Supplementary Methods

Protein expression and purification

The A92E and A204C/D/E/L mutant CA proteins were expressed in *E. coli* cells and purified by ion exchange chromatography, as described previously with a minor modification[29]. Briefly, cell lysate was loaded onto a 5 mL Hi-Trap SP column (GE Healthcare) equilibrated with a buffer containing 25 mM sodium phosphate, pH 5.8, 0.02% sodium azide and 1 mM DTT. Proteins were eluted with a linear gradient of 0-1 M NaCl at a flow rate of 2 mL/min. The peaks containing CA were pooled and loaded onto a Hi-Load Superdex75 26/60 column (GE Healthcare) equilibrated with a buffer containing 25 mM sodium phosphate, pH 7.5, 0.02% sodium azide and 1 mM DTT at a flow rate of 2 mL/min. The CA proteins were concentrated to a stock of 5-10 mg/ml using Amicon concentrators (Millipore).

Cryo-EM specimen preparation and data collection

Full-length CA A92E proteins (10 mg/ml) were diluted to 2 mg/ml in high salt buffer (1M NaCl, 50mM Tris pH 8.0) and incubated at 37°C for 1 hr for tubular assembly. Sample (2.5 μl) was applied to the carbon side of a glow discharged perforated Quantifoil grid (Quantifoil Micro Tools, Jena, Germany), followed by application of 3 μl of low salt buffer (100 mM NaCl, 50mM Tris pH 8.0) on the back side of the grid, and blotting, from the back side, with a filter paper, before plunge-freezing in liquid ethane using a manual gravity plunger. Low dose (~15 $e^-/Å^2$) projection micrographs of tubes embedded in a thin layer of vitreous ice were collected on a Tecnai Polara G2 microscope (FEI Corp., OR.) operated at 200 kV at liquid nitrogen temperature. Images were recorded on Kodak SO163 films at a nominal magnification of 59,000, using under-focus values ranging from 1.0 to 3.5 μm. The best micrographs were selected and digitized using a Nikon super coolscan 9000 ED scanner (Nikon, Japan) at an optical resolution of 4000 dpi.

Three-dimensional reconstruction

We collected more than 1500 films of tubes possessing a wide range of diameters and helical symmetries. Twenty seven tubes with (-12, 11) symmetry were selected and boxed into segments of size 664 × 664 pixels with an overlap of 90% along the helical axis by the EMAN BOXER program[30]. The defocus value of each micrograph was evaluated using the program CTFFIND[31]. An initial density map was reconstructed from binned segments of 332 × 332 pixels using Iterative Helical Real Space Reconstruction[32]. The refined helical parameters corresponded to a rise of 7.24 Å and a rotation angle of 328.87º. The initial map was unbinned to 664 × 664 × 664 to serve as a reference for further refinement using Frealign[33]. For refinement, each segment was corrected for its contrast transfer function (CTF) and aligned with the reference. Several rounds of iterative refinement, using mode 4, were carried out until there was no improvement in the phase residual of each segment. A total of 3210 helical segments, containing 38,520 asymmetric units, were included in the final reconstruction. To eliminate the density gradient along the tube axis, due to the reconstruction algorithm[34], the beautify function in Frealign was used to impose real space symmetry on the final reconstruction. The final reconstruction was masked with a cylindrical mask. The B-factor was estimated to be –500 $Å^2$ using EM-BFACTOR[35] and was applied to the final map with cosine edge mask using the program bfactor. Two-fold symmetry, perpendicular to the surface lattice, was imposed using the software IHRSR++[36]. The Fourier shell correlation (FSC) curve was calculated from two density maps reconstructed from odd and even half data sets; each was aligned and refined separately against a 20 Å initial reference volume (low-pass filtered from the final reconstruction), and reconstructed (with two-fold symmetry imposed) independently using mode 4 in Frealign. The resolution was judged to be 8.6 Å at a 0.5 cut-off value of FSC.

Cryo-ET sample preparation, data collection, and 3D reconstruction

Purified HIV-1 A14C/E45C cores (3 µl) were applied to the carbon side of glow discharged perforated R2/2 Quantifoil grids and quickly mixed with 3 µl of a 15 nm fiducial gold bead solution before plunge-freezing using a manual gravity plunger. Tomography tilt series were collected at a nominal magnification of 39,000 x (effective pixel size of 3.12 Å) on a Tecnai G2 Polara electron microscope operated at 300 kV. A

series of images were recorded on a Gatan 4K×4K CCD camera by tilting the specimen from -70° to 66° in increments of 3° (<45°) and 2° (>45°) tilt angle. Altogether 53 images were collected in one tilt series with a dose of ~120 $e^-/Å^2$. Images were recorded at a defocus value of ~8 μm using FEI batch tomography software.

The IMOD package[37,38] was used to align tilted projection images with the aid of fiducial gold markers. The final aligned image stack was binned to a pixel size of 6.24 Å and low-pass filtered to 25Å. A simultaneous iterative reconstruction technique (SIRT), implanted in program TOMO3D[38], was used to generate the final 3D density map from the aligned image stack. For surface rendering, the tomogram was denoised using the nonlinear anisotropic diffusion, edge enhancing algorithm, available in IMOD and displayed using the program CHIMERA[39]. A total of 26 HIV-1 cores were reconstructed and examined in CHIMERA for structural integrity. The cone angles from these cores were measured in projection images. The best core, which displays a cone angle of 23°, was selected for all-atom model building.

Normal mode flexible fitting

A pseudoatomic model was built by normal mode flexible fitting three crystal NTDs (3H47) and solution NMR CTDs (2KOD) into three asymmetric subunit densities. The normal mode flexible fitting of NTD generates a model that is very close to the crystal structure with a Cα RMSD of 0.8 Å. Three CTD were fitted into the three asymmetric CTD densities using "Fit in Map" in CHIMERA[39]. The density of three CTD was colored using "Color Zone" and cut out using "Split Map". Flexible fitting of CTD to the density was carried out using a normal-mode based software, IMODFIT[40]. The dihedrals of all the residuals within four α-helices were fixed and only the loops were allowed to be flexible. A density cut-off was chosen so that the density contained 100 percent volume of CA protein. The flexible fitting program, Normal Mode Flexible Fitting (NMFF)[41], was also used. Both methods produced similar models. The model presented in Supplementary Fig. 2a was built using IMODFIT.

Molecular dynamics flexible fitting

Molecular dynamics flexible fitting (MDFF)[42] incorporates the EM density map as an external potential in atomistic molecular dynamics (MD) simulations. The MD forcefield preserves the stereochemical quality of the structure, while forces proportional to the EM density gradient flexibly fit the structure to the experimental map. All MDFF and atomistic MD simulations (Supplementary Table 1) were performed using NAMD 2.9[43] and the CHARMM22 forcefield[44] with the CMAP correction[45]. Periodic boundary conditions were assumed to resemble experimental conditions; all simulations were carried out at T=320K and 1 bar, employing the Langevin piston algorithm[46]. Long-range Coulomb forces were computed through the Particle Mesh Ewald method with a grid spacing of 1 Å[47]. All simulations involved the r-RESPA multi-timestep algorithm[48] with an integration timestep of 2 fs, evaluating non-bonded interactions every 2 fs, and evaluating full electrostatics every 4 fs; all hydrogen bonds were constrained with the SHAKE algorithm, unless specified otherwise.

The hexameric structure 3H47[49] was used as the starting point for the MDFF calculations and dimers between hexamers were modeled after the NMR structure 2KOD[50]; missing residues were built by homology modeling using MODELLER[51]. Seven of the modeled hexameric structures were docked to the cryo-EM map using SITUS[52] with a resulting cross correlation coefficient of 0.75.

The resulting 7-hexamer patch was then solvated, using TIP3P water, and ionized to 1M NaCl. Secondary structure restraints were applied to helices 1 to 11, using a force constant k = 200 kcal mol $^{-1}$ Å^{-2} [42]. MDFF was run for 10 ns with an integration timestep of 2 fs and a scaling factor ζ = 0.5 kcal mol $^{-1}$, applying symmetry restraints between CA monomers along the chiral axis[53]. MDFF yielded a structure (Supplementary Table 1, simulation 1) with a cross-correlation coefficient of 0.96.

Equilibrium molecular dynamics simulations

The MDFF-derived HOH structure was equilibrated for 425 ns using 2 fs timesteps (Supplementary Table 1, simulation 2). During the first 10 ns of the simulation, helices 1 to 11 (CTDs and NTDs) were harmonically restrained using an isotropic force constant k

= 10 kcal mol $^{-1}$ Å^{-2}; during the following 15 ns, only helices 1 to 7 (NTDs) were harmonically restrained. For the remaining simulation time, no restraints were applied (Supplementary Video 2).

A total of 71 MDFF-derived hexamers were docked using SITUS to the cryo-EM cylindrical map, solvated, and ionized to 1 M NaCl. The resulting system, consisting of 13 million atoms, was equilibrated following the same protocol as simulation 1, for 125 ns on the NSF-funded Blue Waters supercomputer and employing NAMD2.9 with an integration timestep of 2 fs (Supplementary Table 1, simulation 3).

The POH model was constructed in three stages. First, the pentameric structure 3P05 was employed as a starting point[54]; residues missing in the crystal structure were modeled using MODELLER[51]. The isolated pentamer system was solvated and ionized to 300 mM NaCl. To achieve relaxation of modeled loops, the system was equilibrated for 10 ns, while applying harmonic restraints only to the crystallographically solved residues using an isotropic force constant k = 10 kcal mol $^{-1}$ Å^{-2} and 2 fs timesteps (Supplementary Table 1, simulation 4). Second, a simulation was started from a relatively flat conformation of POH, to capture the quaternary rearrangement of the entire structure (Supplementary Table 1, simulation 5). At the start of simulation 5, five MDFF-derived hexameric models were placed around the equilibrated pentamer model (results from simulation 4) such that dimeric interfaces were aligned with those from the equilibrated HOH structure (results from simulation 1). The resulting structure (POH-0) turned out to be relatively flat; it was then solvated and ionized to 300 mM NaCl involving finally 1,046,161 atoms. The resulting POH model was allowed to equilibrate for 300 ns, without restraints, as shown in Supplementary Video 2 (Supplementary Table 1, simulation 5). The equilibrated structure from simulation 5 (POH-1) adopted the dome-like shape shown in Fig. 3d. The total RMSD change relative to the initial model was 5.5 ± 2.0 Å for the hexamers, 5.6 Å for the pentamer, and 17.2 Å for the entire POH-1 structure.

Once the orientation of the hexamers relative to the pentamer was determined from simulation 5, the resulting quaternary arrangement was used for a second simulation of POH for further refinement of interfaces (Supplementary Table 1, simulation 6). At the start of simulation 6, the hexamers and pentamers from simulation 5 were replaced by the structures derived from simulation 1 and simulation 4, respectively (structure POH-2). Structure POH-2 was further refined and equilibrated for an additional 1200 ns (Supplementary Table 1, simulation 6). In total, the structure of POH was simulated for over 1500 ns. The POH structure simulations (Supplementary Table 1, simulations 4 to 6) were performed using a Graphics Processor Unit-accelerated version of NAMD2.9[55] on the NSF-funded Blue Waters and on the DOE-funded Oakridge Jaguar supercomputers.

VMD[56] was employed for the preparation of figures and analysis of the hexameric tubular assembly, the MDFF-derived HOH structure and the POH structures.

Stability of models tested through molecular dynamics simulations

The structures resulting from the equilibrium molecular dynamics simulations described above (HOH, POH, entire hexameric tubular assembly, and complete capsid) were analyzed for stability. The quality of the models was judged first from a calculation of the root mean square distance (RSMD) of the protein structure, over time, relative to the initial structure (including only Cα atoms). As a second measure of stability, root mean square fluctuations (RMSF) of the protein Cα atoms were determined for the last 100 ns of each simulation. The average RMSF turned out to be 1.7 ± 0.4 Å for the HOH model (Supplementary Table 1, simulation 2), 2.2 ± 0.9 Å for the POH model (Supplementary Table 1, simulation 6), 2.2 ± 0.6 Å for the tubular assembly (Supplementary Table 1, simulation 3), and 2.9 ± 1.1 Å for the entire capsid as shown in Fig. 4 (Supplementary Table 1, simulation 7). The calculated RMSF and RMSD values are similar to those judged acceptable in simulations of several proteins[57]. Therefore, the structures reported here are considered to be stable and biologically relevant.

Isomer search, geometry optimization and density generation

Four classes of isomers of fullerenic cages, containing 166, 186, 216, or 252 hexagons and 12 pentagons, were generated using the spiral algorithm[58,59]. Within each class, main shape differences between members are due to the location of the twelve pentagons (pentagon locations are indicated following the International Union of Pure and Applied Chemistry (IUPAC) nomenclature using the ring spiral pentagon indices - RSPI[58,60]). The program Fullerene 4.4 (kind gift from Drs. Peter Schwerdtfeger, Lukas Wirz and James Avery, The New Zealand Institute for Advanced Study) was used for an extensive neighborhood search around the 12 RSPI to produce satisfactory candidate structures. We employed the tri-dimensional Tutte embedding[61] to obtain the Cartesian coordinates for each isomer from the RSPI. Initial force-field refinement of the embedded structures was performed using the Fletcher-Reeves-Polas-Ribiere geometry optimization[62] with analytical gradients[63]. Resulting structures were further optimized by applying the semiempirical quantum mechanics (QM) method PM7[64] with restricted Hartree-Fock methods together with the Broyden-Fletcher-Goldfarb-Shanno optimizer[65-68], as implemented in MOPAC2012[64]. We note that the quantum chemical calculation served only for the purpose of obtaining a favorable lattice geometry for polyhedral cages; in other words, it was assumed that a polyhedral cage with carbon atoms as vertices and described quantum chemically corresponds closely, except for an overall scale factor, to a geometry-optimized virus capsid.

To generate density maps from the Cartesian coordinates of the cages, the dual of each fullerenic structure was calculated. Each vertex of the fullerenic cage and its dual were replaced by a normalized Gaussian distribution of width 1.0 Å. The various Gaussians were then superimposed on a grid. Simulated densities containing both the fullerenic cage and its dual were generated using the VOLMAP routine in VMD[56]. The simulated density maps of fullerenes were scaled up so that the C-C distance matches the distance between the adjacent center of mass at the three-fold axis of cryoEM density map.

Construction of all-atom capsid models

The capsid density at the outer layer of the core was segmented from the 3D tomographic volume and compared with the simulated density of the fullerene models. For this purpose, cross-correlation values between the capsid density contoured according to the CA tube density, and the simulated density as just described, were calculated for 119 fullerene models from the four classes defined above. The models with the best correlation score from two classes containing 216 hexamers or 186 hexamers and 12 pentamers, were used to construct the atomic MD model. The locations of the pentamers for the selected structures, given in terms of IUPAC canonical ring spiral pentagon indices[58-60] are:

1, 8, 21, 62, 64, 90, 136, 165, 199, 204, 215, and 227.

and

1, 7, 14, 51, 55, 79, 116, 145, 176, 180, 191, and 195

The selected geometry optimized polyhedral cage served as a template for building an atomistic model. The MDFF-derived CA hexamer and equilibrated CA pentamer replaced hexagons and pentagons in the cage structure, respectively. For this purpose, the centers of mass at the three-fold symmetry axes were computed from simulations 1 and 6 (Supplementary Table 1), resulting in 6 and 5 centers of mass for each oligomer; the resulting center of mass rings were then aligned to the hexagons and pentagons in the polyhedral cage, respectively. Given the inherent curvature of the MDFF-derived model (Supplementary Table 1, simulation 1), all six possible orientations of the all-atom hexameric model were matched and the best alignment selected. The first solvation shell, including water and ions, both for pentamers and hexamers (from simulations 6 and 1, Supplementary Table 1), were kept during the modeling; conflicting water and ions between neighboring hexamers and pentamers were removed. The resulting two structures, consisting of the all-atom HIV-1 capsids, together with the first hydration shell, included a total of 13,433,232 and 11,640,222 atoms. Additionally, bulk water and ions were added to the structure, setting the ion concentration to 150 mM NaCl. The fully solvated and ionized structure contained a total of 64,332,521 and 64,423,983 atoms.

Equilibrium molecular dynamics simulation of the HIV-1 capsid

MD simulations of the entire HIV-1 capsid were performed using NAMD 2.9 and the CHARMM36 forcefield[69]. Periodic boundary conditions were assumed and simulations were carried at 310K and 1 bar, employing the Langevin piston algorithm. Long-range Coulomb forces were computed by means of the Particle Mesh Ewald method with a grid spacing of 2.1 Å and 8th-order interpolation.

The all-atom models of the HIV-1 capsid were minimized using velocity quenching for 1 ns, employing 2 fs timesteps, evaluating non-bonded interactions every 2 fs and full electrostatics every 6 fs. All hydrogen bonds were constrained using the SHAKE algorithm. While fixing all protein atoms, the system was allowed to run following the same protocol in an NPT ensemble for 2ns; the protein atoms were restrained using an isotropic force constant k = 10 kcal mol $^{-1}$ Å^{-2}. The system was then allowed to equilibrate for 100 ns. For this purpose restraints were slowly removed at a rate of 1 kcal kcal mol $^{-1}$ Å^{-2} per ns. The system was equilibrated eventually without restraints (Supplementary Table 1, simulation 7 & 8) and remained stable.

Computational performance of multi-million atom molecular dynamics simulations

Simulations of the complete HIV-1 capsids (64 million atoms, including solvent) were performed on the supercomputer Blue Waters, funded by the National Science Foundation and the State of Illinois, with the molecular dynamics program NAMD[43] (version 2.9, compiled with memory optimization based on a new compressed molecular structure data format and parallel input/output). Blue Waters is operated by the National Center for Supercomputing Applications at the University of Illinois. The simulations employed 4000 Cray XE6 nodes with 128,000 cores, each node running one process and 30 threads. The simulation was prepared using the analysis program VMD[56], version 1.9.1a5. To reduce inter-node communication for optimal performance, NAMD employed for long-range electrostatic force calculations the PME method (particle-mesh Ewald), dedicating 4000 cores to the task and utilizing a grid spacing of 2.1 Å and 8th-order interpolation to reduce bandwidth requirements. NAMD made use of hybrid load balancing that was renewed every 20,000 steps. Performance on Blue Waters depended on the activity of other users on the machine and whether the set of nodes assigned by the

queuing system was compact or spread out, yielding between 5 ns/day to 9 ns/day of simulation length.

Coarse-grain (CG) model and Monte Carlo (MC) simulation

A CG model was constructed following a setup similar to one described in an earlier publication[70]. As the CA molecules primarily exist as dimers in solution, the dimer of FL-CA of the cryoEM structural model was used to construct the CG model of subunits. A modified Lennard-Jones (LJ) potential $V(r,\theta) = g(r) \cdot f(\theta) - V_{cutoff}$ was employed to describe the pair-wise interactions between subunit binding interfaces of neighboring molecules, with the configuration observed in the cryoEM structural model corresponding to the energy minimum. The actual LJ potential employed is assumed to be $g(r) = \varepsilon_0 [(\frac{\sigma}{r})^{12} - 2(\frac{\sigma}{r})^6]$. Here, σ is the distance between the interaction points of the two helices at the binding interface, defined as the closest points at the center axis of interacting cylinders, its value being determined from our cryoEM structural model; r is the distance between the interaction points; ε_0 is set to 2.0 kcal/mol for each interface, except for the trimeric interface, in which case ε_0 varies between -4.0 to 4.0 kcal/mol. $f(\theta)$ is defined through

$$f(\theta) = \begin{cases} \frac{|\theta - \theta_0|}{\theta_{\max}} - 1, & |\theta - \theta_0| \le \theta_{\max} \\ 0, & |\theta - \theta_0| > \theta_{\max} \end{cases}$$

and captures the orientation-dependence of the interaction between neighboring subunits (θ_0 is determined from angles between two helices at the binding interfaces as seen in the cryoEM model and θ_{max} is a cutoff angle set to $2\pi/3$); V_{cutoff} is the cutoff value for the standard LJ potential at $r = 4\sigma$. To describe the repulsive interactions between the trimeric interface introduced by mutations, g(r) is replaced by: $g(r) = -4\varepsilon_0(\frac{\sigma}{r})$ - V_{cutoff}, where $\varepsilon_0 < 0$, and V_{cutoff} is the value for the Coulomb potential at the cutoff distance $r = 4\sigma$.

To investigate the assembly simulations were performed, with each subunit moving on a 2D plane with full-3D rotational freedom. During each MC step, a small translation within the plane (≤ 1Å) and a random 3D rotation (≤ 5°) were applied to the consecutive subunit. A fast collision detection program was used to detect collisions or overlaps between molecules after these moves[70]. If collisions or overlaps were identified, new in-plane translational and 3D rotational moves were attempted until no collision or overlaps were detected. The energy of the system was then evaluated, and the system evolved under the standard Metropolis Criterion[71]. Constant-temperature MC simulations were performed with a system of 36 CA dimeric subunits in a periodic square lattice at 300 K and a concentration of $[CA]_{2D}$ = 80 dimer/10^4 nm^2, which are the optimal conditions indicated by the phase diagram from 2D simulations carried out in a previous study[70]. All simulations were initialized by running 100,000 MC steps/subunit at 10,000 K before the temperature was dropped to 300 K. The simulation programs were coded in FORTRAN95 and run on the IBM STOKES cluster at the University of Central Florida. The figures of the simulations were generated with MATLAB (The MathWorks, Natick, MA).

High-speed sedimentation assay

Samples (6 µl) were withdrawn from assembly reaction mixtures and mixed with 4X lithium dodecyl sulfate (LDS) loading buffer (Invitrogen) supplemented with 10 mM DTT for SDS-PAGE analysis (t). The remaining samples were pelleted at 20,000 xg with an Eppendorf centrifuge 5417R for 15 min and supernatants (s) and pellets (p, resuspended in 1/3 of volume) were mixed with 4X LDS loading buffer for gel. Total, supernatant, and pellet samples, without boiling, were loaded on 10% SDS-PAGE and stained with Coomassie Blue.

Cells, viruses and cores

The 293T, HeLa, TZM-bl used in this study were cultured in Dulbecco's modification of Eagle's medium (DMEM; Cellgro), supplemented with 10% fetal bovine serum (FBS), penicillin (50 IU/ml), and streptomycin (50 µg/ml), in a humidified 37˚C incubator containing 5% CO_2. Point mutations in the CA region of the R9 HIV-1 proviral plasmid

were introduced by PCR mutagenesis. The mutant viral fragments generated by PCR were reintroduced into R9 infectious HIV-1 clone with unique restriction site pairs: BssHII-ApaI, Spe1-ApaI, and BssHII-SpeI, depending on the location of the residue in the CA region. The regions of the mutant inserts on the viral plasmids were sequenced to confirm the presence of these mutations and absence of extraneous mutations.

Virus stocks were produced by calcium phosphate transient transfection of 293T cells. 20 μg of plasmid DNA was used per 2.0 x 10^6 cells in 10 cm dishes, as previously described[72]. The supernatant was removed from the plate 36-48 hr after transfection, and the monolayer was gently rinsed with sterile PBS before replenishing the media. The culture supernatants were harvested 24 hr later and filtered through a 0.45 μm pore-size filters. Viral aliquots were frozen at -80°C and virus production quantified by p24 ELISA, as previously described, or by RT activity[73]. HIV-1 A14C/E45C cores were purified using the procedure described previously[74].

Single-cycle assay of HIV-1 infectivity

Viral infectivity was assessed by titration on TZM-bl cells: HIV-1 stocks were serially diluted in culture medium containing polybrene (8 μg/ml), and 100 μl volumes were added to TZM-bl target cells seeded the day before (15,000 cells per well in 96-well plates). The cultures were maintained for an additional 48 hrs prior to lysis and detection with a luciferase substrate (Steady-Glo, Promega). To determine the extent of infection, the luminescence reading in relative light units (RLU) was quantified in a Topcount instrument (Perkin Elmer). The corrected RLU values were normalized by RT activity to calculate the specific infectivity of the viruses in RLU/RT.

Crosslinking assay

Viruses were produced by transfection of the mutant plasmids and pelleted by ultracentrifugation (100,000 x*g*, 30 min) through a 20% sucrose cushion. The pellet was dissolved in SDS sample buffer in the presence or absence of reducing agent, β-mercaptoethanol. Samples were heated at 95˚C for 5 min and then analyzed by SDS-PAGE. CA was detected by immunoblotting using a rabbit polyclonal anti-CA antibody.

Transmission electron microscopy analysis of HIV-1 particles

Cultured HeLa cells were transfected with 2.5 μg of viral DNA plasmids using the TransIT-HeLaMonster transfection kit (Mirus), according to the manufacturer's protocol. 24 hr after transfection, cells were washed with phosphate buffered saline (PBS) and fixed with freshly made 2.5% glutaraldehyde in 0.1 M cacodylate buffer (pH 7.4) at room temperature for 2 hr. Following fixation, the cells were sent to the Electron Microscopy BioServices (Frederick, MD) for embedding, sectioning, and microscopic examination.

Biochemical analysis of HIV-1 capsid stability

HIV-1 cores were isolated and characterized as previously described[72]. For each virus, the level of core-associated CA was determined as a percentage of the total CA in the gradient (based on p24 ELISA). Uncoating of purified cores *in vitro* was assayed as previously described[72].

Supplementary References:

29 Yang, R. & Aiken, C. A mutation in alpha helix 3 of CA renders human immunodeficiency virus type 1 cyclosporin A resistant and dependent: rescue by a second-site substitution in a distal region of CA. *J. Virol.* **81**, 3749-3756 (2007).

30 Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82-97 (1999).

31 Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334-347 (2003).

32 Egelman, E. H. The iterative helical real space reconstruction method: surmounting the problems posed by real polymers. *J. Struct. Biol.* **157**, 83-94 (2007).

33 Grigorieff, N. FREALIGN: high-resolution refinement of single particle structures. *J. Struct. Biol.* **157**, 117-125 (2007).

34 Sachse, C. *et al.* High-resolution electron microscopy of helical specimens: a fresh look at tobacco mosaic virus. *J. Mol. Biol.* **371**, 812-835 (2007).

35 Fernandez, J. J., Luque, D., Caston, J. R. & Carrascosa, J. L. Sharpening high resolution information in single particle electron cryomicroscopy. *J. Struct. Biol.* **164**, 170-175 (2008).

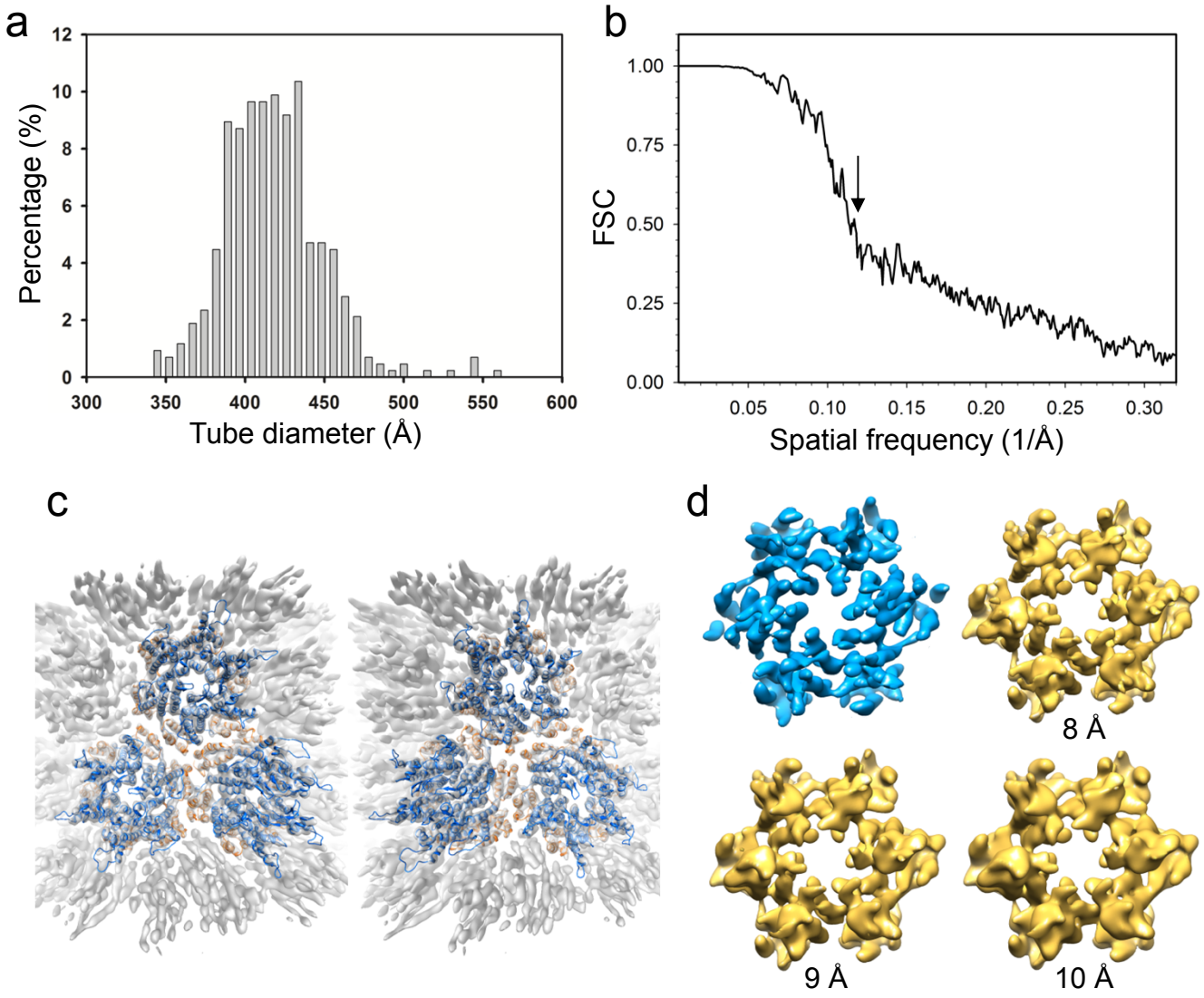
36 Parent, K. N. *et al.* Cryo-reconstructions of P22 polyheads suggest that phage assembly is nucleated by trimeric interactions among coat proteins. *Phys. Biol.* **7**, 045004 (2010).

37 Mastronarde, D. N. Correction for non-perpendicularity of beam and tilt axis in tomographic reconstructions with the IMOD package. *J. Microsc.* **230**, 212-217 (2008).
38 Agulleiro, J. I. & Fernandez, J. J. Fast tomographic reconstruction on multicore computers. *Bioinformatics* **27**, 582-583 (2011).
39 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612 (2004).
40 Lopez-Blanco, J. R., Garzon, J. I. & Chacon, P. iMod: multipurpose normal mode analysis in internal coordinates. *Bioinformatics* **27**, 2843-2850 (2011).
41 Tama, F., Miyashita, O. & Brooks, C. L., 3rd. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. *J. Struct. Biol.* **147**, 315-326 (2004).
42 Trabuco, L. G., Villa, E., Schreiner, E., Harrison, C. B. & Schulten, K. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods* **49**, 174-180 (2009).
43 Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781-1802 (2005).
44 MacKerell, A. D. *et al.* All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **102**, 3586-3616 (1998).
45 Mackerell, A. D., Feig, M. & Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400-1415 (2004).
46 Feller, S. E., Zhang, Y. H., Pastor, R. W. & Brooks, B. R. Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method. *J. Chem. Phys.* **103**, 4613-4621 (1995).
47 Darden, T., York, D. & Pedersen, L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **98**, 10089-10092 (1993).
48 Tuckerman, M., Berne, B. J. & Martyna, G. J. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **97**, 1990-2001 (1992).
49 Pornillos, O. *et al.* X-ray structures of the hexameric building block of the HIV capsid. *Cell* **137**, 1282-1292 (2009).
50 Byeon, I. J. *et al.* Structural convergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. *Cell* **139**, 780-790 (2009).
51 Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779-815 (1993).
52 Wriggers, W. Using Situs for the integration of multi-resolution structures. *Biophys. Rev.* **2**, 21-27 (2010).
53 Chan, K. Y. *et al.* Symmetry-restrained flexible fitting for symmetric EM maps. *Structure* **19**, 1211-1218 (2011).
54 Pornillos, O., Ganser-Pornillos, B. K. & Yeager, M. Atomic-level modelling of the HIV capsid. *Nature* **469**, 424-427 (2011).
55 Stone, J. E. *et al.* Accelerating molecular modeling applications with graphics processors. *J. Comput. Chem.* **28**, 2618-2640 (2007).

56 Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **14**, 33-38 (1996).
57 Arkhipov, A., Freddolino, P. L. & Schulten, K. Stability and dynamics of virus capsids described by coarse-grained modeling. *Structure* **14**, 1767-1777 (2006).
58 Fowler, P. W. & Manolopoulos, D. E. Magic numbers and stable structures for fullerenes, fullerides and fullerenium ions. *Nature* **355**, 428-430 (1992).
59 Fowler, P. W. & Manolopoulos, D. E. *An atlas of fullerenes*. (Clarendon Press, 1995).
60 Godly, E. W. & Taylor, R. Nomenclature and terminology of fullerenes: A preliminary survey. *Fullerenes, Nanotubes, and Carbon Nanostructures* **5**, 1667-1708 (1997).
61 Tutte, W. T. How to draw a graph. *Proc. London Math. Soc* **13**, 743-768 (1963).
62 Flannery, B. P., Press, W. H., Teukolsky, S. A. & Vetterling, W. Numerical recipes in C. *Press Syndicate of the University of Cambridge, New York* (1992).
63 Wu, Z., Jelski, D. A. & George, T. F. Vibrational motions of buckminsterfullerene. *Chem. Phys. Lett.* **137**, 291-294 (1987).
64 Stewart, J. J. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* **19**, 1-32 (2013).
65 Broyden, C. G. The convergence of a class of double-rank minimization algorithms 2. The new algorithm. *IMA J. Appl. Math.* **6**, 222-231 (1970).
66 Fletcher, R. A new approach to variable metric algorithms. *Comput. J.* **13**, 317-322 (1970).
67 Goldfarb, D. A family of variable metric methods derived by variational means. *Math. Comput.* **24**, 23-26 (1970).
68 Shanno, D. F. Conditioning of quasi-Newton methods for function minimization. *Math. Comput.* **24**, 647-656 (1970).
69 Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J. Chem. Theory Comput.* **8**, 3257-3273 (2012).
70 Chen, B. & Tycko, R. Simulated self-assembly of the HIV-1 capsid: protein shape and native contacts are sufficient for two-dimensional lattice formation. *Biophys. J.* **100**, 3035-3044 (2011).
71 Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **21**, 1087-1092 (1953).
72 Shah, V. B. & Aiken, C. In vitro uncoating of HIV-1 cores. *J. Vis. Exp.* **57, e3384, DOI: 10.3791/3384** (2011).
73 Aiken, C. Pseudotyping human immunodeficiency virus type 1 (HIV-1) by the glycoprotein of vesicular stomatitis virus targets HIV-1 entry to an endocytic pathway and suppresses both the requirement for Nef and the sensitivity to cyclosporin A. *J. Virol.* **71**, 5871-5877 (1997).
74 Zhao, G. *et al.* Rhesus TRIM5alpha disrupts the HIV-1 capsid at the inter-hexamer interfaces. *PLoS Path.* **7**, e1002009 (2011).

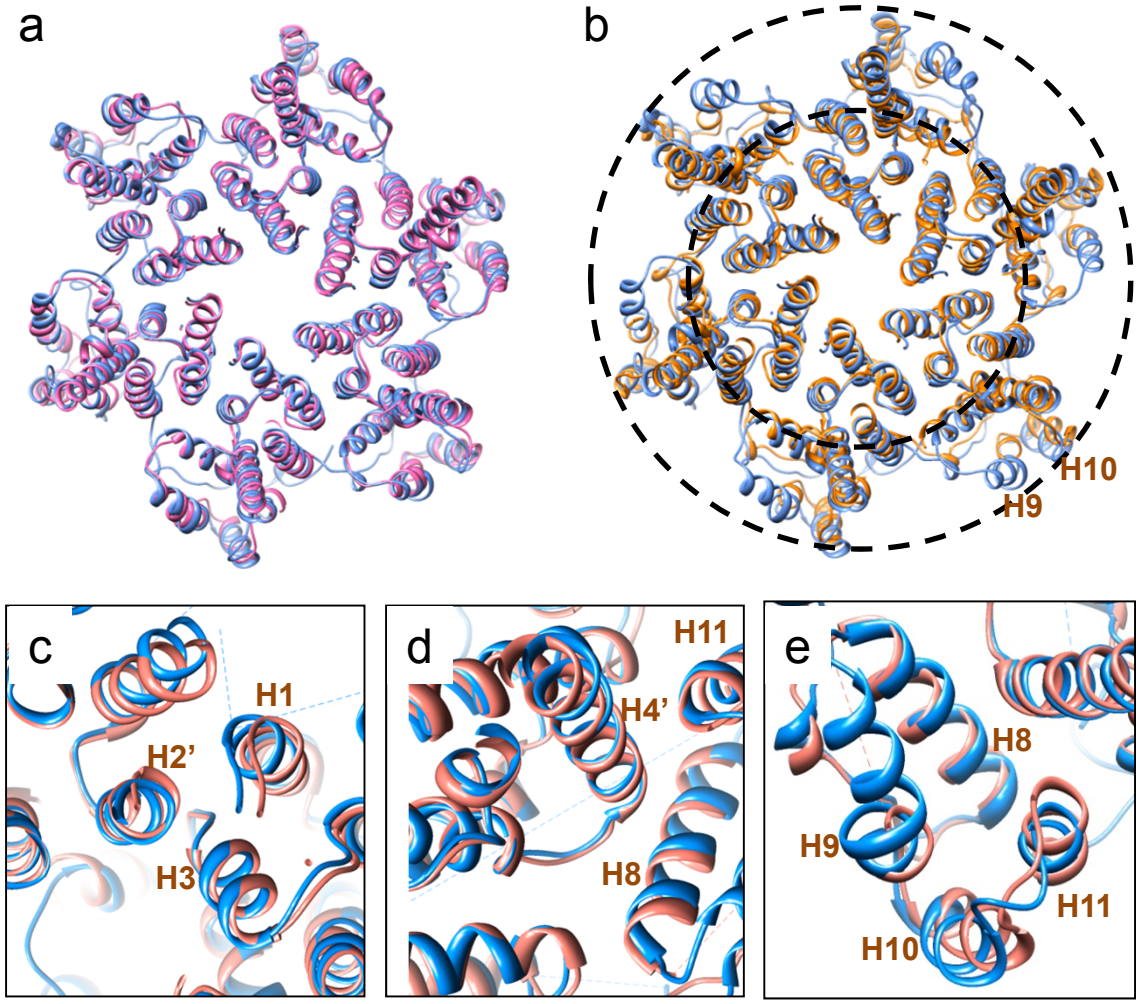
Table S1. Atomistic MD simulations performed.

	Simulation	System Size (Number of atoms)	Length of the simulation (ns)
1	Hexamer-of-Hexamers (MDFF)	1,319,202	10
2	Hexamer-of-Hexamers	1,319,202	425
3	Entire hexameric tubular assembly	12,894,735	125
4	Pentamer (3P05)	130,229	10
5	Pentamer-of-Hexamers (POH-1)	1,046,161	300
6	Pentamer-of-Hexamers (POH-2)	1,111,025	1200
7	216 Hexamers-12 Pentamers	64,332,531	100
8	186 Hexamers-12 Pentamers	64,423,983	100

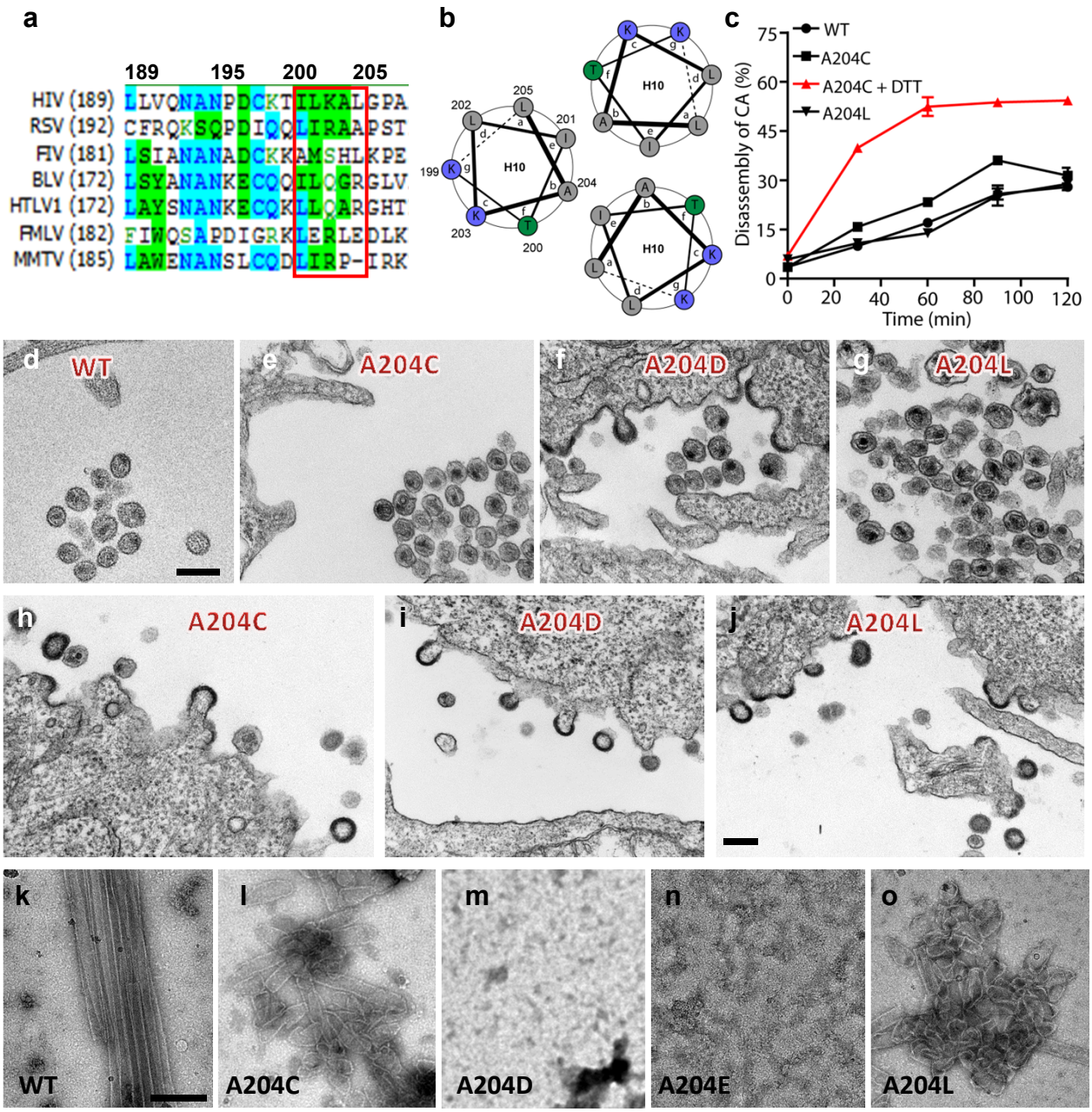


Supplementary Figure 1 | Image processing and 3D reconstruction of HIV-1 CA hexameric tubular assembly.

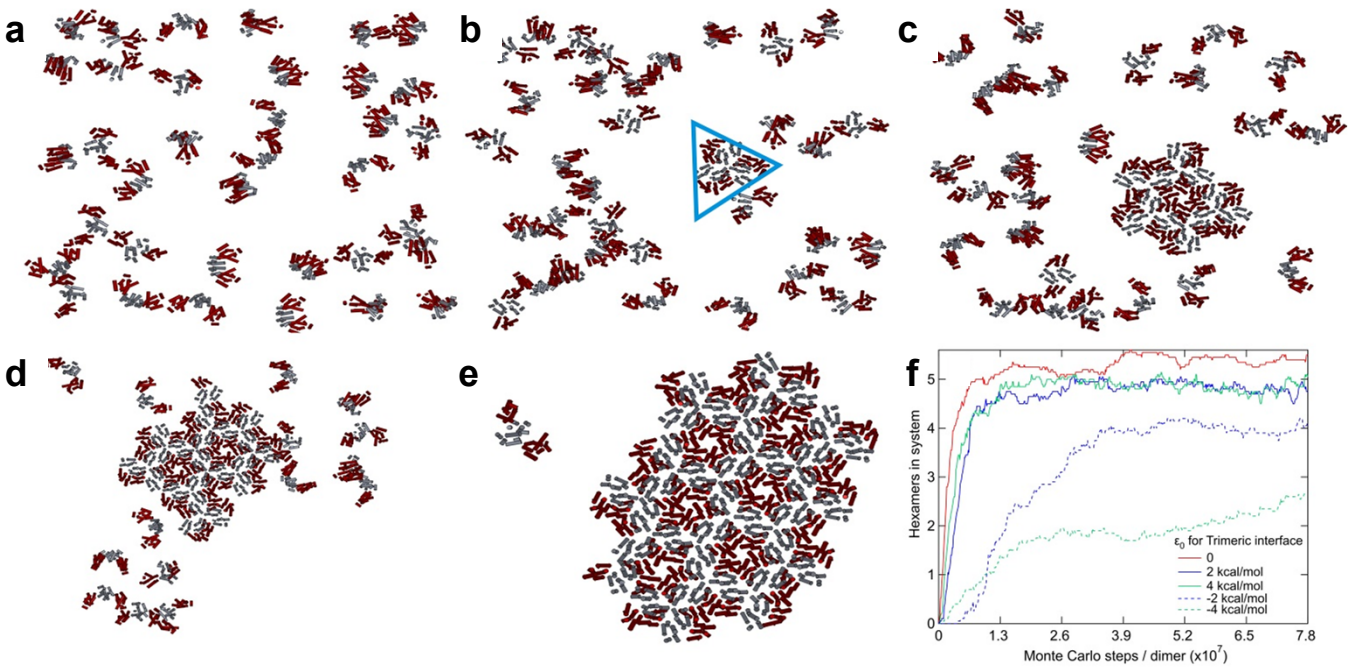
a, Histogram of tube diameter distribution (n = 425). b, Fourier shell correlation (FSC) of the CA density map, calculated from odd and even data sets. The data sets were independently aligned and reconstructed using a 20 Å resolution density map as the initial reference. The resolution of the map is 8.6 Å at FSC=0.5 (arrow). c, Stereo view of the density map, with the MDFF model of three adjacent hexamers superimposed. d, Comparison of the reconstruction (blue hexamer) with the MDFF model (gold) filtered to 8, 9, or 10 Å resolution.



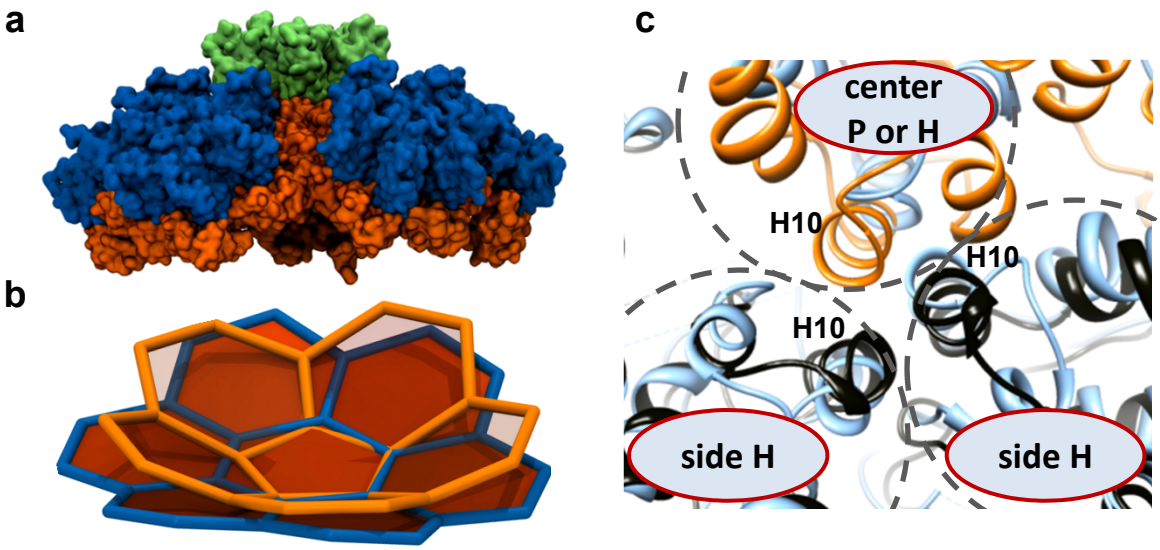
Supplementary Figure 2 | Comparison of an MDFF-derived hexamer with the X-ray crystal structure.
a, Superposition of an MDFF hexamer (blue) with a normal mode-based flexible fitting model (pink). The overall RMSD for Cα is 1.3 Å. b, Superposition of the MDFF hexamer (blue) with the crystal structure (PDB code: 3H47, orange). The average RMSD between the two models is 2.3Å, with an RMSD of 1.9 Å for the NTD (within the inner circle) and of 2.8 Å for the CTD (between the two circles). CTD helices H9 and H10, located at the periphery of the hexamer, are labeled. c-e, Comparison of NTD-NTD (c) and NTD-CTD interfaces (d) and of the CTD helices H8-H11 (e) from MDFF model (blue) and crystal structure (orange).



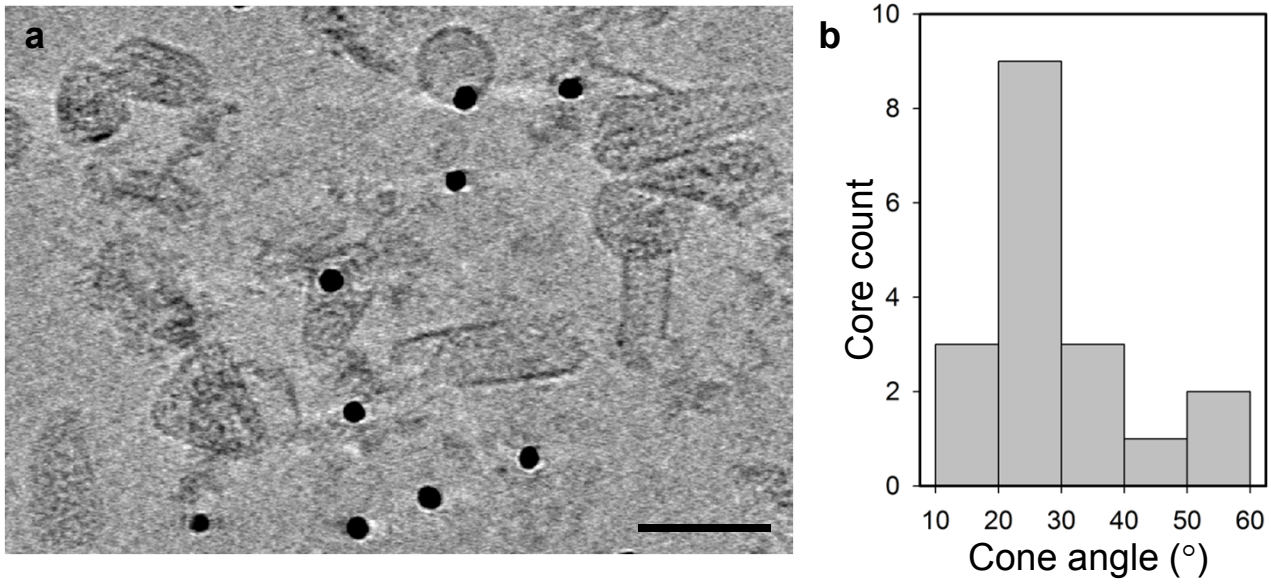
Supplementary Figure 3 | Mutational study of the CTD trimer interface.
a, Sequence alignment of the H10 region among retroviruses. Identical positions and conserved positions are highlighted in blue and green, respectively. Weakly conserved positions are shown with a green letter. Residues at the CTD trimer interface (201-205) are boxed in red. b, Helical wheel presentation of the three H10 helix bundle at the CTD trimer interface. The orientations of the H10 helices are derived from the MDFF model. Hydrophobic, basic, and polar non-charged residues are shown in gray, blue, and green, respectively. c, *In vitro* uncoating kinetics of purified WT and A204 mutant HIV-1 cores. Addition of DTT to WT cores has no significant effect on the uncoating kinetics. d-j, Electron micrographs of fixed, plastic-embedded, and thin sectioned samples containing released mature virus particles (d-g) and budding immature particles (h-j). WT (d), A204C (e) and A204L (g) mature virions display cone-shaped cores and A204D (f) particles contain aberrant cores. k-o, Negatively stained EM micrographs of WT (h), A204C (i), A204D (j), A204E (k) and A204L (l) mutant CA assemblies. Scale bars, 200 nm.



Supplementary Figure 4 | Coarse grain Monte Carlo simulation of CA assembly.
a, Initial configuration of a system consisting of 36 dimeric CA subunits at a density $[CA]_{2D}$ = 0.016 nm^{-2} and at T = 300K. ε_0 = 2 kcal/mol for all subunit binding interfaces, except for the trimeric interface. CA NTDs and CTDs are depicted in red and gray, respectively. b-e, Snapshots of assembly at the following evolution points: 6.8 (b), 7.3 (c), 8.6 (d) and 13.4 × 10^7 (e) steps/dimer (corresponding to 6.8, 7.3, 8.6, and 13.4 ms [70]). A trimer of three CA dimers is formed (b, highlighted by the blue triangle) and quickly grows into a full hexameric lattice (e). f, Rate of hexamer formation under different trimer interface binding strengths, from ε_0 = -4 (repulsive) to 4 (attractive) kcal/mol. Each trace is an average of 20 simulations to remove the noise associated with MC simulations.

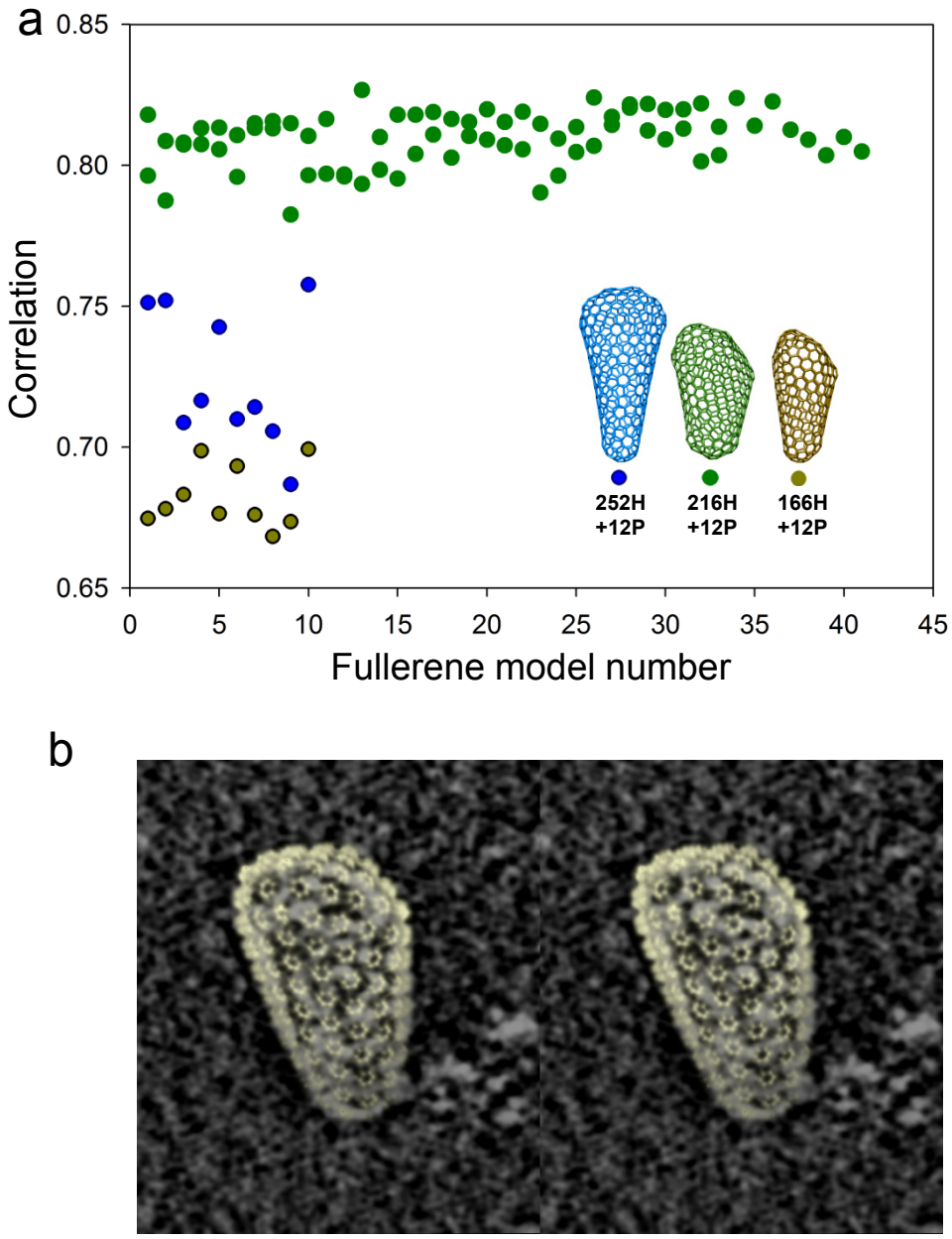


Supplementary Figure 5 | All-atom MD simulation of POH CA assembly
a, Surface rendering of the POH, viewed along the tube axis (NTD in green for the pentamer and blue for the hexamers with CTDs all in orange). b, Comparison of curvatures between POH (orange) and HOH (blue). c, Comparison of CTD trimer interfaces arising in HOH (light blue) and POH (orange and black). The three CTDs at the trimer interface are within the dashed grey circles.

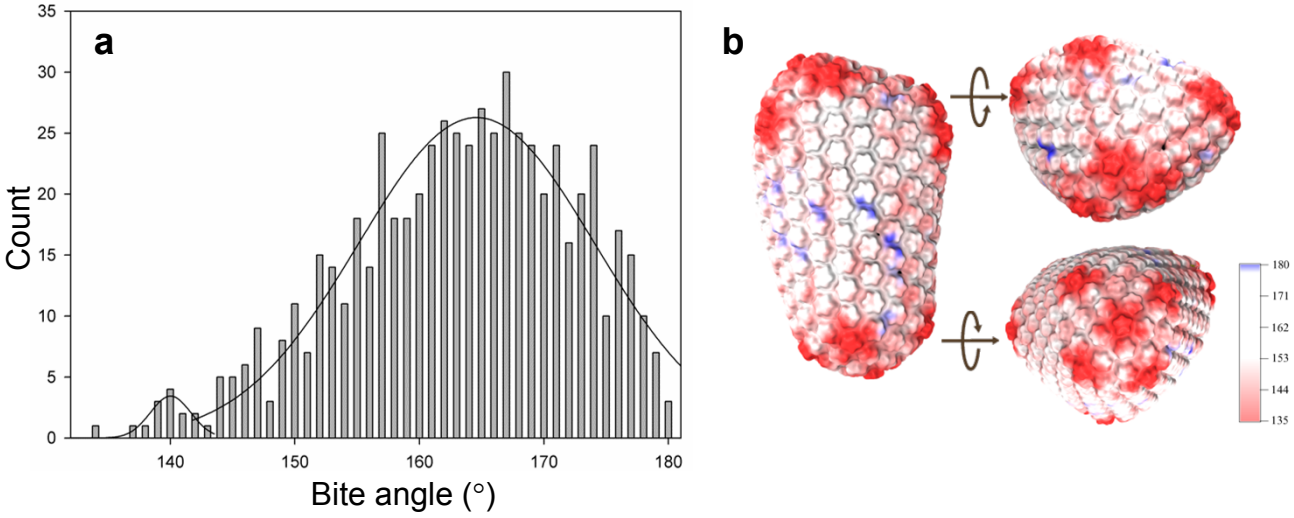


Supplementary Figure 6 | Cryo-ET analysis of native HIV-1 cores.

a, Projection view of purified HIV-1 cores from which 3D tomographic data were collected. Scale bar, 100 nm. b, Distribution of the cone angles measured from tomography reconstructions of HIV-1 cores.



Supplementary Figure 7 | Fitting the fullerene models to the HIV-1 capsid structure. a, Cross-correlation between HIV-1 core density and three classes of fullerene models: 252H+12P (blue), 216H+12P (green), and 166H+12P (brown). Each model includes 10, 74, or 10 isomers, respectively. b, Stereo view of the simulated density (light yellow) of the all-atom HIV-1 capsid model comprising 186H+12P (pdb code 3J3Y) overlaid with a planar slice of the HIV-1 core tomogram. At the intersection between model and tomogram one can recognize for the bottom right periphery agreement between the two densities.



Supplementary Figure 8 | Analyses of the HIV-1 capsid model.
a, Histogram of bite angle distribution. The bite angle distribution was fitted to two Gaussian curves centered at 138° and 164.8°, respectively. b, Orthogonal views of an equilibrated mature HIV-1 capsid model. The surface was colored according to bite angles with red corresponding to 135° and blue corresponding to 180°.