PNAS

Supporting Information for

Mechanistic Insights into Lenacapavir-Induced Off-Pathway HIV-1 Capsid Assembly

Manish Gupta,[1] Curt Waltmann,[1] Nadine Renner,[2] Yihang Wang,[1] Leo C James,[2] David A Jacques,[3] Till Böcking,[3] and Gregory A. Voth,[1,*]

[1]Department of Chemistry, Chicago Center for Theoretical Chemistry, Institute for Biophysical Dynamics, and James Franck Institute, The University of Chicago, Chicago, IL 60637
[2]MRC Laboratory of Molecular Biology, Cambridge, United Kingdom
[3]EMBL Australia Node in Single Molecule Science, School of Biomedical Sciences, UNSW, Sydney, Australia

*Gregory A. Voth
E-mail: gavoth@uchicago.edu

This PDF file includes:

Supporting text
Table S1
Figs. S1 to S7
Legends for Movies S1 to S2
SI References

Other supporting materials for this manuscript include the following:

Movies S1 to S2

Supporting Information Text

Coarse-grained (CG) model details

In this study, we used a previously developed solvent-free CG model of HIV-1 CA and IP6 (1, 2). Here, we briefly describe the key features relevant to our simulations. HIV-1 CA monomer was represented by two independent elastic network models (ENM) for the NTD and CTD domains, collectively comprising 130 Cα atoms corresponding to α–helices of the CA subunit. The CG representation of the NTD was constructed using Cα positions from PDB structure 3H4E (3). The potential energy of the bonds (E_{bond}) was given by equation 1. To preserve the local secondary structure of the protein, CG particles within each α–helix were connected using harmonic springs with $K = 10\ \mathrm{kcal/mol/Å^2}$. CA CTD was modeled from PDB 2KOD using a similar approach (4). Additionally, to preserve the relative orientation of the α–helices, each CG particle was connected to its nearest neighbors from other helices as well. A weak ENM with $K = 0.01\ \mathrm{kcal/mol/Å^2}$ was used to preserve inter-domain connectivity, allowing the model to access curvature dependent NTD/CTD orientations. CA dimers were then formed by connecting the CG particles within the CTD/CTD interface of two CA subunits with harmonic spring constant $K = 1\ \mathrm{kcal/mol/Å^2}$. Note that our primary objective was to maintain the overall molecular shape of CA dimers, rather than to reproduce atomic level details.

$$E_{bond} = \frac{K}{2}(x - x_0)^2 \tag{1}$$

In mature lattice, CA monomers are highly constrained and the intermonomer interactions are strongly influenced by local curvature, orientation of neighboring monomers, and ring architecture (3, 5, 6) which can only be partially sampled by simulations of isolated dimers or small fragments. Moreover, all-atom simulations underestimate the long-range interactions that drive CA multimerization. To avoid these challenges, our HIV-1 CA CG model incorporates structural information directly from experimentally determined hexamers and CA tubes. Our main goal is to reproduce the experimentally observed organization of CA in lattices. To achieve this, we include two conceptual elements: a steric envelope that defines the molecular shape of CA and a minimal set of attractive interactions representing the experimentally well characterized interfaces that drive capsid self-assembly.

To reproduce the molecular shape of CA, all CG particles interact through soft repulsive potentials (eq. 2).

$$E_{excl}(x_{ij}) = A\left[1 + \cos\left(\frac{\pi x_{ij}}{x_{cut}}\right)\right], \qquad x_{ij} < x_{cut} \tag{2}$$

Here, x_{ij} is the separation between CG particles, $A = 10\ \mathrm{kcal\,mol^{-1}\,Å^{-1}}$ and default $x_{cut} = 10\ \mathrm{Å}$ is the onset of excluded volume repulsion for CG sites *i* and *j*. Pairwise cutoffs are tuned by placing CG monomers onto experimental structures (PDB 3H4E and 3P0A) and adjusting radii to eliminate steric clashes (3). Restrained CG hexamers were superposed onto cylindrical CA density (EMD 5136) to ensure that the coarse model respects the overall steric packing as observed in capsomers and in CA lattices. This ensured CG particles have an effective radius and prevented any unphysical overlap between them.
CA assembly is governed by NTD/NTD contacts within a ring, the NTD/CTD interface between adjacent monomers, and the CTD/CTD trimeric interface at threefold symmetry axes (3). As detailed in Grime et al. 2016, inter-monomer positional statistics were extracted directly from experimental structures (2). To capture these key intra-capsomere interactions relevant to CA assembly, inert "ghost" particles were embedded into the CG structure of CA at selected positions; residues 51, 57 and 63 on the NTD and residue 204 on the CTD. These ghost sites facilitated assembly by interacting exclusively with specified binding interfaces on adjacent CA subunits (assembly competent) when active and remaining inert when inactive (assembly incompetent). The interaction between ghost and specific non-ghost CG particles was governed by an attractive potential defined by a double-Gaussian function:

$$E_{attr}(x_{ij}) = -A\exp\left[-B{x_{ij}}^2\right] - C\exp\left[-D{x_{ij}}^2\right] \tag{3}$$

Here, x_{ij} is the separation between CG particles *i* and *j* with a cutoff radius $x_{cut} = 25$ Å. The parameters A and B describe the depth and width of short-range inter-protomer interaction. Similarly, the parameters C and D represent the depth and width of long-ranged component of Gaussian well. In our simulations, $A = 1\ \mathrm{kcal\ mol^{-1}}$, $B = 0.1\ \mathrm{Å^{-2}}$, $C = 2\ \mathrm{kcal\ mol^{-1}}$ and $D = 0.01\ \mathrm{Å^{-2}}$. Using this approach, we were able to captures the broad morphological diversity of HIV-1 capsids observed experimentally as shown in Fig. S1 (1, 7).

To represent macromolecular crowding, an inert Ficoll70 molecular crowding was used (8). Ficoll70 was represented by two CG spheres, each of which consisted of 42 CG beads and $\sim 5\ nm$ diameter with excluded volume separation of 10 Å. Excluded volume interactions were modeled using a soft cosine potential described in eq. 2 with $A = 10\ \mathrm{kcal\ mol^{-1}\ Å^{-1}}$ and $x_{cut} = 14$ Å. Further details on the CG models for CA and the crowding agents are provided in previous CG studies (1, 2). IP6 was modeled using a single-site CG particle placed at the center of mass and the interaction between IP6 and the CA was modeled using a single-well Gaussian potential:

$$\mathrm{E}_{CA-IP6}(x_{ij}) = -\frac{H}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x_{ij} - x_0)^2}{2\sigma^2}\right] \tag{4}$$

where $x_0 = 8.3$ Å corresponds to the average distance from the center of the R18 ring to the IP6 binding site, as determined from preassembled capsid structure. The value of *H* and σ are -2.8 kcal/mol and 0.4 respectively. Excluded volume interactions between IP6 and CA were modeled using eq. 2 with $A = 10\ \mathrm{kcal\ mol^{-1}\ Å^{-1}}$ and $x_{cut} = 10$ Å.

LEN is a small molecule featuring four ring structures (R1-R4), which form extensive hydrophobic, electrostatic and hydrogen-bonding network within the FG pocket of the neighboring CA subunits (9, 10). R1 interacts with the NTD of both CA subunits, while R2 engages with the CA1-NTD residues. Importantly, R3 ring binds strongly with sidechains of L56, V59, M66 and other residues within the CA1-NTD hydrophobic pocket (9). Finally, R4 ring intercalates with CA1-NTD and CA2-CTD residues N63, M66, T169, L172, R173. The all-atom structure of LEN, derived from its co-crystal complex with the HIV-1 hexamer PDB 6VKV (10), was mapped onto a CG model such that each site corresponded to critical functional groups mediating CA-LEN interactions. For this, we employed the graph-based coarse-graining (GBCG) scheme, a spectral clustering algorithm designed to optimally group atoms into CG sites (11). We performed a spectral decomposition of the atomic-level connectivity matrix of LEN, derived from its bonded interaction network. Guided by the dominant eigenmodes, the molecular graph was then iteratively coarse-grained by merging atoms with the strongest mutual connectivity. Clustering of these modes yielded chemically coherent domains, from which we defined a rigid seven-site CG representation of LEN. Excluded volume parameters were set to $A = 10\ \mathrm{kcal\ mol^{-1}\ Å^{-1}}$ and $x_{cut} = 10$ Å for LEN/LEN and CA/LEN pairs.

Attractive Gaussian potentials were used to represent interaction between LEN CG sites corresponding to R1-R4 rings and CA CG sites corresponding to residues S41, N57, Q67, K70, N74, T107, Y130, R173 to replicate binding modes observed in high-resolution crystal structures (9). These interactions are given by

$$\mathrm{E}_{CA-LEN}(x_{ij}) = -A \exp\left[-\frac{1}{2}\left(\frac{x_{ij} - x_0}{\sigma}\right)^2\right] \tag{5}$$

Here, x_{ij} is the separation distance of CG sites and x_0 is the location of energy minima determined from the co-crystal structure of LEN bound to HIV-1 capsid hexamer (PDB 6VKV) (10). The parameters $A = -2.5\ kcal/mol$ and $\sigma = 0.6$ describe the depth and width, respectively, of a Gaussian energy well used to model LEN R1 interaction with CA1-NTD and CA2-NTD in the capsid binding pocket. Similarly, van der Waals interaction between LEN sites R3/R4 and CA1-NTD or CA2-CTD are defined by $A = -2.0\ kcal/mol$ and $\sigma = 0.6$. To parameterize the CA-LEN attractive interactions, we simulated freely diffusing LEN molecules in the presence of a CG hexamer and identified values that produced the expected occupancy

of roughly 2-3 LEN molecules per hexamer. To further evaluate model fidelity, we performed three independent simulations of a preassembled capsid with 200 molecules of free LEN for 10^8 CG steps. We examined the number of LEN molecules bound to the capsid. In agreement with experimental observations, LEN binding occurred only at hexamer pockets, with no stable binding to pentamers (Fig. S2).

We modeled the viral RNP complex as 2 polymers, which are each 3000 beads in length (the first 112 are of a different type to represent the 5' UTR), and 35 integrase tetramers. Each polymer bead represents 3 bases and bound nucleocapsid protein in the dimeric RNA genome which is approximately 9000 base pairs long. Each bead has a diameter which is enforced by a soft exclusion potential shown below in eq. 2, where A is 10 kcal/mol and x_{cut} is 3 nm. Consecutive CG particles, regardless of type, were connected by harmonic springs to maintain polymer continuity, with the bonding potential defined in eq. 1.

where x is the separation distance, and $x_0 = 1.0\,nm$ is the equilibrium bond distance. The bond coefficient is $\mathrm{K} = 10\ \mathrm{kcal/mol}/nm^2$. We also apply these bonds to the secondary structure contacts found in the 5' UTR according to the map from Goodsell et al (12-14). This includes the contacts between the two different chains in the dimerization sequence of the RNA. We also add angle potentials.

$$U_{Angle}(r) = \frac{K_\theta}{2}\,(\theta - \theta_0)^2 \qquad (\mathbf{6})$$

To prevent unphysical bending of the RNA chains. The equilibrium angle (θ_0) is completely straight, 180°, although the bending constant is relatively weak, $K_\theta = 3\frac{kcal}{mol}/rad^2$ (eq. 6). We use a gaussian attraction to model the condensation of the RNA globule due to electrostatic interactions between NC and the RNA as in Eq.4. Differing attraction strengths are used between the structured 5' UTR and the rest of the RNA polymer.

The integrase tetramers consist of 4 beads each representing one integrase unit. They are bonded together and to the RNA/NC using harmonic bonds of the form shown in Eq. 1, where $K = 1000\frac{kcal}{mol}/nm^2$ and the bond lengths vary. Interactions between the integrase units and RNA/NC beads are given by a gaussian attraction shown in Eq. 4. Here the width, $\sigma = 0.5\,nm$, the equilibrium distance,$x_0 = 4\,nm$, and the interaction strength, $H = 0.9\frac{kcal}{mol}/nm$, which was calibrated to control the amount of surface accessible integrase that CA can interact with.

Experiments have suggested that integrase (IN) is necessary for proper encapsulation of the RNP by the growing capsid and thus we place attractive interactions between CA and the integrase while only volume-excluding interactions exist between CA and the rest of the RNP (15). Interactions between CA and IN are applied through each of the 4 integrase sites on the tetramer and site 115 on each CA. Interactions between CA and the RNA/NC are given by soft repulsions as in Eq. 2, where the diameter, $x_{cut} = 2\,nm$. Attractive interactions come from a Gaussian potential in Eq. 4 where $\sigma = 0.2\,nm, x_0 = 3\,nm$, and $H = 2.25\frac{kcal}{mol}/nm$. Interactions between the other RNP beads and CA are left as a soft repulsion with a cutoff, $x_{cut} = 3\,nm$. These interactions and interactions between RNA/NC and the integrase were calibrated together to ensure successful assembly and RNP encapsulation in the absence of LEN. The two main parameters that control the assembly of CA on the RNP are the interactions between CA and IN and the interactions between IN and the rest of the RNP (R). CA-IN interactions control the amount of IN that is on the surface of the RNP and thus accessible to the CA, which will interact with the IN with a certain strength. In Fig. S4, we tested various combinations of these parameters with only IP6 and no LEN present showing the capsids that encapsulate the RNP, if any. We also show the number of integrases that end up interacting with CA. We found that too many integrases interacting too strongly with the capsid led to defective capsids while not enough integrases interacting too weakly led to failed encapsulation. We choose the following values to balance these factors. $H'(IN{-}R)$ = 0.56 $kcal/mol$ and $H'(IN{-}CA)$ = 1.8 $kcal/mol$ correspond to the values of H mentioned previously.

Tables

# of CA dimers	# of IP6	# of LEN	# of RNP	# of crowders	replicates
880	-	200	-	2500	3
880	200	200	-	2500	4
880	600	200	-	2500	4
880	600	220	-	2500	10
880	600	440	-	2500	10
880	600	880	-	2500	10
880	600	1760	-	2500	10
800	600	400	+	2000	10

Table S1. Summary of the HIV-1 CA assembly simulations performed under different IP6, LEN and RNP conditions.

Figures

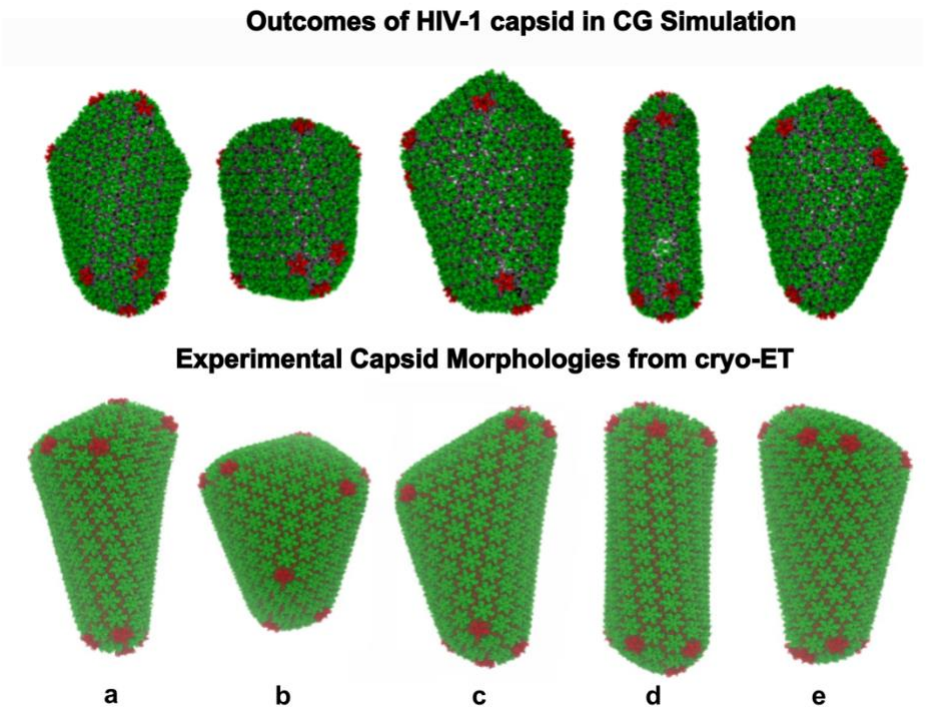


Fig. S1. HIV-1 capsid morphological diversity. (a-e) Our CG model recapitulate the morphological diversity observed in cryo-ET reconstructions of HIV-1 capsid. The top panel is reprinted from ref. 1, which is licensed under CC BY 4.0.

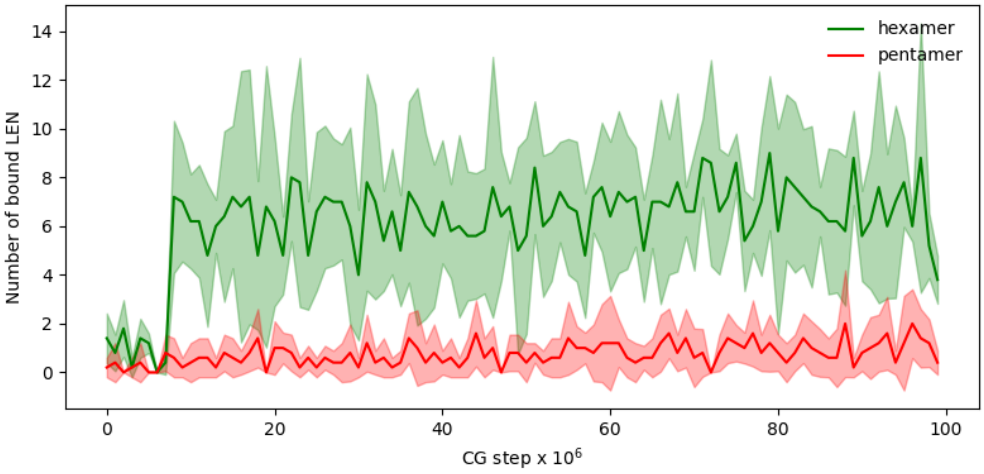


Fig. S2. Time series profile of LEN binding with capsid. Number of LEN molecules bound to the capsid hexamer (green) and pentamer (red) within the simulation period.

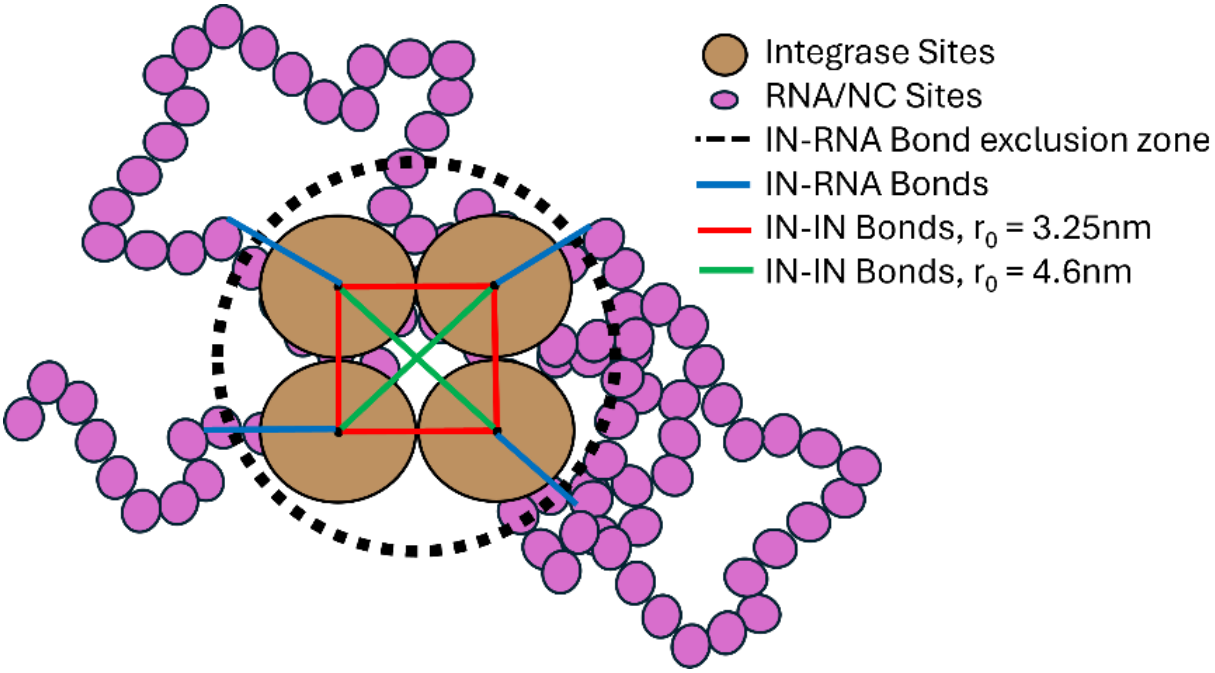


Fig. S3. Model for the bonded topology of integrase crosslinking the RNA. As mentioned in the main text, the integrase tetramers consist of 4 beads bonded together, with each representing one integrase unit. The bond topology is shown in Fig. S3, where the red lines represent bonds where r_0 is 3.25nm and the green lines represent bond which are 4.6 nm long. Integrase units are also bonded to the RNA/NC beads. Using the first model with integrase tetramers and self-avoiding gag from Goodsell et al. as the initial structure, each individual integrase is bonded to the RNA/NC bead they are closest to, not including beads which are less than 4.6 nm from the center of the tetramer. This ensures that no RNA/NC bead is bound to more than integrase unit. The equilibrium distance for these bonds is set to the distance in the initial structure.

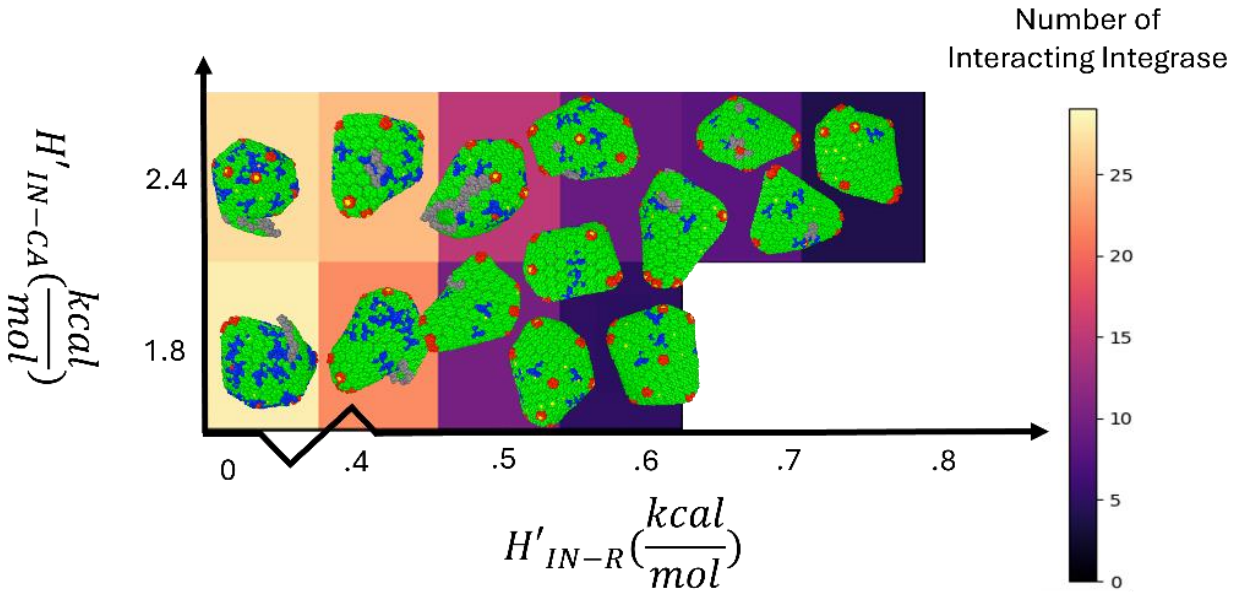


Fig. S4. Assembly results in the absence of LEN. different values of the integrase-RNP interaction strength, H'_{IN-R} and integrase-CA interaction strength, H'_{IN-CA}. CA are shown.

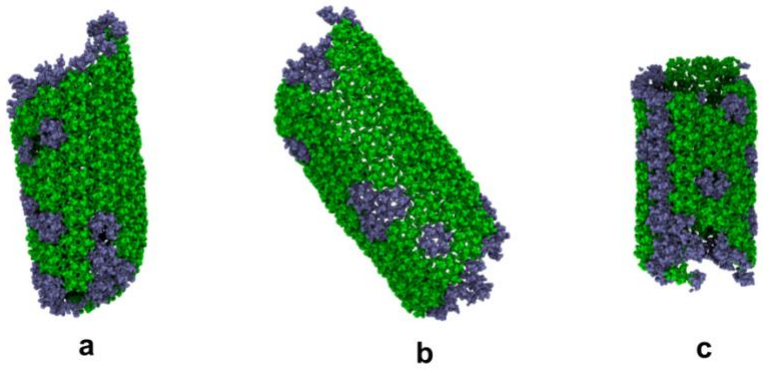


Fig. S5. LEN induced HIV-1 CA assembly in the absence of IP6. (a-c) Simulation snapshots taken after 2×10^8 CG steps shows that LEN promoted tubular structures in the absence of IP6.

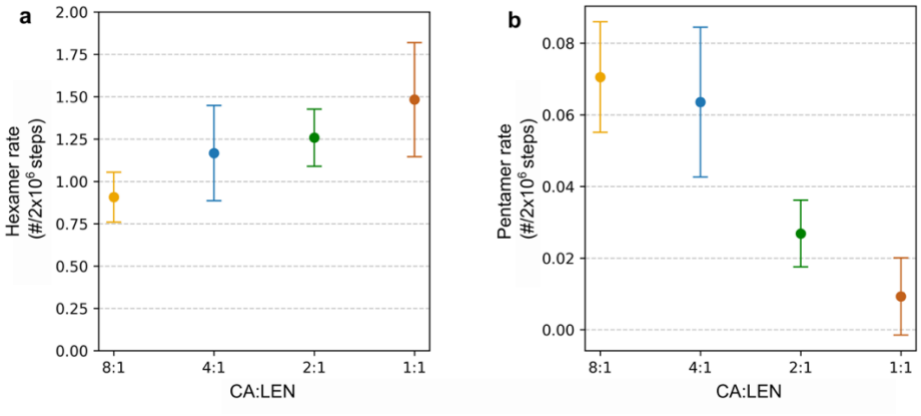


Fig. S6. HIV-1 capsid early assembly statistics under varying CA/LEN conditions. (a-b) Mean formation rate (per 2×10^6 CG steps) for (a) hexamers and (b) pentamers.

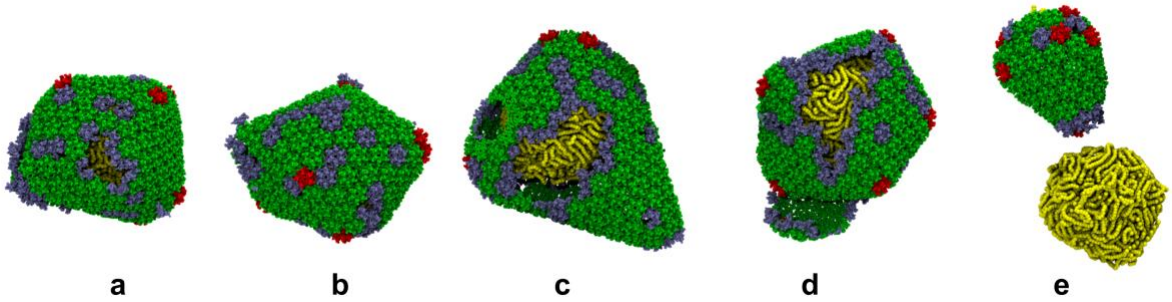


Fig. S7. Additional simulations of capsid-RNP packaging in the presence of LEN. (**a-e**) Simulation snapshots showing that capsid assembled malformed architectures in the presence of LEN. On few occasions LEN treated capsid failed to enclose RNP partially or completely.

Movie S1: Capsid morphogenesis at low LEN levels. All four independent trajectories demonstrate that capsid can adopt conical geometry, with pentamers forming during both early and late stages of assembly. CA assembly is like WT, with the broad end forming first and growing into a hook-like structure. However, CA lattice closure is delayed and substantial defects emerge at the high curvature regions.

Movie S2: Capsid morphogenesis at high LEN levels. In one trajectory, a partially conical capsid forms, while in others, the CA lattice grows rapidly from both ends and anneals through end-to-end joining, characteristic of tubular structures. Pentamers fail to form during the late stages, preventing proper lattice closure. In all cases, an additional CA layer assembles on the lattice, reflecting elevated assembly kinetics at high LEN concentrations.

SI References

1. M. Gupta, A. J. Pak, G. A. Voth, Critical mechanistic features of HIV-1 viral capsid assembly. *Sci Adv* **9**, eadd7434 (2023).
2. J. M. A. Grime *et al.*, Coarse-grained simulation reveals key features of HIV-1 capsid self-assembly. *Nat Commun* **7**, 11568 (2016).
3. O. Pornillos *et al.*, X-ray structures of the hexameric building block of the HIV capsid. *Cell* **137**, 1282-1292 (2009).
4. I. J. Byeon *et al.*, Structural convergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. *Cell* **139**, 780-790 (2009).
5. G. Zhao *et al.*, Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **497**, 643-646 (2013).
6. J. C. Stacey *et al.*, Two structural switches in HIV-1 capsid regulate capsid curvature and host factor binding. *Proceedings of the National Academy of Sciences* **120**, e2220557120 (2023).
7. S. Mattei, B. Glass, W. J. Hagen, H.-G. Kräusslich, J. A. Briggs, The structure and flexibility of conical HIV-1 capsids determined within intact virions. *Science* **354**, 1434-1437 (2016).
8. M. del Alamo, G. Rivas, M. G. Mateu, Effect of macromolecular crowding agents on human immunodeficiency virus type 1 capsid protein assembly in vitro. *J. Virol.* **79**, 14271-14281 (2005).
9. S. M. Bester *et al.*, Structural and Mechanistic Bases of Viral Resistance to HIV-1 Capsid Inhibitor Lenacapavir. *mBio* **13**, e0180422 (2022).
10. S. M. Bester *et al.*, Structural and mechanistic bases for a potent HIV-1 capsid inhibitor. *Science* **370**, 360-364 (2020).
11. M. A. Webb, J.-Y. Delannoy, J. J. De Pablo, Graph-based approach to systematic molecular coarse-graining. *Journal of chemical theory and computation* **15**, 1199-1208 (2018).
12. D. S. Goodsell, A. Jewett, A. J. Olson, S. Forli, Integrative modeling of the HIV-1 ribonucleoprotein complex. *PLoS computational biology* **15**, e1007150 (2019).
13. A. I. Jewett, Z. Zhuang, J.-E. Shea, Moltemplate a Coarse-Grained Model Assembly Tool. *Biophys. J.* **104**, 169a (2013).
14. A. I. Jewett *et al.*, Moltemplate: A tool for coarse-grained modeling of complex biological matter and soft condensed matter physics. *Journal of molecular biology* **433**, 166841 (2021).
15. J. J. Kessl *et al.*, HIV-1 integrase binds the viral RNA genome and is essential during virion morphogenesis. *Cell* **166**, 1257-1268. e1212 (2016).