two different non-targeting siRNAs, and lanes 3 and 4 from cells transfected with two different NUP153-targeting siRNAs.

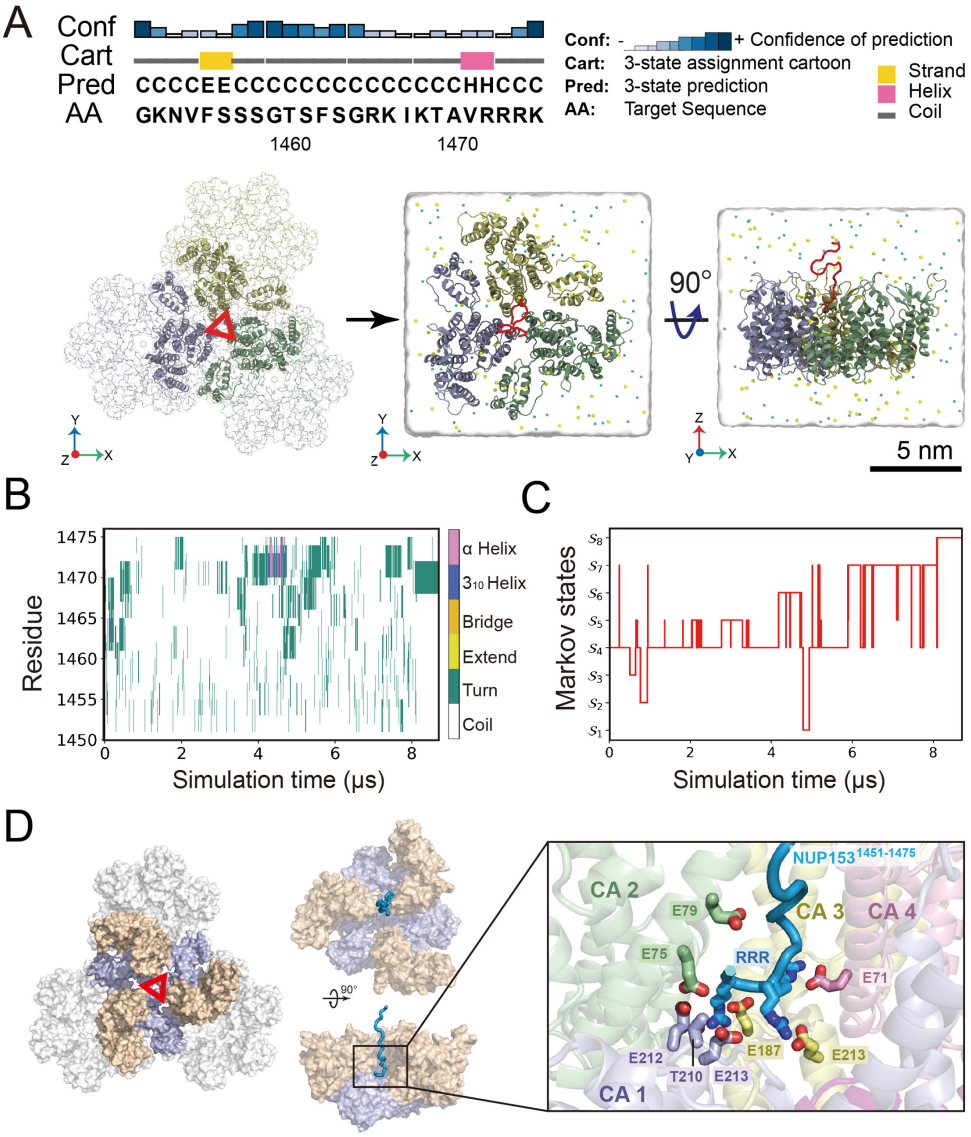

Fig. S6. All-atom MD simulations identify NUP153$_{CTD}$ interactions at the CA tri-hexamer center. (A) Top: the sequence of NUP153$^{1451\text{-}1475}$ and predicted secondary structure by PSIPRED (31). Bottom: the RRR-motif NUP153$^{1451\text{-}1475}$-CA tri-hexamer$_{center}$ complex model used in MD simulations. Three dimers from different hexamers are in blue, yellow, and green (left). The NUP153$^{1451\text{-}1475}$ peptide in red was placed above the tri-hexamer region before simulations (middle and right). (B) The evolution of NUP153$^{1451\text{-}1475}$ secondary structures during the MD simulation, which was computed by the STRIDE program (32) in VMD (33). (C) Conformational transitions and distributions of NUP153$^{1451\text{-}1475}$ during the simulation. Eight Markov states of NUP153$^{1451\text{-}1475}$ in the CA tri-hexamer interface were identified. (D) Surface representation of the tri-hexamer center region (left) and simulated NUP153$^{1451\text{-}1475}$ (light blue cartoon) binding to tri-hexamer$_{center}$ (middle) in state 8. The RRR-motif is surrounded by negatively charged glutamate residues (sticks) from 4 CA monomers (right).

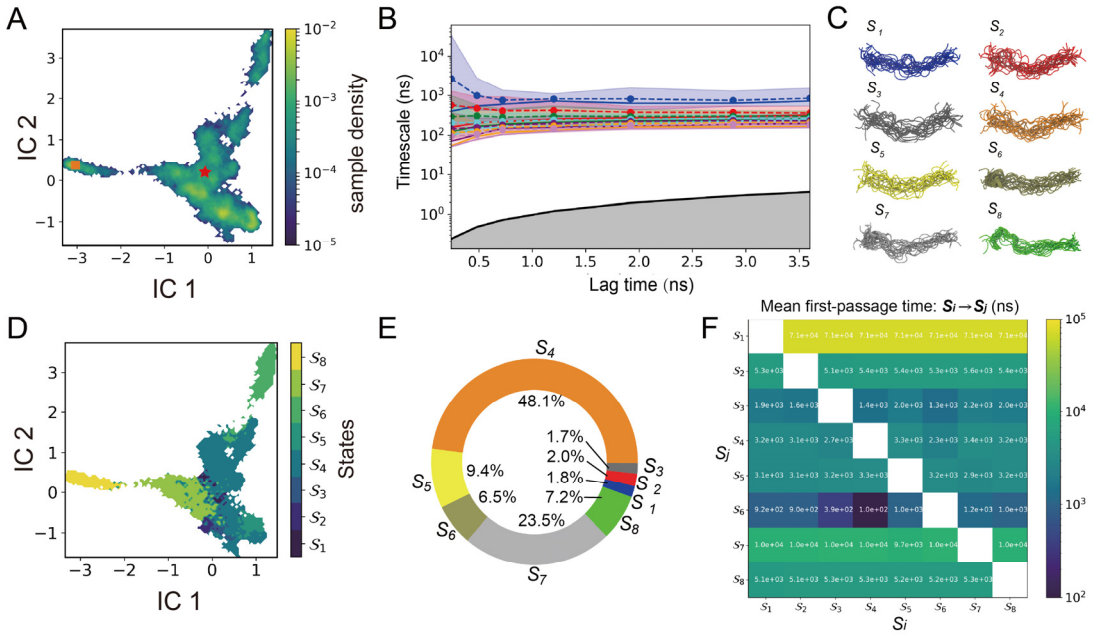

Fig. S7. The conformational dynamics of NUP153$^{1451\text{-}1475}$ at the CA tri-hexamer interface. (A) The distribution of NUP153$^{1451\text{-}1475}$ conformations projected onto the first two independent components (ICs) space. The red star and orange square mark the starting and ending conformations of NUP153$^{1451\text{-}1475}$ in the MD simulation. (B) Optimization of the lag-time for constructing Markov state models (MSMs). The seven slowest relaxation timescales of the MSMs as a function of the lag-time are shown. The MSM results reported in the present study were obtained with a lag-time of 1.2 ns, after which the timescales are nearly constant. (C) Eight Markov states of NUP153$^{1451\text{-}1475}$ at the CA tri-hexamer interface were identified, each shown with twenty representative NUP153$^{1451\text{-}1475}$ structures. (D) Projections of eight Markov states onto the first two IC space. (E) Relative populations of the Markov states. (F) The mean free-passage times (MFPTs) of transitions between Markov states.

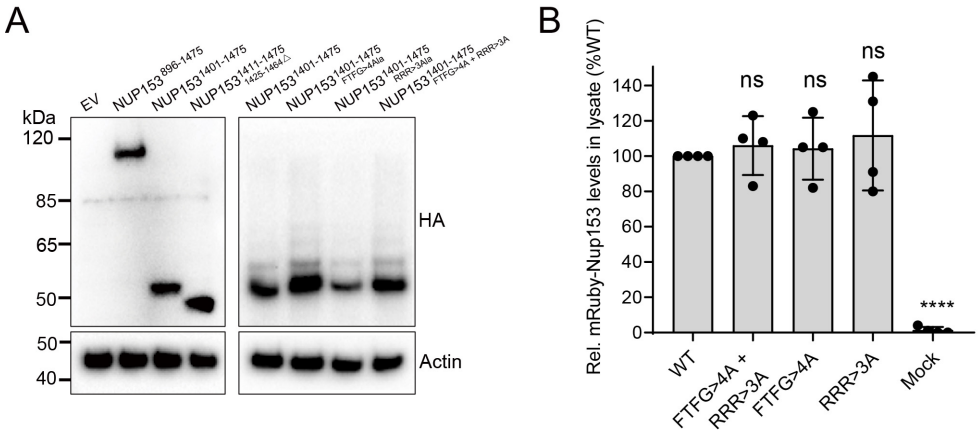

Fig. S8. (A) Immunoblot showing expression of TRIM-NUP153$_{CTD}$ fusion constructs in HEK293T cells detected with anti-HA antibody; β-actin was blotted as a loading control. (B) Relative mRuby-NUP153 and mutant expression levels for the CA co-pelleting assay, normalized by WT full-length mRuby-NUP153.

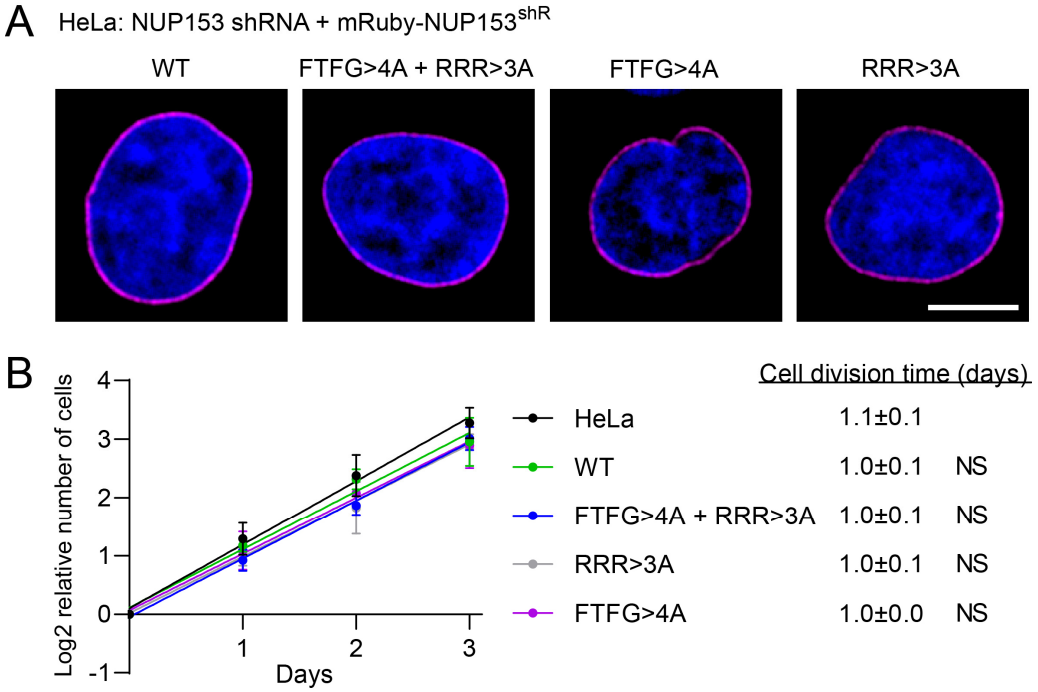

Fig. S9. Characterization of mRuby-NUP153 expressing HeLa cell lines. (A) Confocal images of representative cells showing localization of mRuby-NUP153 WT or mutants (magenta) to the nuclear envelope. Cell nuclei were counterstained using DAPI (blue). Scale bar, 10 µm. (B) Growth kinetics of parental HeLa and mRuby-NUP153 expressing HeLa cell lines. Cell growth kinetics were measured by cell counts over a three-day period. The relative values were log2-transformed, plotted, and fitted using a simple linear regression (solid lines). The slope of the line was used to determine the cell division time. Results are mean ± standard deviation of 3 independent experiments. Statistical significance was determined by two-tailed Welch's t-test; NS, not significant ($P \geq 0.05$).

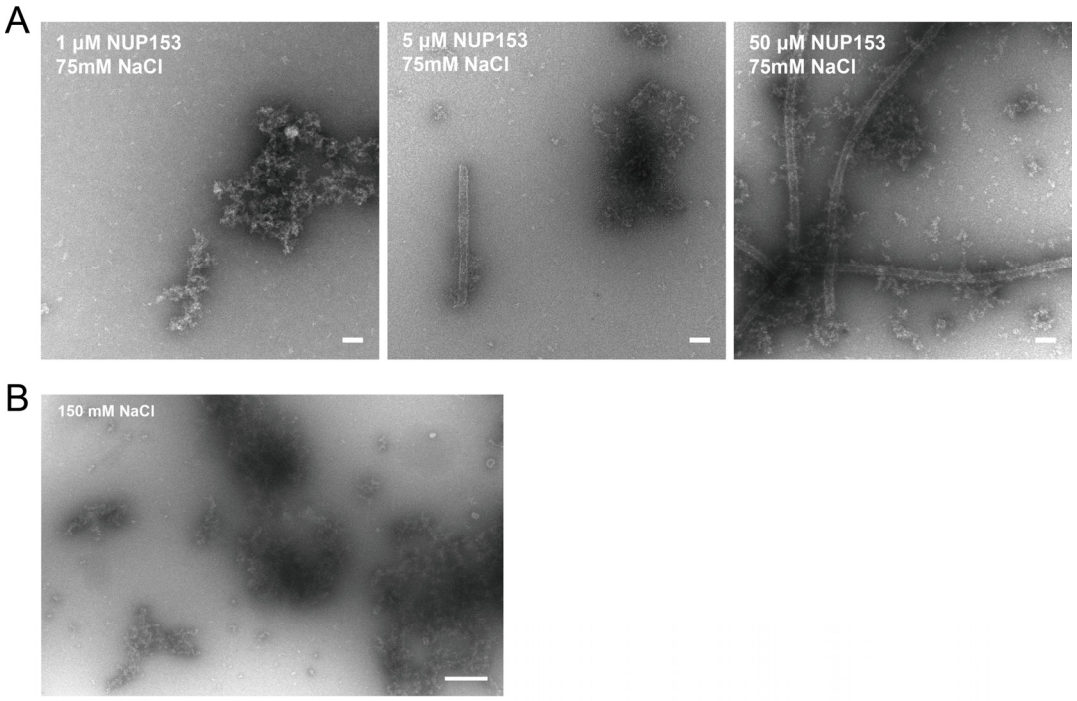

Fig. S10. Characterizations of CA assemblies stabilized by NUP153 fragments. (A) The pre-assembled WT CA tubes (6 μM CA) incubated with different concentrations of NUP153$^{1411\text{-}1475}$ (1, 5, or 50 μM) at 75 mM NaCl. Scale bars: 100 nm. (B) Negative-stain electron micrographs of WT CA tubes under 150 mM NaCl conditions. Scale bars: 500 nm.

Table S1. NUP153$^{1451\text{-}1475}$-CA tri-hexamer$_{center}$ inter-residue contacts categorized by Markov states.

Markov states	Nup153 residues	CA residues	Occupancy (%)	Types
State 1	GLY1464	ARG82	100.0	
	ARG1465	ASP81	98.4	
	GLY1464	ASP81	96.9	
	SER1463	ASP81	93.8	
	SER1463	ARG82	93.8	
State 2	ARG1465	TRP80	100.0	
	ARG1465	LEU83	98.6	
	ARG1465	HIS84	97.3	
	PHE1462	PRO125	95.9	
	ILE1467	LEU83	94.5	
	ARG1465	GLU98	94.5	
	PHE1462	ILE129	91.8	
	GLY1464	TRP80	90.4	
State 3	PHE1462	PRO99	96.7	
	PHE1462	PRO125	96.7	
	PHE1462	ILE129	96.7	
	PHE1462	GLU98	95.1	
	LYS1466	GLU79	95.1	
	LYS1466	GLU76	95.1	
	ARG1474	ARG82	95.1	
	PHE1462	ILE124	91.8	
	PHE1462	VAL126	90.2	
	ILE1467	LEU83	90.2	
	GLY1464	ILE129	90.2	
	ARG1473	ARG82	90.2	
State 6	ILE1467	ARG82	92.4	
	THR1469	ARG82	90.7	
State 8	ILE1467	ASP81	100.0	Hydrogen bond
	ARG1472	THR210	98.9	Hydrogen bond
	LYS1466	ARG82	98.1	Hydrogen bond / Electrostatic
	ILE1467	ARG82	98.1	Hydrogen bond
	ARG1473	GLU213	95.8	Hydrogen bond / Electrostatic
	ARG1473	GLY208	94.6	Hydrogen bond
	ILE1467	ALA78	93.5	Hydrophobic
	ALA1470	ASN74	92.0	Hydrogen bond

Only contacts with occupancies greater than 90% are shown. In Markov state 4, 5, and 7, NUP153 was found to have no residue contact with an occupancy greater than 90% with CA. The NUP153-CA residue contacts were defined as the NUP153-CA inter-residue distance smaller than a cutoff distance (3.4 Å), longer range electrostatic interactions were not shown. The contact occupancy was calculated by $100\% \times N_{ij}/n$, where N_{ij} is the number of occurrences of a residue contact between NUP153 residue 'i' and CA residue 'j' in the simulation and 'n' is the total number of the MD simulation snapshots. The types of top residue contacts in state 8 are mostly hydrogen bonds.

Table S2. The top inter-residue contacts from the 8.7 μs MD simulation of the NUP153$^{1451-1475}$-CA tri-hexamer$_{center}$ interaction.

NUP153 RRR-motif	CA	Occupancy (%)
LYS1468	GLU79	62.5
ILE1467	ARG82	41.3
ARG1465	LEU83	39.2
THR1469	ARG82	36.2
ARG1472	THR210	35.7
ARG1474	GLU213	33.7
ARG1474	THR210	33.3
ARG1473	THR210	32.4
ARG1473	GLU213	32.3
ARG1472	GLU213	31.9
ARG1465	TRP80	31.6
ARG1474	GLY208	31.0
PHE1462	PRO99	30.5
PHE1462	ILE124	30.4
ARG1465	PRO85	29.4
PHE1462	PRO125	29.3
ARG1472	GLU75	29.2
PHE1462	ILE129	28.1

Only contacts from the top 18 occupancies are shown. The NUP153-CA residue contacts were defined as the NUP153-CA inter-residue distance smaller than a cutoff distance (3.4 Å). The contact occupancy was calculated by $100\% \times N_{ij}/n$, where N_{ij} is the number of occurrences of a residue contact between NUP153 residue 'i' and CA residue 'j' in the simulation and 'n' is the total number of the MD simulation snapshots.

Table S3. Primers used to generate NUP153 variants

Primer	Sequence
NUP153-F	ATGCCATATGGAAAATTTATACTTCCAAGGTAACTCAGCAG CCTCCTCATCCTTCA
NUP153-R	ATGCCTCGAGTCAGCAAAGCTTTTTCCTGCGTCTAACAGC AGTCTTT
SNAP-F	ATGTGCCCAAGCTTATGGACAAAGATTGCGAAATGAAACG TACCACCC
SNAP-R	ATGCCGCTCGAGTCATCCCAGACCCGGTTTACCCAGACG A
NUP153 1475 F	ATGGACAAAGATTGCGAAATGAAACGTAC
NUP153 1400 R	GCTGCTGCCAAACTGGAAAGCCGAA
NUP153 1300 R	TGCAGAGCTAGATGTGGTTGTGGCT
NUP153 1401 F	ACTACAAATTTCAACTTCACAAACAACAGTCCATCAG
NUP153 1411 F	CCATCAGGAGTGTTCACATTTGGTGCAAATTC
NUP153 1425 F	GCAGCCTCAGCCCAGCCTTC
NUP153 1436 F	TTTCCATTTAACCAGTCTCCAGCAGCAT
NUP153 1465 F	CGCAAGATAAAGACTGCTGTTAGACGCA
NUP153 896 R	ACCTTGGAAGTATAAATTTTCCATATGTCCACCAATCT
NUP153 1424 R	AGGTGTGCTAGAATTTGCACCAAATGTG
NUP153 1464 R	ACCAGAGAATGAAGTTCCAGAAGAAGAGA
NUP153 1415FTFG F	AACAGTCCATCAGGAGTGGCAAATTCTAGCACACC
NUP153 1415FTFG R	GGTGTGCTAGAATTTGCCACTCCTGATGGACTGTT
Trim F	GATCTCGAGCTCAAGCTTCGAATTC
Trim R	GTAATCTGGAACATCGTATGGGTAGCCACCTCCAGATCCC CAGTAGCGTCGG
NUP153 F	CATACGATGTTCCAGATTACGCTGGAGGTGGATCTCCGCG GAACTCAGCAGCCTCCTC
NUP153 R	GATCCCGGGCCTTTATTTCCTGCGTCTAACAG
Trim-HA-NUP153 F	GATCTCGAGCTCAAGCTTCGAATTC
Trim-HA-NUP153 R	GATCCCGGGCCTTTATTTCCTGCGTCTAACAG
NUP153 1401-1475 F	CAGCCGCGGACTACAAATTTCAACTTCACAAACAACAG
NUP153 1401-1475 R	GATCCCGGGCCTTTATTTCCTGCGTCTAACAG
NUP153 25AA	CAGCCGCGGCCATCAGGAGTGTTCACATTTGGTGCAAAT TCTAGCACACCTCGCAAGATAAAGACTGCTGTTAGACGCA GGAAATAAAGGCCCGGGATC
NUP153 RRR deletion F	AAAAAGCTTATGGACAAAGATTGCGAAATGA
NUP153 RRR deletion R	AACAGCAGTCTTTATCTTGCGACCA

F, forward primer; R, reverse primer

Movie S1. Dynamics of NUP153$^{1451-1475}$ at CA tri-hexamer interface from MD simulation. The NUP153 peptide is shown as a red ribbon, while the HIV-1 CA N-terminal domain (NTD) and CTD are shown as golden and blue ribbons, respectively. NUP153$^{1451-1475}$ was found to be mostly unstructured during the simulations, except for a short α-helix at the C-terminus (residue 1470 to 1475) that appeared temporarily.

SI References

1. B. J. Summers *et al.*, Modular HIV-1 Capsid Assemblies Reveal Diverse Host-Capsid Recognition Mechanisms. *Cell Host Microbe* **26**, 203-216 e206 (2019).
2. J. Wen, T. Arakawa, J. S. Philo, Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions. *Anal Biochem* **240**, 155-166 (1996).
3. P. J. Wyatt, Light-Scattering and the Absolute Characterization of Macromolecules. *Anal Chim Acta* **272**, 1-40 (1993).
4. N. Eswar *et al.*, Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* **Chapter 5**, Unit-5 6 (2006).
5. M.-y. Shen, A. Sali, Statistical potential for assessment and prediction of protein structures. *Protein Science* **15**, 2507-2524 (2006).
6. J. C. Phillips *et al.*, Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802 (2005).
7. W. L. Jorgensen, C. Jenson, Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density. *Journal of Computational Chemistry* **19**, 1179-1186 (1998).
8. D. E. Shaw *et al.* (2014) Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (IEEE Press, New Orleans, Louisana), pp 41-53.
9. J. Huang *et al.*, CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* **14**, 71-73 (2017).
10. R. A. Lippert *et al.*, Accurate and efficient integration for molecular dynamics simulations at constant temperature and pressure. *J Chem Phys* **139**, 164106 (2013).
11. Y. Shan, J. L. Klepeis, M. P. Eastwood, R. O. Dror, D. E. Shaw, Gaussian split Ewald: A fast Ewald mesh method for molecular simulation. *J Chem Phys* **122**, 54101 (2005).
12. G. Pérez-Hernández, F. Paul, T. Giorgino, G. D. Fabritiis, F. Noé, Identification of slow molecular order parameters for Markov model construction. *The Journal of Chemical Physics* **139**, 015102 (2013).
13. J. D. Chodera, F. Noé, Markov state models of biomolecular conformational dynamics. *Curr Opin Struc Biol* **25**, 135-144 (2014).
14. B. E. Husic, V. S. Pande, Markov State Models: From an Art to a Science. *J Am Chem Soc* **140**, 2386-2396 (2018).
15. W. Wang, S. Cao, L. Zhu, X. Huang, Constructing Markov State Models to elucidate the functional conformational changes of complex biomolecules. *WIREs Computational Molecular Science* **8**, e1343 (2018).
16. M. K. Scherer *et al.*, PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J Chem Theory Comput* **11**, 5525-5542 (2015).
17. J. MacQueen (1967) Some methods for classification and analysis of multivariate observations. in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Oakland, CA, USA), pp 281-297.
18. S. Röblitz, M. Weber, Fuzzy spectral clustering by PCCA+: application to Markov state models and data classification. *Advances in Data Analysis and Classification* **7**, 147-179 (2013).
19. F. Noe, C. Schutte, E. Vanden-Eijnden, L. Reich, T. R. Weikl, Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc Natl Acad Sci U S A* **106**, 19011-19016 (2009).
20. J.-H. Prinz *et al.*, Markov models of molecular kinetics: Generation and validation. *The Journal of Chemical Physics* **134**, 174105 (2011).
21. M. C. Shun *et al.*, LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes & Development* **21**, 1767-1778 (2007).
22. K. A. Matreyek, S. S. Yucel, X. Li, A. Engelman, Nucleoporin NUP153 phenylalanine-glycine motifs

engage a common binding pocket within the HIV-1 capsid protein to mediate lentiviral infectivity. *Plos Pathog* **9**, e1003693 (2013).
23. J. K. Yee *et al.*, A General-Method for the Generation of High-Titer, Pantropic Retroviral Vectors - Highly Efficient Infection of Primary Hepatocytes. *P Natl Acad Sci USA* **91**, 9564-9568 (1994).
24. E. Serrao, P. Cherepanov, A. N. Engelman, Amplification, Next-generation Sequencing, and Genomic DNA Mapping of Retroviral Integration Sites. *Jove-J Vis Exp* 10.3791/53840 (2016).
25. S. Jang *et al.*, Differential role for phosphorylation in alternative polyadenylation function versus nuclear import of SR-like protein CPSF6. *Nucleic Acids Res* **47**, 4663-4683 (2019).
26. R. Lu *et al.*, Class II integrase mutants with changes in putative nuclear localization signals are primarily blocked at a postnuclear entry step of human immunodeficiency virus type 1 replication. *Journal of Virology* **78**, 12735-12746 (2004).
27. R. C. Burdick *et al.*, HIV-1 uncoats in the nucleus near sites of integration. *P Natl Acad Sci USA* **117**, 5486-5493 (2020).
28. H. Mochizuki, J. P. Schwartz, K. Tanaka, R. O. Brady, J. Reiser, High-titer human immunodeficiency virus type 1-based vector systems for gene delivery into nondividing cells. *Journal of Virology* **72**, 8873-8883 (1998).
29. W. Hubner *et al.*, Sequence of human immunodeficiency virus type 1 (HIV-1) gag localization and oligomerization monitored with live Confocal Imaging of a replication-competent, fluorescently tagged EIV-1. *Journal of Virology* **81**, 12596-12607 (2007).
30. C. L. Li, R. C. Burdick, K. Nagashima, W. S. Hu, V. K. Pathak, HIV-1 cores retain their integrity until minutes before uncoating in the nucleus. *P Natl Acad Sci USA* **118**, e2019467118 (2021).
31. L. J. McGuffin, K. Bryson, D. T. Jones, The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404-405 (2000).
32. D. Frishman, P. Argos, Knowledge-based protein secondary structure assignment. *Proteins-Structure Function and Genetics* **23**, 566-579 (1995).
33. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *J Mol Graph Model* **14**, 33-38 (1996).