# Strain and crack propagation of HIV-1 capsids during uncoating

**Authors:** Alvin Yu,[1] Elizabeth M.Y. Lee,[2] John A.G. Briggs,[3] Barbie K. Ganser-Pornillos,[4] Owen Pornillos,[4] and Gregory A. Voth[1,*]

**Affiliations:**

[1]Department of Chemistry, Chicago Center for Theoretical Chemistry, Institute for Biophysical Dynamics, and James Franck Institute, The University of Chicago; Chicago, USA.

[2]Pritzker School of Molecular Engineering, The University of Chicago; Chicago, USA

[3]Department of Cell and Virus Structure, Max-Planck-Institute of Biochemistry, Martinsried, Germany

[4]Department of Molecular Physiology and Biological Physics, University of Virginia; Charlottesville, USA

**Corresponding author**

*Gregory A. Voth
Department of Chemistry
The University of Chicago
5735 S. Ellis Ave, SCL 123
Chicago IL, 60637

Phone: (773) 702-9092
Fax: (773) 795-9106
**Email**: gavoth@uchicago.edu

**Classification: Primary** Biological Sciences – Biophysics and Computational Biology
**Secondary** Physical Sciences – Chemistry

**Keywords:** HIV capsid, all-atom molecular dynamics, strain, uncoating

**Author Contributions:** A.Y., E.M.Y.L., J.A.G.B., B.K.G-P., O.P., and G.A.V. designed research. A.Y. performed research. A.Y., B.K.G-P., and O.P. contributed data, methods, simulation code and analytic tools. A.Y., E.M.Y.L., J.A.G.B., B.K.G-P., O.P., and G.A.V. analyzed data. A.Y., E.M.Y.L., J.A.G.B., B.K.G-P., O.P., and G.A.V. wrote the paper.

**This PDF includes:**
Main Text
Figures 1 to 4

**Abstract**

Viral replication in HIV-1 relies on a fullerene-shaped capsid to transport genetic material deep into the nucleus of an infected cell. Capsid stability is linked to the presence of cofactors, including inositol hexakisphosphate ($IP_6$) that bind to pores found in the capsid. Using extensive all-atom molecular dynamics simulations of HIV-1 cores imaged from cryo-electron tomography (cryo-ET) in intact virions, which contain $IP_6$ and a ribonucleoprotein complex, we find markedly striated patterns of strain on capsid lattices. The presence of these cofactors also increases rigidity of the capsid. Conformational analysis of capsid (CA) proteins show CA accommodates strain by locally flexing away from structures resolved using x-ray crystallography and cryo-electron microscopy. Then, cryo-ET of HIV-1 cores undergoing endogenous reverse transcription demonstrate that lattice strain increases in the capsid prior to mechanical failure and that the capsid ruptures by crack propagation along regions of high strain. These results uncover HIV-1 capsid properties involved in their critical disassembly process.

**Significance statement**

The mature capsids of HIV-1 are transiently stable complexes that self-assemble around the viral genome during maturation, and uncoat to release preintegration complexes that archive a double-stranded DNA copy of the virus in the host cell genome. However, a detailed view of how HIV cores rupture remains lacking. Here, we elucidate the physical properties involved in capsid rupture using a combination of large-scale all-atom molecular dynamics simulations and cryo-electron tomography. We find that intrinsic strain on the capsid forms highly correlated patterns along the capsid surface, along which cracks propagate. Capsid rigidity also increases with high strain. Our findings provide fundamental insight into viral capsid uncoating.

## Main Text

### Introduction

The mature capsids of HIV-1 are large fullerene-like protein complexes that are comprised of more than a thousand copies of the capsid protein (CA) (1). During replication, viral particles that bud from host cells are initially immature and composed primarily of a spherical array of Gag proteins that link essential viral proteins and enzymes into a linear polypeptide chain (2). As the virus matures extracellularly, proteolytic cleavage of Gag, releases the capsid protein, which self-assembles in a cone-shaped geometry and packages two copies of the retroviral genome and associated enzymes into the capsid. Mature capsids are deposited during the fusion of HIV particles with the plasma membrane, in which the lipid envelope and embedded proteins are lost, leaving behind the core of the virus in the cytoplasm of cells.

Capsids play essential roles during replication, by transporting viral genetic material deep into the host cell (3). Pores in the capsid can bind or import small molecules, including inositol phosphates ($IP_6$) and nucleotides (dNTPs) (4, 5), and the binding of $IP_6$ increases the stable lifetimes of the capsid and promotes the assembly of CA into fullerene structures (6, 7). Cryo-electron tomography (cryo-ET) and other techniques have recently demonstrated that viral cores are imported with apparently intact capsids across the nuclear pore of infected cells (8–11). Reverse transcription processes inside the capsid can rupture the core as seen in both atomic-force microscopy and cryo-ET experiments (12, 13) owing to an increased internal pressure on the capsid during the conversion of RNA to DNA. Uncoating of the capsid is a critical replication event, releasing enzymes and nucleic acids that integrate a copy of the virus in the host genome. Yet, little is known about the physical properties underlying capsid rupture.

To investigate the structural and mechanical properties of HIV-1 capsids that lead to rupture and disassembly, we have performed large-scale all-atom molecular dynamics (AAMD) simulations of HIV-1 capsids, containing native cofactors including $IP_6$, and a ribonucleoprotein complex (RNP). Analysis of the strain induced on the capsid, reveals spatially correlated patterns, indicating that the capsid collectively cracks open along regions of high strain rather than slowly disassembles. Local fluctuations in the capsid volume decrease concomitantly with increased strain, consistent with a mechanical rigidification of the capsid in response to $IP_6$ and the RNP. Calculated free energy landscapes also reveal shifts in the conformational ensembles of CA in response to increased strain. Cryo-ET imaging of *in vitro* reconstituted HIV-1 cores incubated with nucleotides then add further insight into the temporal sequence of events during disassembly to demonstrate that strain is maximal prior to the formation of cracks in the capsid. These results show that capsids are intrinsically strained and help to characterize the molecular processes by which viral cores rupture.

**Results**

Our AAMD simulations of HIV-1 cores contained a total simulation size ranging from 44 to 76 million atoms. In prior AA MD simulations (14, 15), an atomic level model for the HIV-1 capsid was constructed from low-resolution cryo-ET maps that did not distinguish individual CA hexamers or pentamers and contained positions for these structures estimated from local capsid curvature. The latter model was also an empty capsid shell, enclosing only water with neither nucleic acid contents nor the binding of inositol phosphates (i.e., $IP_6$). In contrast, we derived six fullerene lattices for the capsid from 6.8 Å cryo–ET densities imaged in intact HIV-1 virions, in which positions and orientations of lattice components were resolved (Fig. 1A) (16, 17). $IP_6$ polyanions were added to each CA hexamer or pentamer pore at the binding site, a location 2.7 Å above the R18 ring (5,

7). In the absence of atomic resolution into the structure of the ribonucleoprotein complex (RNP), we constructed a model of two copies of the 9-kilobase pair RNA genome in complex with nucleocapsid proteins consistent with experimental secondary structure constraints from selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) analysis (18), in order to mimic the HIV-1 RNP. Core particles were simulated at successively increasing levels of detail ranging from capsid shells containing only water to more realistic capsids containing the RNP model and $IP_6$ molecules (Fig. 1 B to E) (see Materials and Methods for a more complete description). In aggregate, the AA MD simulations totaled 1.6 μs across the HIV-1 core particles (table S1).

**Capsid Strain and Rigidity**

During the AAMD simulations, the viral capsids remained intact. Empty capsids in bulk solvent without $IP_6$ did not dissociate nor reassemble into more stable helical or spherical arrangements within the timescales simulated. Local deformations in materials and proteins have elucidated the mechanical properties of materials under stress and conformational changes resulting from protein–ligand interactions (19). To quantify these features, we computed per-particle strain tensors ($\boldsymbol{\varepsilon}$), for the center-of-masses of five amino acid residue segments in the CA domain. The volumetric strain ($s_{\mathrm{V}} = \frac{1}{3}\mathrm{Tr}(\boldsymbol{\varepsilon})$) measures the propensity for a particular region of the capsid to either swell or shrink (Fig. 2A). Empty cores containing liquid water had relatively little strain (CA; $\langle|s_{\mathrm{V}}|\rangle = 2.4 \times 10^{-2}$), which was distributed randomly across the capsid. Strain increased In the presence of the RNP complex (CA-RNP; $\langle|s_{\mathrm{V}}|\rangle = 3.3 \times 10^{-2}$). $IP_6$ binding induced even larger effects that were also more spatially correlated than the capsid containing just the RNP (CA-$IP_6$; $\langle|s_{\mathrm{V}}|\rangle = 3.6 \times 10^{-2}$). Cores containing an RNP complex and $IP_6$ molecules were the most strained with patterns that formed unexpected striations (see Fig. 2A) along the capsid surface (CA-RNP-$IP_6$; $\langle|s_{\mathrm{V}}|\rangle = 5.1 \times 10^{-2}$). Different capsid

structures showed a similar trend of increased strain in the presence of $IP_6$ and RNP, but with slightly altered strain patterns that are attributable to differences between core morphologies (Fig. S1).

To assess whether capsid properties change in response to $IP_6$ or RNP, we monitored the internal volumes of each core during the simulations. Although each pleomorphic capsid differed in size (Fig. 1 A), average core volumes ranged from $1.15–1.53 \times 10^5$ $nm^3$, with fluctuations of several tens of cubic nanometers about the equilibrium. Mean-free volume fluctuations for the largest core are shown in Fig. 2B. Fluctuation amplitudes decreased markedly in the presence of either $IP_6$ or the RNP, consistent with higher core rigidity (Fig. 2C, CA-$IP_6$ vs. CA-RNP) (std. dev.: $\sigma_{CA}$= 41.1 $nm^3$; $\sigma_{CA\text{-}RNP}$= 36.3 $nm^3$; $\sigma_{CA\text{-}IP6}$= 31.2 $nm^3$; $\sigma_{CA\text{-}RNP\text{-}IP6}$= 29.2 $nm^3$). Fourier analysis showed a shift in the peak-to-peak frequencies of the dominant mode towards lower frequencies with the RNP, indicating that the capsids fluctuate more slowly (CA: $\upsilon = 57.8$ MHz; CA-RNP: $\upsilon = 39.4$ MHz) (Fig. S3). As the core rigidifies further upon $IP_6$ binding, the dominant low frequency mode broadens, and the fluctuations are distributed to higher frequency modes, consistent with a stiffer capsid. Negatively charged $IP_6$ molecules bind tightly to an arginine ring in pores distributed throughout the capsid (4, 5, 7). RNP interactions with CA, on the other hand, could occur through interactions of positively charged residues on the flexible CTD tail of CA with the RNP (7). The cofactor interactions at the pore and CTD tail change the conformational flexibility of CA resulting in an increase in capsid rigidity that introduces strain on the underlying lattice.

**Conformational Analysis of CA Proteins**

To examine the conformations CA proteins adopt in the actual capsid, we computed a 3D free energy landscape (Fig. 3A to B) from the spatial distribution occupied by the non-hydrogen atoms of CA across the pleomorphic lattice for the capsids containing $IP_6$

and RNP. At high contours ($\Delta G = 0.5\ \mathrm{kcal/mol}$), density for the protein backbone of CA is clearly visible, whereas at low contours ($\Delta G = 4.4\ \mathrm{kcal/mol}$), variability in the ensemble of CA structures undergoing dynamical motion results in larger volumes (Fig. 3B). Less structured regions of the protein, including the CTD tail, β hairpin, and CypA binding loop, had higher variability, and were more mobile. The hinge connecting the CA NTD and CTD was more ordered than the other unstructured regions, possibly owing to NTD–NTD and CTD–CTD contacts in the adjacent CA domains of the capsid. The conformations of individual CA monomers in pentamers and hexamers resolved by cryo-ET (16) show small differences in the relative orientations of the NTD and CTD (Fig 3C). Relative free energy differences between the conformations were quite small ($\Delta G < 0.2\ \mathrm{kcal/mol}$), indicating that CA can switch between the two states in the capsid under thermal motion.

Interfacial contacts between the NTD of one CA domain and the CTD of the adjacent CA domain differed appreciably during the simulations in the pentamer and hexamer (Fig. 3D to G). We employed an angle parameter, $\phi$, defined as the angle between amino acid backbone center-of-masses in the NTD helices and CTD helices, using the base of the NTD helices as a pivot, to assess the relative orientation of the NTD and adjacent CTD (Fig. 3E, G, and Supplementary Materials). The angle, $\phi$, was larger in the cryo-ET structure (16) of the pentamer ($\phi = 157°$) than in the hexamer ($\phi = 113°$) reflecting a greater curvature of pentamers over hexamers. Pore sizes were monitored using a parameter, $\xi$, defined as the distance between the center-of-masses of A22 and N21 on adjacent CA monomers (Fig. 3D, F). Pentamers in flexible, empty capsids had high curvatures and small pore sizes $(\xi_{\mathrm{CA}}, \phi_{\mathrm{CA}}) = (7\text{–}11\ \text{Å}, 115\text{–}167°)$, although a subpopulation ($f = 2.5\text{–}3.5$) shifted to a lower curvature closer to that of the x-ray crystal structure of a flat pentamer (20) with engineered disulfide crosslinks between A22 and N21 (PDB ID: 3P05) $(\xi, \phi) = (7.5\ \text{Å}, 107°)$ (Fig. 3H). $IP_6$ binding led to

a conformational shift in pentamer distributions towards larger pore sizes ($\xi_{\mathrm{RNP-IP6}} = 7-12.5$ Å), and capsids with both $IP_6$ binding and the RNP increased pentamer curvature ($\phi_{\mathrm{RNP}} = 119-168°$, $\phi_{\mathrm{IP6}} = 121-170°$, $\phi_{\mathrm{RNP-IP6}} = 128-171°$).

CA hexamers on the other hand occupied wider pores, but flatter curvatures (lower $\phi$), $(\xi_{\mathrm{CA}}, \phi_{\mathrm{CA}}) = (6.5-15.6\ \text{Å}, 92-136°)$. $IP_6$ binding and the RNP did not significantly affect pore size in hexamers, although there was a greater variation in hexamer curvature ($\phi_{\mathrm{IP6}} = 90-147°$, $\phi_{\mathrm{RNP-IP6}} = 91-150°$), owing to the strain induced on the overall capsid lattice (Fig 3I). X-ray crystal and cryo-EM structures agree with the minima of both curvature and pore size distributions, and resided within the lowest two contours of the $(\xi, \phi)$ distributions ($f < 1.5$), but CA conformations deviate away from these minima under increasing lattice strain. Notably, pentamers exhibited larger relative conformational shifts than hexamers (Fig. 3H, I), suggesting that pentamers may be more susceptible to destabilization during rupture.

**Core Rupture During Endogenous Reverse Transcription**

Cryo-ET imaging and lattice mapping were used to probe how core structure changed during reverse transcription (13). In brief, HIV-1 cores were released from purified virions by permeabilizing the viral membranes with the pore-forming melittin peptide and then stabilized by the addition of $IP_6$ at native cellular concentrations. Reverse transcription was initiated by adding dNTPs at concentrations found in $CD4^+$ T cell cytoplasms (21). After four hours of incubation at 37 °C, capsid structures showed a range of intermediates, ranging from intact cones to partially cracked capsids to nearly completely disassembled (Figure 4A, images 1-7). We examined lattice separations in each of the cores, using a local order parameter, $\chi$,, which measures the near-neighbor contacts of a particle (see Supplementary Material). In our AAMD simulations, lattice separations correlated with the strain, as expansive strain caused slight, but measurable deviations

in lattice separation ($\chi < 1.0$), whereas compressive strain induced more closely packed lattices ($\chi > 1.0$) (Fig. 4B). No neighboring particles are present at low values of $\chi$, which provided a quantitative metric for the degree of cracking across the core ($\chi < 0.55$). Interestingly, tubular "pill" shaped capsids (Fig. 4A image 1) were nearly perfect fullerenes with no cracks present, whereas other more conical capsids had cracks that were not visually obvious prior to computing $\chi$ in initial inspections (Fig. 4C). Intermediate values for $\chi$ correspond to small deviations and separations that reflect the expansive strain on the capsid ($0.75 < \chi < 0.9$). Analysis of $\chi$ at these intermediate values revealed a bimodal distribution, with an initial spike in the early stages of rupture (Fig. 4D 2,3). Mechanical failure of the capsid and loss of capsid integrity lowers the strain, which then increases as the pre-integration complex exits the core (Fig. 4D 6,7). This suggests reverse-transcribing capsids are strained prior to and during rupture of the capsid.

## Discussion

We simulated entire HIV-1 core particles derived from cryo-ET imaging, at an atomic level of detail and computed the intrinsic strain on the capsid in the presence of native cofactors, including $IP_6$ and RNP, which increased the mechanical rigidity of the capsid. In biochemical assays (13), overstabilization of the capsid at high $IP_6$ concentrations results in cores that do not produce reverse transcription products, and the binding of small-molecule inhibitors such as GS-CA1 (22) and its close analog GS-6207 (23) accelerates capsid fracturing (13), suggesting HIV-1 cores have mechanical properties that can be altered to disrupt retroviral life cycle processes. Our conformational analysis of CA proteins in the capsid also indicates that pentamers and hexamers have a surprising degree of dynamic flexibility that changes under strain. Cryo-ET images of cores undergoing reverse transcription show that lattice strain is locally maximal prior to

mechanical failure, and that capsids rupture in a fashion consistent with the propagation of cracks along highly strained regions, as observed in the AAMD simulations.

During trafficking from the cytoplasm to the nucleus of infected cells, viral capsids are exposed to interactions with a variety of host proteins. Cyclophilin A binds to and enhance stability of the capsid in a concentration dependent fashion (24) in the cytoplasm. Nucleoporins bind to and import viral cores across the nuclear pore of host cells (25, 26). The cleavage and polyadenylation specific factor 6 (CPSF6) interacts with capsid complexes (27) and is involved in nuclear import and trafficking to integration sites (28, 29). Could these interactions with host proteins along with increasing internal pressure from reverse transcription weaken capsid integrity and rupture the capsid? Additional investigations should therefore further probe the molecular mechanisms involved in viral core interactions within the host cell.

## Materials and Methods

### All-atoms models of the capsid core

Initial atomic models for the CA hexamer and pentamer were constructed from the cryo-ET structure of the CA hexamer (PDB ID: 5MCX) and CA pentamer (PDB ID: 5MCY) derived from intact virus particles. Amino acid side chain conformations were modeled based on the x-ray crystal structure of the CA hexamer (PDB ID: 4XFX) and disulfide-crosslinked CA pentamer (PDB ID: 3P05). Missing protein backbones residues were built using MODELLER (30), and missing side chains were built using SCWRL4 (31). C-terminal domain (CTD) tails for CA were transplanted on to the model using the NMR structure for the tail (residues: 220–231) (PDB ID: 2KOD) onto the model. Atomic models for the CA hexamer and pentamer were then coarse-grained (CG) at a resolution of one amino acid residue per CG particle. Each CG hexamer and pentamer subunit was constrained as a rigid body, and positioned at the Cartesian coordinates and Euler angles that maximized electron density overlap between the atomic model and the cryo-ET fullerene capsid structure derived from intact virions (16).

CG models of the capsids were briefly relaxed in a 20 ps Langevin dynamics run under the canonical (constant NVT) ensemble with the large-scale atomic/molecular massively parallel simulator (LAMMPS) (32). To maintain overall protein shape, each CG particle included excluded volume interactions using a soft cosine potential $U_{\mathrm{Vol}} = A\,[1+\cos\,(\pi\, r/r_c)]$, when $r < r_c$. The cutoff ($r_c$) is the onset of excluded volume repulsion between CG particles, and was set to the separation distances found in the CTD interfaces if the distances were less than 10 Å, or otherwise 10 Å. The parameter *A* was taken to be $A = 100\ \mathrm{kcal\ mol^{-1}}$. The mass of the protein was evenly distributed among the CG beads. CTD contacts were maintained using an elastic network model (ENM) that connected the two groups of CG particles in neighboring CTD dimers (residues: 178–194) in a cutoff distance, $r_{\mathrm{cut}}$, with harmonic bonds; the potential energy

of each bond is: $U_{\mathrm{bond}} = K_{\mathrm{bond}}\,(r - r_0)^2$, where $r$ is the separation distance and $r_0$ is the equilibrium bond length, set to the distance found in the crystal structure ($r_{\mathrm{cut}} = 10$ Å, $K_{\mathrm{bond}} = 0.1\ \mathrm{kcal\ mol^{-1}\ Å^{-2}}$). Temperature was mainted at 300K with a Langevin thermostat with a damping constant ($t_{\mathrm{damp}} = 5$ ps). All-atom (AA) models for each CA hexamer and pentamer were then aligned with each CG capsomere subunit to construct an initial AA model for the capsid. Six complete capsids were constructed corresponding to the fullerene geometries derived from the cryo-ET structures, two of which were selected for additional modeling.

For each of the capsid systems selected, $IP_6$ molecules were placed at positions corresponding to the bound x-ray crystal structure, 3.5 Å above an R18 ring that lines the pore of the CA hexamer and pentamer. Approximately 200 $IP_6$ molecules were used for each system, corresponding to a single $IP_6$ molecule for each capsomere subunit in the fullerene structure. A model of the ribonucleoprotein core was generated using a simplified CG model of RNA with a resolution of four beads per nucleotide base pair. CG particles for the atom types in each base pair (adenine: C3' C8 N6 C2; cytosine: C3' C6 O2 N4; guanosine: C3' C8 O6 N2; uracil: C3' C6 O2 O4) were templated to generate the viral RNA. Bond lengths were maintained using harmonic restraints between neighboring CG particles ($K_{\mathrm{bond}} = 1.0\ \mathrm{kcal\ mol^{-1}\ Å^{-2}}$). Secondary structure constraints were implemented using harmonic restraints between the nearest-neighbor CG particles forming base-pairing interactions on the basis of high-throughput selective 2-hydroxyl acylation analyzed by primer extension (SHAPE) reactivity data (18) ($K_{\mathrm{SS}} = 0.1\ \mathrm{kcal\ mol^{-1}\ Å^{-2}}$). Two copies of the 9 kilobase pair genome were modeled and positioned in the interior of the capsid. CG particles for nucelocapsid proteins (NC) were added randomly bound to RNA, until there was a 1:1 correspondence between the number of CA and NC domains. All-atom models for the RNA and nucleocapsid proteins

(PDB ID: 1A1T) were fit to the CG models, after a brief energy minimization and 100 ps Langevin dynamics to generate a model of the RNP complex.

**MD simulations of viral capsid cores**

Solvated capsids were of sizes ranging from 44–76 million atoms including water molecules and ions. $Na^+$ and $Cl^-$ ions were added to the bulk solution until the salt concentration was 150 mM NaCl to produce an electrostatically neutral system. Periodic boundary conditions were imposed on an orthorhombic unit cell ranging from $55.2\ \text{nm} \times 58.9\ \text{nm} \times 128.5\ \text{nm} - 72.8\ \text{nm} \times 76.3\ \text{nm} \times 141.9\ \text{nm}$, and contained a solvent buffer of 10 nm in the (x, y, z) dimensions away from nonsolvent atoms. The AA potential energy function CHARMM36m (33, 34) for proteins and the TIP3P (35) potential energy function for water were used. The AA systems were energy minimized and equilibrated under constant pressure and temperature (NPT) conditions. Simulations in the constant NPT ensemble were performed using a Langevin thermostat at 310 K and a Nosé-Hoover Langevin barostat at 1 atm. Bond-lengths for hydrogen atoms were constrained using the SHAKE algorithm (36). An r-RESPA integrator was used with a timestep of 2 fs; long-range electrostatics were computed every 4 fs (37). Long-range electrostatics were calculated using the particle mesh Ewald algorithm, and short-ranged, nonbonded interactions were truncated at 12 Å (38). $IP_6$ molecules were parameterized using the CHARMM General Force Field (CGenFF) (39). All simulations used the AAMD simulation package NAMD 2.14 (40). Production-level runs were performed on six capsid containing liquid water and two capsids containing $IP_6$, the RNP core, or both $IP_6$ and RNP. For all subsequent analysis AAMD trajectories were sampled at 0.04-ns intervals.

**Strain calculations**

The center-of-masses for every five amino acid residues in the CA domains were used to subsample the capsid structure. For each center-of-mass, a deformation tensor that describes the local deformation of a point particle in a system relative to it's neighboring particles was calculated: $\boldsymbol{F}_{i,t} = \left(\sum_{j\in N_i^0} |r_{ji}^0\rangle\langle r_{ji}^0|\right)^{-1} \sum_{j\in N_i^0} |r_{ji}^0\rangle \langle r_{ji}^t|$, where $|r_{ji}^0\rangle = \vec{x}_j^0 - \vec{x}_i^0$ is the difference between the Cartesian coordinates of neighboring particles $j$ and $i$ with respect to the reference configuration, similarly $|r_{ji}^t\rangle$ is the difference in coordinates at time $t$, and $N_i^0$ is the set of particles in the local neighborhood of $i$ in the reference configuration, within a cutoff distance of 10 nm. Note that the deformation tensor, $\boldsymbol{F}_{i,t}$, can be determined by minimizing $\sum_{j\in N_i^0} \left|\langle r_{ji}^0|\boldsymbol{F}_{i,t} - \langle r_{ji}^t|\right|^2$.The reference configuration was the fitted atomic model of the fullerene capsid after CG relaxation. The per-particle Green-Lagrange strain tensor is then, $\boldsymbol{\varepsilon}_i = \left(\boldsymbol{F}_i\boldsymbol{F}_i^T - \boldsymbol{I}\right)/2$, with volumetric and shear (von Mises) strain invariants given by: $s_{\mathrm{V}} = \mathrm{Tr}(\boldsymbol{\varepsilon})/3$ and $s_{\mathrm{s}} = [\mathrm{Tr}(\boldsymbol{\varepsilon}^2) - \mathrm{Tr}(\boldsymbol{\varepsilon})^2/3]^{1/2}$ respectively. Strain tensors were computed for each frame of the trajectories, until the invariants converged to an equilibrium value (e.g., $\sum_i s_{V,i}^{t+1} - s_{V,i}^t < 0.01$) at frame $t \sim 32$ ns. Average strain magnitudes, $\langle|s_{\mathrm{V}}|\rangle$, were measured using the strain corresponding to residue 110.

**Core volume analysis**

Core volumes of the capsid were computed using a Monte-Carlo sampling strategy. Cartesian coordinates for the Cα atoms of the CA domains were monitored and used to subsample the capsid structure. For each iteration, a point within the periodic simulation cell were randomly generated ($\vec{x}_{\mathrm{MC}}$), and assigned to the capsid interior, if $\|\vec{x}_{\mathrm{MC}}\| < \|\vec{x}_{\mathrm{C}\alpha}\|$, where $\vec{x}_{\mathrm{C}\alpha}$ is the position of the Cα atom that maximizes the projection of the point onto the capsid structure ($\vec{x}_{\mathrm{MC}}/\|\vec{x}_{\mathrm{MC}}\| \cdot \vec{x}_{\mathrm{C}\alpha}/\|\vec{x}_{\mathrm{C}\alpha}\|$). $10^9$ iterations were performed, and core volumes were calculated from the total fraction of points inside the capsid,

accumulated for each 0.04 ns interval frame. Fluctuations from the mean volume ($V - \langle V \rangle$) were measured in each capsid trajectory. Normalized fluctuation probabilities were obtained by binning the amplitudes with a 15 nm step-size, and subsequently fit to Gaussian distributions. Fourier analysis was performed by computing the discrete Fourier transform on the mean-free volume fluctuations (41).

**CA Conformational distributions**

AAMD trajectories from production-level simulations of CA systems were subsampled at 0.04-ns intervals and aggregated. CA monomers in the capsid were aligned by minimizing the RMSD of the protein backbone atoms with respect to the crystallographic structure (PDB ID: 4XFX). Cartesian coordinates for the nonhydrogen atoms of CA were recorded, and the density of atomic positions $\rho(\vec{r})$ was calculated, using a hard-sphere van der Waals approximation on a discretized grid with a spacing of $0.5\,\text{Å} \times 0.5\,\text{Å} \times 0.5\,\text{Å}$ as in (42). The 3D density map for the CA monomer was subsequently contoured at $G = 0.5$, 2.3, and 4.4 kcal/mol, where $G = -k_\text{B}\,T \log \rho$. The parameter, $\xi$, was used to specify the distance between the center-of-masses of the amino acid backbones of residue 21 and residue 22 on the pore helices (H1) of adjacent CA domains. The angle, $\phi$ is defined as the angle between three center-of-masses in the amino acid backbone, two of which were positioned on the CA NTD (residues 50–52, 78–80, and 128–130; residues 63–65, 144–146, and 58–60) and the third positioned at the adjacent CA CTD (residues 188–192, 166–170, 196–200). $(\xi, \phi)$ values were recorded for each 0.04-ns frame and aggregated across each of the simulated capsid systems (CA, CA-RNP, CA-$IP_6$, and CA-RNA-$IP_6$). Successive contours for the $(\xi, \phi)$ distributions are given in units of $-\log p$, where $p$ is the probability.

**Cryo-ET imaging of fractured cores**

Endogenous reverse transcription (ERT) starting with purified HIV-1 virions was performed as described (13). Samples of cores undergoing ERT for 4 h at 37 °C were mixed with equal volumes of 10-nm BSA Gold Tracer (Electron Microscopy Sciences), and 3.5 uL aliquots were applied onto glow-discharged Quantifoil grids (Electron Microscopy Sciences), blotted to near-dryness, and then plunge-frozen into liquid ethane. Cryotomograms were acquired using an FEI Titan Krios electron microscope operating at 300 kV and equipped with either a Falcon III camera or a K3/GIF with a slit width of 20 eV. Tilt series were collected using the data collection software Tomography 4.0 (FEI) with an angular range of −60° to +60°, an angular increment of either 1 or 2°, defocus values of 5 to 10 μm, and a nominal magnification of ×29,000 (Falcon III) or ×33,000 (K3), which correspond to a pixel (px) size of 2.92 (Falcon III) or 2.69 (K3) Å. Tilt series were aligned by using IMOD (43). Weighted back-projection or SIRT was used to reconstruct tomograms in IMOD. Lattice mapping was performed as described (13).

**Local order parameter ($\chi$) analysis**

Atomic models for CA hexamers were aligned with the Cartesian coordinates and Euler angles for each subtomogram in the cryo-ET density. For each structure, Cα positions were used to compute a per-particle, order parameter, $\chi$, which quantifies the separation in the capsid lattice. $\chi$ is defined as the number of neighboring particles (within a cutoff distance of 10 nm) of the same chemical identity in a CA monomer (e.g., residue number) $N_{\mathrm{P}}$, normalized by the average $N_{\mathrm{P}}$ for a given residue. $\chi$ values < 1 correspond to lattice separations that were higher than average for the capsid structure, whereas $\chi$ values > 1 correspond to lattice separations that were lower than average. The capsid with the highest degree of strain (CA-RNP) from the MD simulations was used to measure correlations between $\chi$ and $s_{\mathrm{V}}$. The values of $\chi$ and $s_{\mathrm{V}}$ were measured for the protein residue with the largest variation in lattice separation (residue 110). The total

number of CA monomers that contain a particle with $\chi < 0.55$ or $n_{\mathrm{CA}}$, and the total number of particles with $0.75 < \chi < 0.9$ or $n_{\mathrm{CG}}$ were measured for each capsid structure.

**Acknowledgements**

This research was supported by the National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health under grant R01 AI154092 (G.A.V.), grant P50 AI150464 for the Center for the Structural Biology of Cellular Host Elements in Egress, Trafficking, and Assembly of HIV (B.K.G.-P., O.P.), and the Medical Research Council as part of UK Research and Innovation (MC_UP_1201/16 to J.A.G.B.). B.K.G.-P. and O.P. were also supported by NIH-NIAID grants R01 AI129678 and R01 AI150479. Computational resources were provided by Frontera at the Texas Advanced Computer Center funded by the National Science Foundation (OAC-1818253). A.Y. gratefully acknowledges support from the NIH-NIAID under grant F32 AI150208. The authors thank Simone Mattei for preparing and sharing cryo-ET data on HIV-1 capsids.

## References


1. B. K. Ganser, S. Li, V. Y. Klishko, J. T. Finch, W. I. Sundquist, Assembly and Analysis of Conical Models for the HIV-1 Core. *Science* **283**, 80–83 (1999).

2. F. K. M. Schur, *et al.*, Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution. *Nature* **517**, 505–508 (2015).

3. W. I. Sundquist, H.-G. Kräusslich, HIV-1 Assembly, Budding, and Maturation. *Cold Spring Harb. Perspect. Med.* **2**, a006924 (2012).

4. D. A. Jacques, *et al.*, HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis. *Nature* **536**, 349–353 (2016).

5. R. A. Dick, *et al.*, Inositol phosphates are assembly co-factors for HIV-1. *Nature* **560**, 509–512 (2018).

6. D. L. Mallery, *et al.*, IP6 is an HIV pocket factor that prevents capsid collapse and promotes DNA synthesis. *eLife* **7**, e35335 (2018).

7. A. Yu, E. M. Y. Lee, J. Jin, G. A. Voth, Atomic-scale characterization of mature HIV-1 capsid stabilization by inositol hexakisphosphate (IP6). *Sci. Adv.* **6**, eabc6465 (2020).

8. V. Zila, *et al.*, Cone-shaped HIV-1 capsids are transported through intact nuclear pores. *Cell* **184**, 1032-1046.e18 (2021).

9. C. 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. *Proc. Natl. Acad. Sci.* **118**, e2019467118 (2021).

10. R. C. Burdick, *et al.*, HIV-1 uncoats in the nucleus near sites of integration. *Proc. Natl. Acad. Sci.* **117**, 5486–5493 (2020).

11. T. G. Müller, *et al.*, HIV-1 uncoating by release of viral cDNA from capsid-like structures in the nucleus of infected cells. *eLife* **10**, e64776 (2021).

12. S. Rankovic, J. Varadarajan, R. Ramalho, C. Aiken, I. Rousso, Reverse Transcription Mechanically Initiates HIV-1 Capsid Disassembly. *J. Virol.* **91**, e00289-17.

13. D. E. Christensen, B. K. Ganser-Pornillos, J. S. Johnson, O. Pornillos, W. I. Sundquist, Reconstitution and visualization of HIV-1 capsid-dependent replication and integration in vitro. *Science* **370**, eabc8420 (2020).

14. G. Zhao, *et al.*, Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **497**, 643–646 (2013).

15. J. R. Perilla, K. Schulten, Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations. *Nat. Commun.* **8**, 15959 (2017).

16. S. Mattei, B. Glass, W. J. H. Hagen, H.-G. Kräusslich, J. A. G. Briggs, The structure and flexibility of conical HIV-1 capsids determined within intact virions. *Science* **354**, 1434–1437 (2016).

17. A. Yu, *et al.*, TRIM5α self-assembly and compartmentalization of the HIV-1 viral capsid. *Nat. Commun.* **11**, 1307 (2020).

18. J. M. Watts, *et al.*, Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711–716 (2009).

19. M. R. Mitchell, T. Tlusty, S. Leibler, Strain analysis of protein structures and low dimensionality of mechanical allosteric couplings. *Proc. Natl. Acad. Sci.* **113**, E5847–E5855 (2016).

20. O. Pornillos, B. K. Ganser-Pornillos, M. Yeager, Atomic-level modelling of the HIV capsid. *Nature* **469**, 424–427 (2011).

21. E. M. Kennedy, *et al.*, Ribonucleoside Triphosphates as Substrate of Human Immunodeficiency Virus Type 1 Reverse Transcriptase in Human Macrophages *. *J. Biol. Chem.* **285**, 39380–39391 (2010).

22. S. R. Yant, *et al.*, A highly potent long-acting small-molecule HIV-1 capsid inhibitor with efficacy in a humanized mouse model. *Nat. Med.* **25**, 1377–1384 (2019).

23. J. O. Link, *et al.*, Clinical targeting of HIV capsid protein with a long-acting small molecule. *Nature* **584**, 614–618 (2020).

24. C. Liu, *et al.*, Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site. *Nat. Commun.* **7**, 10714 (2016).

25. K. A. Matreyek, S. S. Yücel, 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).

26. M. Kane, *et al.*, Nuclear pore heterogeneity influences HIV-1 infection and the antiviral activity of MX2. *eLife* **7**, e35738 (2018).

27. A. Bhattacharya, *et al.*, Structural basis of HIV-1 capsid recognition by PF74 and CPSF6. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 18625–18630 (2014).

28. G. A. Sowd, *et al.*, A critical role for alternative polyadenylation factor CPSF6 in targeting HIV-1 integration to transcriptionally active chromatin. *Proc. Natl. Acad. Sci.* **113**, E1054–E1063 (2016).

29. A. C. Francis, *et al.*, HIV-1 replication complexes accumulate in nuclear speckles and integrate into speckle-associated genomic domains. *Nat. Commun.* **11**, 3505 (2020).

30. N. Eswar, *et al.*, Comparative Protein Structure Modeling Using Modeller. *Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis Al* **0 5**, Unit-5.6 (2006).

31. G. G. Krivov, M. V. Shapovalov, R. L. Dunbrack, Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **77**, 778–795 (2009).

32. S. Plimpton, Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).

33. A. D. MacKerell, *et al.*, All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).

34. J. Huang, *et al.*, CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2017).

35. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).

36. H. C. Andersen, Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* **52**, 24–34 (1983).

37. M. Tuckerman, B. J. Berne, G. J. Martyna, Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **97**, 1990–2001 (1992).

38. U. Essmann, *et al.*, A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).

39. K. Vanommeslaeghe, *et al.*, CHARMM General Force Field (CGenFF): A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690 (2010).

40. J. C. Phillips, *et al.*, Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).

41. E. M. Y. Lee, A. J. Mork, A. P. Willard, W. A. Tisdale, Including surface ligand effects in continuum elastic models of nanocrystal vibrations. *J. Chem. Phys.* **147**, 044711 (2017).

42. A. Yu, H. Salazar, A. J. R. Plested, A. Y. Lau, Neurotransmitter Funneling Optimizes Glutamate Receptor Kinetics. *Neuron* **97**, 139-149.e4 (2018).

43. J. R. Kremer, D. N. Mastronarde, J. R. McIntosh, Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).

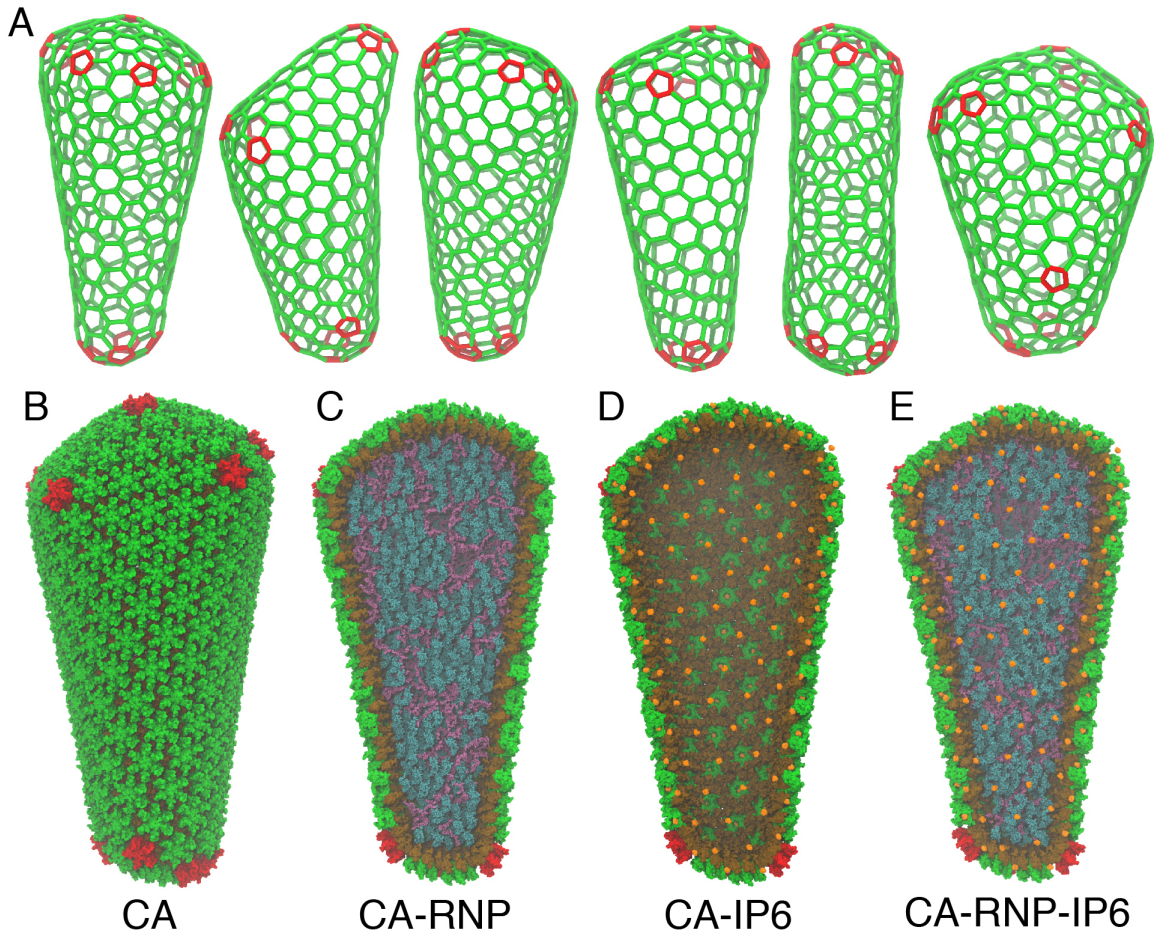


**Figure 1. Mature HIV-1 capsids are pleomorphic. (A)** Fullerene geometries for the HIV-1 capsid were derived from cryo-ET images of intact virions (16). Atomic models for the capsid contain either **(B)** liquid water in the capsid interior, **(C)** a ribonucleoprotein (RNP) complex model, **(D)** inositol hexakisphosphate ($IP_6$) molecules bound to the capsid pores, or **(E)** both the RNP and $IP_6$. The CA amino terminal domain (NTD) and carboxyl terminal domain (CTD) are colored in green and brown, whereas genomic RNA, nucleocapsid proteins, and $IP_6$ molecules are in purple, teal, and orange, respectively. Pentamer defects are colored in red.

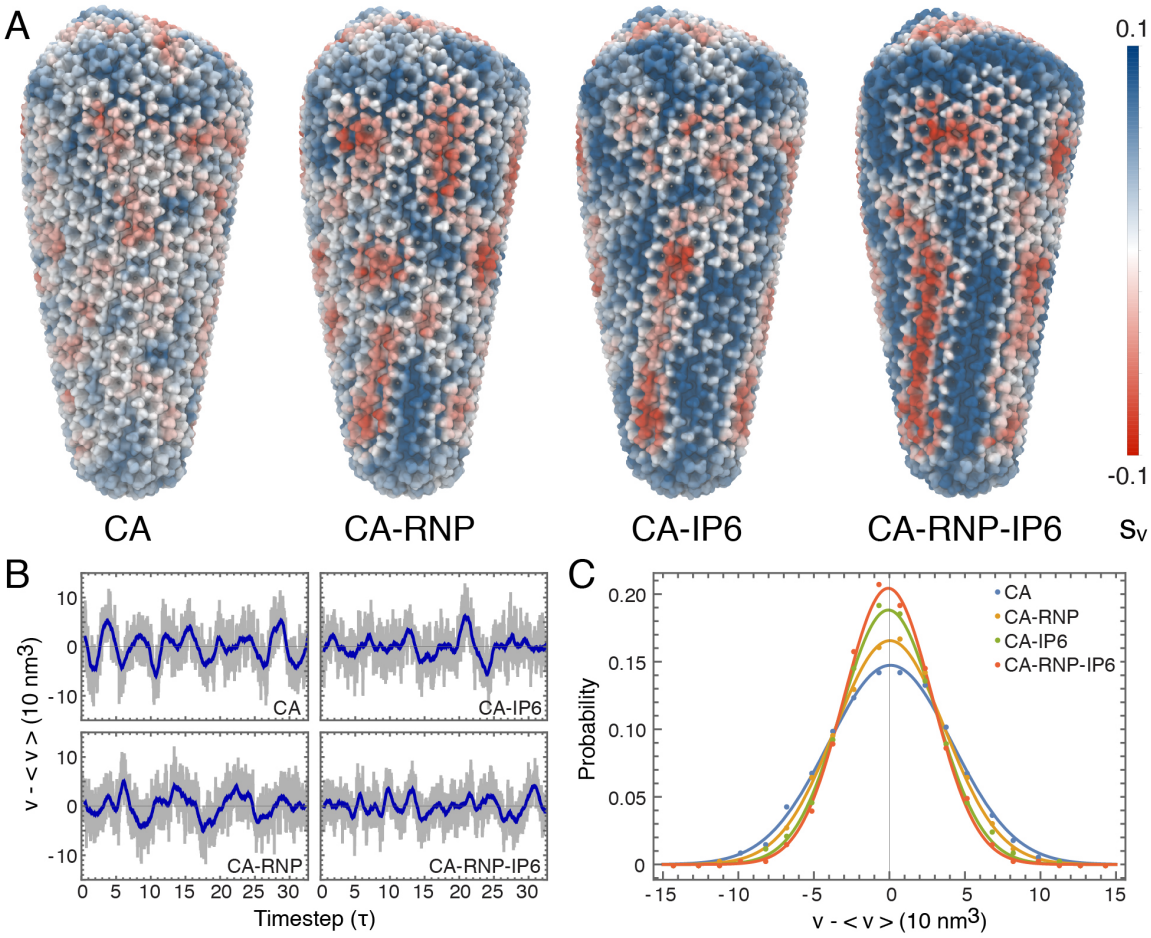


**Figure 2. Strain and fluctuation analysis of the HIV-1 capsid. (A)** The first invariant of the strain tensor ($s_V$) or volumetric strain was computed for each capsid core complex. Red and blue colors correspond to compressive and expansive strain, respectively. **(B)** Fluctuations in the HIV-1 core volumes, measured as deviations from the average capsid volume ($\langle V \rangle = 1.45 \times 10^5$ nm$^3$). The subsampled timestep, τ, is 4 ns. Gray colors denote the instantaneous volume, whereas the blue line denotes a moving average within a one-timestep window. **(C)** Probability distributions for the volume fluctuation amplitudes in each capsid. Closed circles correspond to measured data, while the solid line indicates a Gaussian fit.

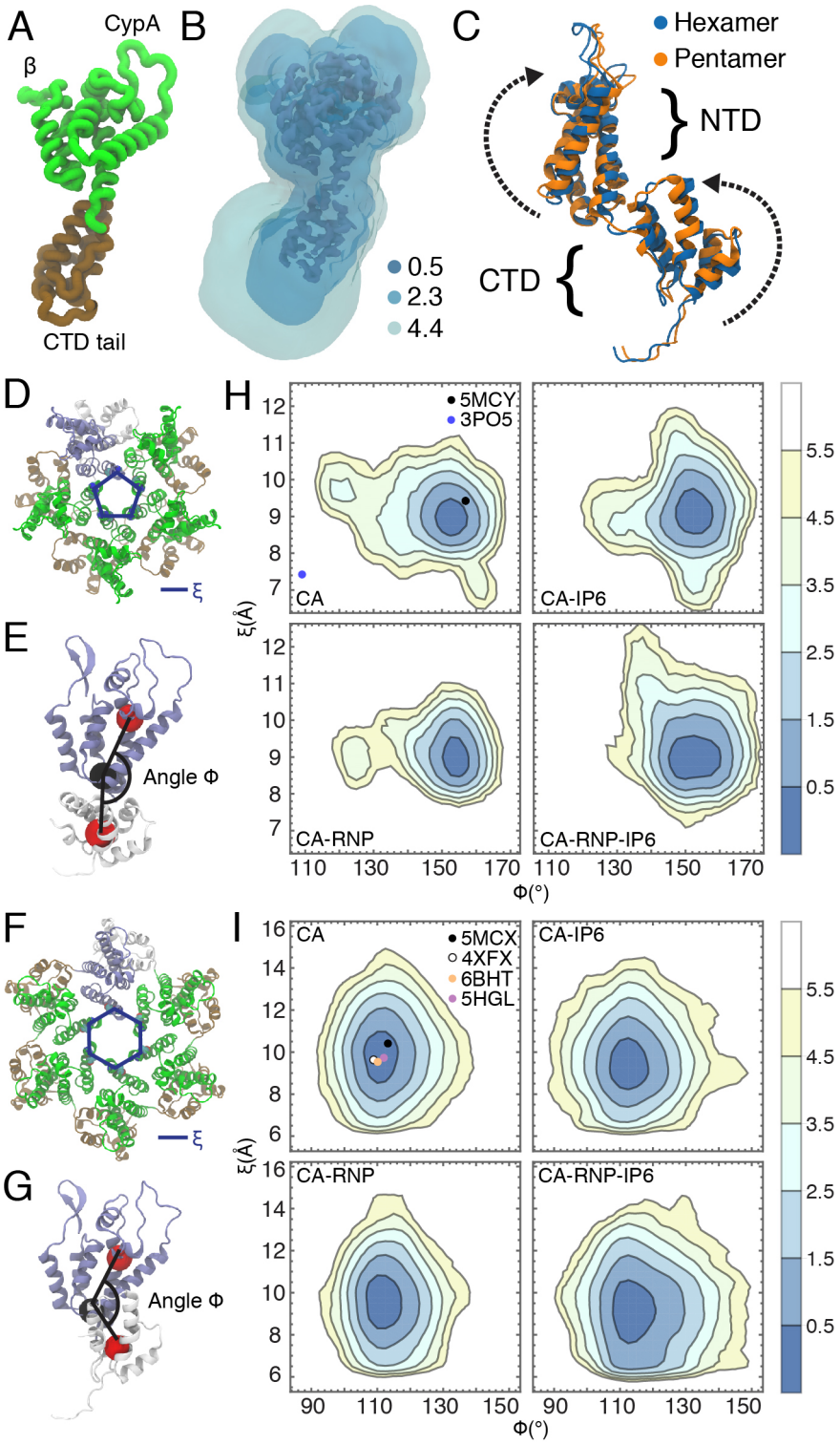


**Figure 3. Conformational analysis of CA domain proteins. (A)** A monomer of the CA protein. **(B)** The 3D potential of mean force (PMF) for CA conformations contoured at 0.5, 2.3 and 4.4 kcal/mol with respect to the Cartesian coordinates of the protein heavy atoms. **(C)** Differences in CA monomer conformations between the cryo-EM structure of the hexamer (PDB ID: 5MCX) and cryo-EM structure of the pentamer (PDB ID: 5MCY) consist of rotations in the NTD and CTD. The parameter, ξ, is used to describe the pore size and is defined as the center-of-mass distance between residues L20 and N21. The

angle Φ is used to describe the relative orientation of a NTD and adjacent CTD and is defined as the angle between three center-of-masses across the NTD and CTD (red and black spheres). ξ and Φ are shown for the CA hexamer **(D, E**) and pentamer **(F,G)**. **(H)** (ξ, Φ) distributions for the hexamers in each capsid. **(I)** (ξ, Φ) distributions for the pentamers in each capsid. Contour lines correspond to increments of $f = -\log p$, where $p$ is the probability. Closed circles correspond to (ξ, Φ) values from experimental structures in the Protein Data Bank.

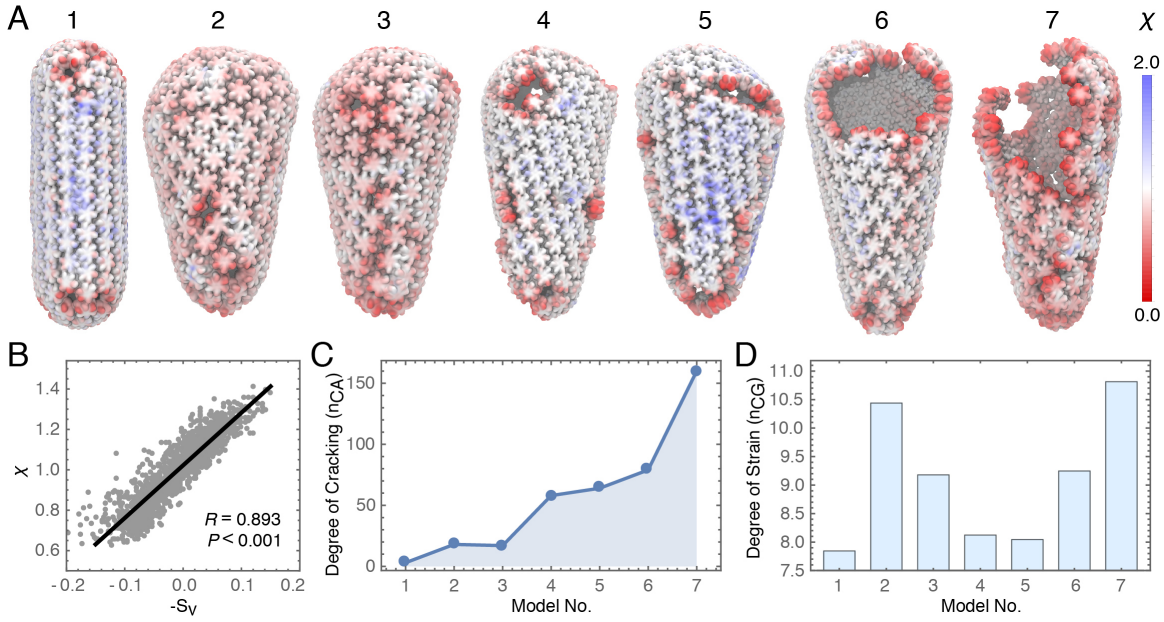


**Figure 4. Cryo-ET of HIV-1 cores during rupture. (A)** Cores were imaged during endogenous reverse transcription. Each structure (images 1-7) is colored by a local order parameter ($\chi$) that quantifies the degree of separation in the lattice. An ideal lattice has a separation of $\chi = 1.0$. Red colors correspond to particles further apart on average from near-neighbors, whereas blue colors correspond to particles closer to near-neighbors. **(B)** Correlation between the strain ($s_{\mathrm{V}}$) and lattice separation ($\chi$) measured in the MD simulations. **(C)** The degree of cracking in each capsid model measured by the number of CA monomers containing a particle with $\chi < 0.55$. **(D)** The degree of strain in each capsid model measured by the number of particles ($10^3$) with $0.75 < \chi < 0.9$. Capsids 1–7 in panel A were ordered from no fractures to completely ruptured.