Title

Critical Mechanistic Features of HIV-1 Viral Capsid Assembly

Authors

Manish Gupta,[1] Alexander J. Pak, [1,⊥] 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, IL 60637, United States

[⊥]Present address: Department of Chemical and Biological Engineering, Colorado School of Mines, Golden, CO 80401, United States

[*]Corresponding author: gavoth@uchicago.edu

Abstract

The maturation of HIV-1 capsid protein (CA) into a cone-shaped lattice encasing the condensed RNA genome is critical for viral infectivity. During HIV-1 maturation, CA can self-assemble into a wide range of capsid morphologies made of ~175-250 hexamers and 12 pentamers. Most recently, the cellular polyanion inositol hexakisphosphate (IP6) has been demonstrated to facilitate conical capsid formation by coordinating a ring of arginine residues within the central cavity of capsid hexamers and pentamers. However, the precise kinetic interplay of events during IP6 and CA co-assembly is unclear. In this work, we use Coarse-grained Molecular Dynamics (CGMD) simulations to elucidate the underlying molecular mechanism of capsid formation, including the crucial role played by IP6. We show that IP6, in relatively small quantities at first, promotes curvature generation by trapping pentameric defects in the growing lattice and shifts assembly behavior towards kinetically favored outcomes. Our analysis also suggests that IP6 can stabilize metastable capsid intermediates and can induce structural pleomorphism in mature capsids.

Teaser

Computer simulations reveal that IP6 promotes fullerene-like capsid formation by stabilizing high curvature regions of the capsid.

Introduction

The HIV-1 replication process requires budding of immature viral particles that subsequently undergo a crucial structural rearrangement called maturation.(*1, 2*) HIV-1 maturation is a string of biochemical and morphological changes induced by proteolytic processing of Gag by viral protease (PR),(*1, 3*) leading to self-assembly of released capsid protein (CA) into a cone-shaped capsid(*4*) (mature core) encasing a dense RNA/nucleocapsid (NC) complex.(*5*) The mature capsid executes several critical functions during reverse transcription (RT)(*6*), cellular trafficking(*7*) and nuclear import.(*8, 9*) Thus, mature capsid generation is essential for HIV-1 infection, with failure resulting in loss of viral infectivity.(*10-12*)

The CA monomer contains two independent domains, the N-terminal domain (NTD) comprising seven α-helices and a β-hairpin and the C-terminal domain (CTD) comprising four α-helices, that are connected by a flexible linker.(*13*) Interestingly, CA can adopt a dimeric form (Fig. 1a) in solution due to interactions between CTD helices(*13, 14*) and exists in a dynamic monomer/dimer equilibrium.(*14*) CA can self-assemble into a wide variety of morphologies while a typical capsid is a closed fullerene shell (Fig. 1d), composed of hexamer and pentamer building blocks (Fig. 1c) with more than 1000 copies of CA.(*4, 15*) Typical HIV-1 virions contain 2500-5000 Gag monomers in a non-trivial and highly crowded environment.(*16-19*) Therefore, a mature capsid, which contains only half of the CA present in an immature virion, exists in equilibrium with solution state CA. As the assembly condition is sensitive to variation in CA concentration, pH and salt concentration, biochemical reconstitution of CA assembly has been particularly challenging.(*20*) Under favorable conditions, a mixture of cone and cylinder formation was observed in most of the previous *in vitro* studies.(*21, 22*) Cryo-EM and image reconstruction studies have identified several key structural motifs for capsid assembly that involve interactions between the NTD/NTD, NTD/CTD and CTD/CTD interfaces of CA.(*23, 24*) Characterization of CA tubular assemblies showed that the NTDs form a hexagonal arrangement while inter-hexamer CTD interfaces connect adjacent assembly units.(*23*) Based on various *in vitro*(*23, 24*) and intact virion(*25*) analyses, the atomic structure of mature cores, following a fullerene cone model, with exactly 12 pentamers(*22, 24, 26*) in agreement with Euler's theorem, has also been determined. Recent X-ray crystallography studies reveal that a narrow pH gated pore with a ring of R18 at its center is distributed throughout the capsid surface.(*27, 28*) This ring of R18 is essential for RT and avidly binds several host cell metabolites.(*27*) Interestingly, HIV-1 packages more than 300 ions of negatively charged IP6 per virion,(*29*) which has been found to influence immature Gag assembly.(*29, 30*) Pioneering work by Dick *et al*. demonstrated that IP6 is also critical for enhanced assembly of cone-shaped

mature cores.(*31*) IP6 has been shown to interact with the ring of R18 and removal of electropositive repulsion in the central pore by R18A mutation precludes viral infectivity.(*31*) Therefore, it is evident that IP6 plays a central role in infectious capsid formation. In a previous work from this group,(*32*) the structural basis for interaction between IP6 and capsid building blocks was illustrated using all-atom (AA) molecular dynamics (MD). Importantly, free energy calculations indicated that IP6 preferentially stabilizes CA pentamers over hexamers. Therefore, it was hypothesized that IP6 promotes capsid formation by stabilizing pentamers, which are necessary for inducing curvature in the capsid. However, the detailed molecular mechanism of IP6 and CA co-assembly and several important questions regarding the kinetic mechanism of pentamer inclusion and curvature generation during the HIV-1 self-assembly remain elusive.

All-atom simulations offer complementary insight to experimental studies. AAMD simulations of full-capsid have revealed several key physical and chemical properties of capsid dynamics and structure,(*33, 34*) nonetheless, these techniques are computationally demanding and, therefore inadequate for studying the large-scale functional dynamics of capsid assembly processes that occur over long time scales. Alternatively, coarse-grained (CG) approaches reduce system complexity to a simpler molecular description while retaining the ability to predict emergent behavior.(*35*) While caution should always be exercised when considering large-scale molecular assembly, general agreement with experimental results is often considered a reliable indicator of CG model fidelity. Previous CG simulations(*36, 37*) have offered valuable insight into the effects of CA conformational freedom, concentration, and molecular crowding on self-assembly, thus explaining the initial stages of capsid growth. In this work, however, we extend our prior CG model(*36*) to include, and investigate, the effects of IP6 on HIV-1 capsid formation. In particular, we examine the kinetic pathways promoted by IP6 that lead to capsid morphogenesis. Importantly, we observe that IP6 is essential for the formation of high curvature lattice regions representing kinetically favored intermediates. We therefore propose that a physiological role for IP6 is to induce lattice curvature by stabilizing a small population of CA pentamers during the early stages of the capsid self-assembly. Subsequent to that process, a larger number of IP6 ions then intercalate into the formed or nearly formed capsid lattice to add further stabilization.

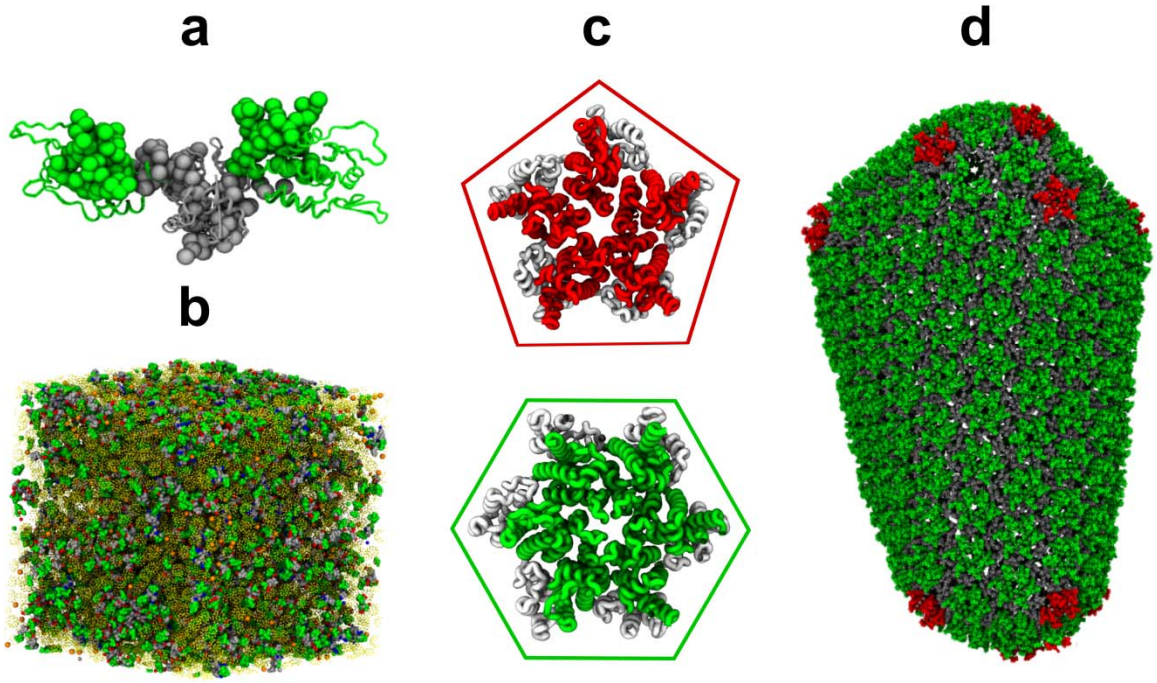

Figure 1. CG model of HIV-1 CA dimer and CGMD simulation setup. (a) Overlay of a CG CA dimer (NTD green and CTD gray) over an all-atom CA dimer (PDB 2ML8).(*14*) Shown in this figure are the CG sites or "beads". (b) Snapshot of an initial simulation setup; CA dimer (green, gray beads and blue/red for active/inactive), IP6 (orange beads), inert crowder (transparent yellow). (c) Quasi-equivalent pentamers (PDB 3P05) and hexamers (PDB 5HGL) after ref 15 and 26 respectively. CTDs are gray and NTDs are red for capsid pentamer and green for hexamer. (d) Mature core as made of pentamer (red) and hexamer (green) building blocks.

Results

CA assembly behavior in the absence of IP6

HIV-1 CA monomers consist of two globular domains (NTD and CTD) connected by a flexible linker. The monomer is conformationally flexible and adopts an assembly competent (i.e. mature-like) state ($[CA]_+$) with only 5-10% probability in solution.(*14*) To emulate this behavior in our CG simulations, CA was partitioned into assembly competent ($[CA]_+$) and assembly incompetent ($[CA]_-$) populations and all free CA were periodically assigned to either the $[CA]_+$ or $[CA]_-$ state with a fixed $[CA]_+$ probability; the state switching rate was chosen to maintain assembly relevant correlation times (conformational switching interval), i.e., inter-domain motions. This approach is the so-called "ultra" CG (UCG) model in which various physical or chemical processes can implicitly be overlaid on the kinematic motion of the CG particles in

order to describe (and include) features that are impossible to include at the explicit CG resolution of the model. UCG modeling can be exceptionally powerful and, in this case, has enabled the simulation of HIV-1 capsid formation from more than 1000 proteins. As noted earlier, herein the UCG model for HIV-1 capsid has been modified to include explicit CG IP6 anions in the simulation.

Several *in vitro* experiments have demonstrated that, in the absence of IP6, CA predominantly forms hollow tubes made of hexamers.(*24, 31*) While our prior CG model(*36*) of CA predicted cone formation, we tested whether our current modified CG model was also capable of producing tubes, resulting in a mixed population of tubes and cones. To simulate CA assembly, 4 independent trajectories with 600 CA dimers were propagated for 5×10^8 CG steps and 200 mg/mL inert crowder density with $[CA]_+/[CA]_{free}$ = 15% and conformational switching interval of 5×10^5 time steps. The CA concentrations in our simulations were about 50% of a typical virion due to computational expense. However, by increasing $[CA]_+$ to 15% in comparison to our prior studies(*36, 38*) that used $[CA]_+$ = 10%, we compensated for the possible undersupply of assembly competent CA. From our CG simulations, we observed 3 tubular structures and 1 truncated cone (Figs. 2a and S1). CA dimers quickly produced metastable trimer of dimer populations ($< 5 \times 10^6$ CGMD steps), but nucleation of these trimers was slow and did not occur before 10^7 CG steps. The growing lattice curled into a hollow cylinder and grew rapidly by CA dimer association at both ends generating a sheet of hexamers (Fig. 2a). After 3×10^8 CG steps, pentamer defect incorporation at the edge of the lattice was observed forming a narrow end (Fig. S1). Interestingly, CA self-assembly at the exposed lattice edges was dynamic and exhibited a degree of local remodeling by “error correction”. The other end of the CA tube continued to grow until the population of assembly competent CA was exhausted in the simulation box. Snapshots taken at various stages of CA assembly in the absence of IP6 are shown in Fig. 2a and the full process is shown in the movie S1. As shown in Fig. 2b, the final CA tube containing 845±40 CA monomers was primarily composed of hexamers (126 ± 7). The diameter of the final structure was 38 ± 6 nm, which is consistent with tube diameters from *in vitro* experiments.(*23*) The rarer appearance of truncated cones (Fig. S1) reflects the ability of our CA model to assemble spontaneously into both helical tubes and cones with a preference for tubular structures in the absence of IP6. These results are reproducible in additional simulations albeit with a statistical spread.

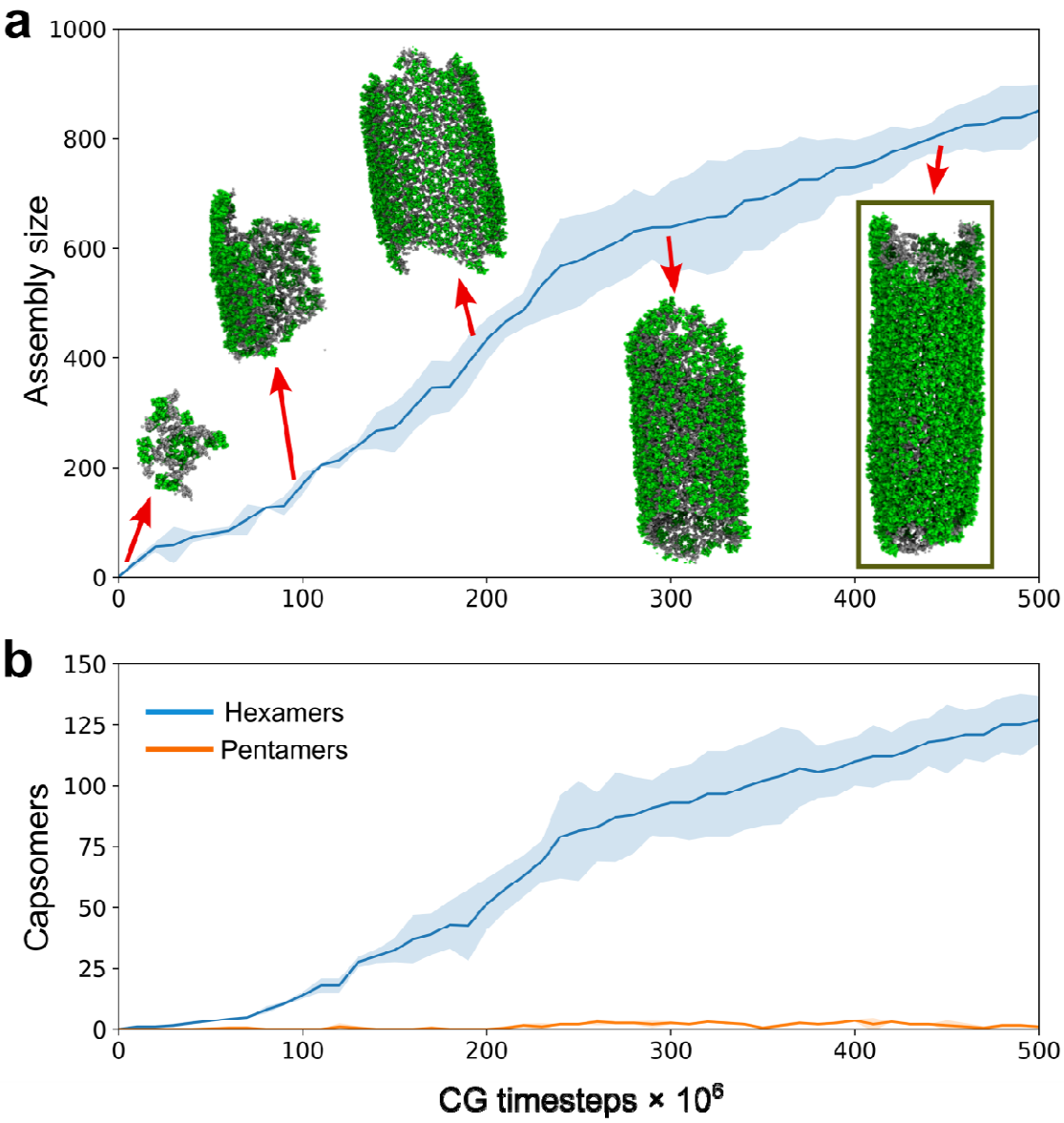

Figure 2. CGMD simulations of CA dimer assembly in the absence of IP6. (a) Assembly time-series plot that depicts cluster size as a function of CG time. CA dimers preferably assemble into tubular structures in the absence of IP6 (CA NTDs are green and CTDs are gray). Simulation snapshots taken at , , , , CGMD time steps. (b) Time-series profile of the total number of hexamers and pentamers as a function of CG time step. In both panels, the shaded region represents the standard deviation across all four replicas.

Molecular mechanism of IP6 mediated CA assembly

Recent studies have shown that the negatively charged host metabolite IP6 can stabilize the six helix-bundle(*31*) in the immature Gag and is efficiently recruited by HIV-1 virions.(*39*) After proteolysis, IP6 is known to stabilize the mature capsid by binding at the central ring of R18.(*32*) IP6 is also critical for enhanced mature capsid formation.(*31*) Thus, IP6 shifts the population of assembled CA morphologies towards mature capsid cones through as-yet unknown mechanisms. A typical mature capsid contains 1000-1500 CA monomers.(*4, 15*) Therefore, we increased the simulation box domain to 90×90×90 nm^3 with 880 CA dimers to allow the formation of complete capsids. To examine the effects of IP6 during self-assembly, simulations were performed with 4mM CA and 600 IP6 molecules using the aforementioned assembly conditions, including the concentration of inert crowders and the CA conformational switching rate described earlier. Figure 3 presents different stages of IP6 mediated capsid assembly. A small and curled hexameric lattice formed initially with small curvature since the dihedral angle between adjacent hexamer-hexamer planes was close to 180°. This explains the formation of the relatively flatter middle region in the initial stages of capsid assembly (~ CGMD steps). Subsequently, stable pentamers formed at the broad end of the capsid and were stabilized by IP6, imparting curvature to the capsid shell. The growth of the capsid lattice followed a sigmoidal profile, a characteristic also observed experimentally.(*40*)

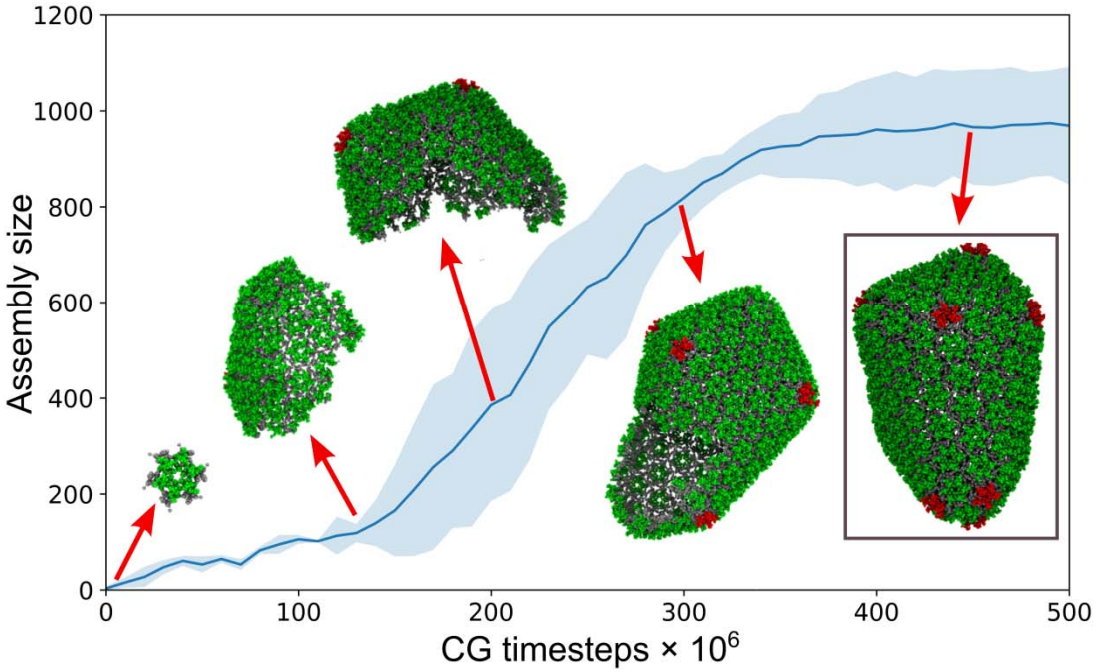

Figure 3. CGMD simulations of CA dimer assembly in the presence of IP6. (a) Time-series plot of assembled lattice size as a function of CG time step. CA dimers assemble into a conical capsid in the presence of IP6 (CA hexamer NTDs are green, pentamer NTDs are red and all CTDs are gray). Simulation snapshots taken at , , , , CGMD time steps. The shaded region represents the standard deviation across six simulation replicas.

Interestingly, the narrow end of the cone formed last in our simulations where five proximal pentamers emerged, leading to the high-curvature cap shown in Figure 3 and movie S2. For comparison, our prior CG model(*36*) in the absence of IP6 tended to form the narrow high curvature cap first. In the present case, the complete capsid contained 978±121 CA subunits and was observed after CGMD steps. We analyzed time series profiles of CG trajectories in Figure 4, in which we find that the complete capsid core is composed of 151±21 CA hexamers and exactly 12 pentamers.

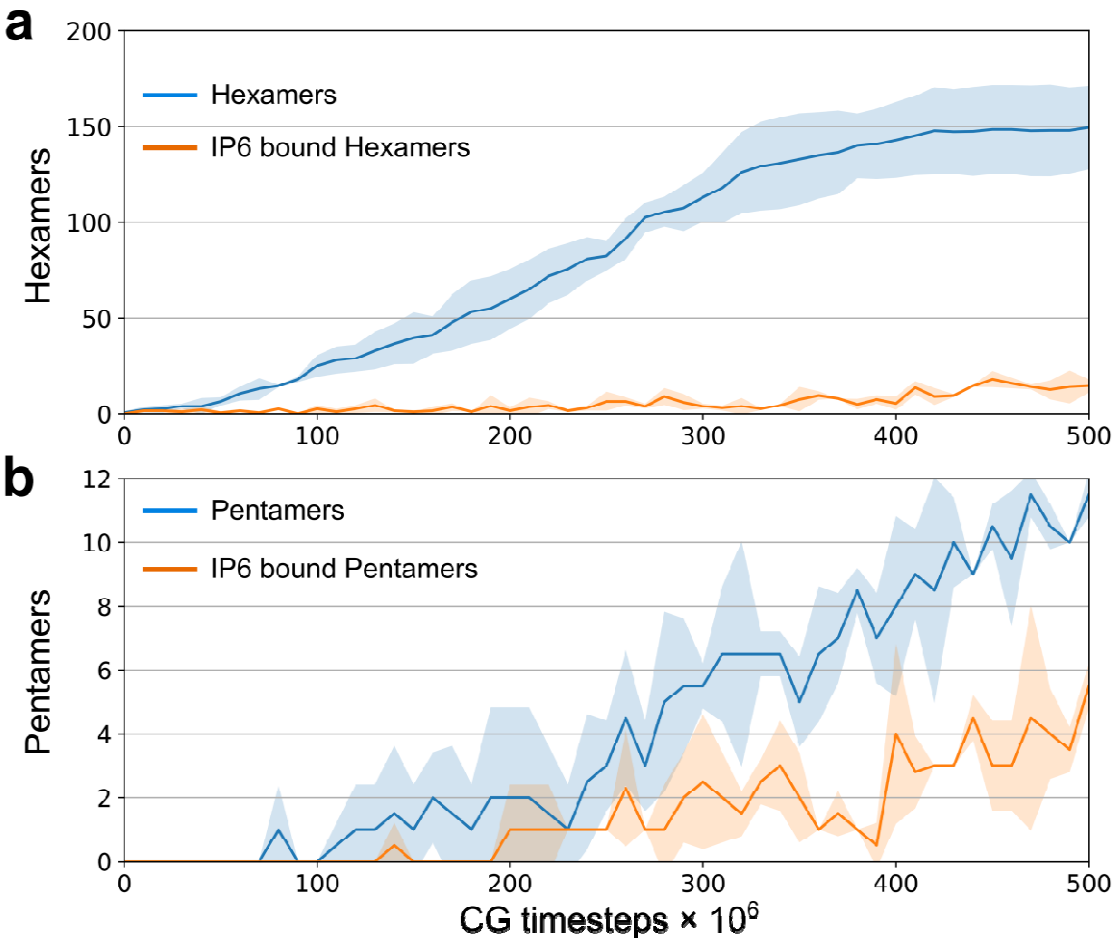

Figure 4. Assembly time-series profiles illustrate the preference of IP6 towards pentamers over hexamers. (a) The total number of hexamers and the total number of IP6 bound hexamers in the capsid

as a function of CGMD time, (b) The total number of pentamers and the total number of IP6 bound pentamers in the capsid as a function of CG time. In both panels, the shaded region represents the standard deviation across six simulation replicas.

To further understand the role of IP6 in CA assembly kinetics, we counted the number of IP6 bound capsid hexamers and pentamers as the simulation progressed. While IP6 bound with only 15-20% of the hexamers, by contrast at least 50% of the pentamers were bound with IP6 (Fig. 4a & b), indicating that IP6 shows preference towards pentamers in our simulations even under such complex conditions.

IP6 accelerates CA assembly rate and stabilizes regions of high curvature

To further investigate the precise behavior of CA assembly, we compared the early stages of assembly in the absence and presence of IP6. We performed 10 additional short simulations, 5 each for CA without IP6 and CA with IP6 starting from a pre-nucleated hexameric cluster with 6 CA dimers. We depict the assembly rate of CA in Figure 5 and analyze the two cases. In the presence of IP6, the assembled clusters contain 150 ± 44 CA monomers after CGMD steps. During the early stages of assembly, IP6 binds with the nucleated cluster and accelerates CA assembly compared to that of the IP6-absent condition. While local remodeling at the growing lattice edges is a key feature of capsid assembly, we observed that IP6 binding stabilizes defects in capsid intermediates, thereby curtailing extensive error correction and facilitating CA association.

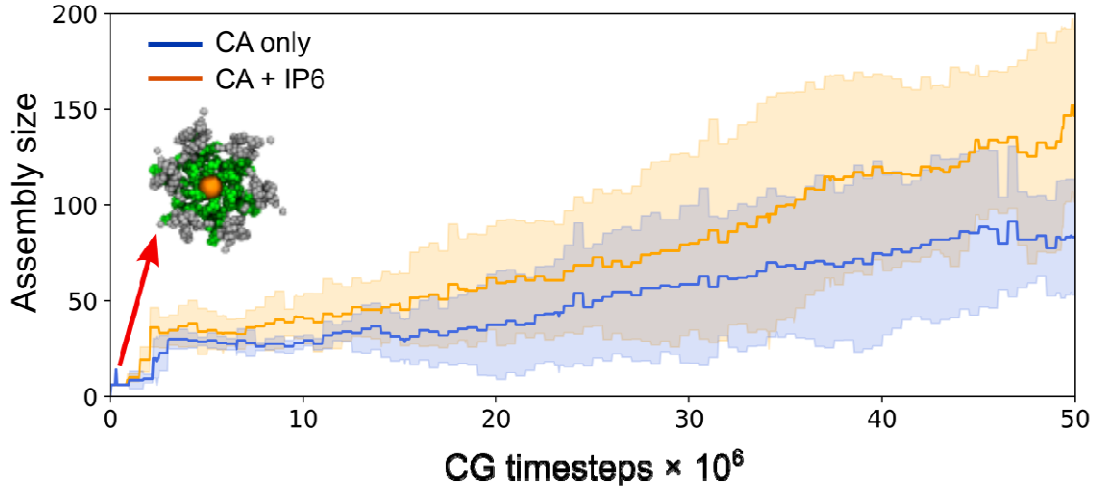

Figure 5. IP6 accelerates CA assembly rate by kinetically trapping capsid intermediates (color scheme is same as Figure 3. IP6 is orange here). Time series profiles of cluster size as a function of

CGMD time step in the presence and absence of IP6. In both panels, the shaded region represents the standard deviation across five simulation replicas.

It has been noted that HIV-1 capsids possess a caged polyhedral arrangement composed of CA hexamers and exactly 12 pentamers.(*4, 24*) These pentamers introduce curvature on the capsid and are necessary for capsid closure. On the basis of 2.5 Å X-ray crystallography structures, hexamers and pentamers were determined to be quasi-equivalent(*15*) with the positively charged R18 sidechains throughout the central pore juxtaposed more closely in the pentamer compared to the hexamer, yielding stronger electrostatic repulsion in the former. Pentamers are considered defects and are disfavored compared to regular hexamers. Our CA simulations also show a certain degree of error correction where defective pentamers are annealed into hexamers. Intriguingly, at the end of the CGMD runs when the capsid had formed, the high curvature regions both at the broad and narrow ends of the capsid had more IP6 population than the comparatively flatter middle region as shown in Figure 6.

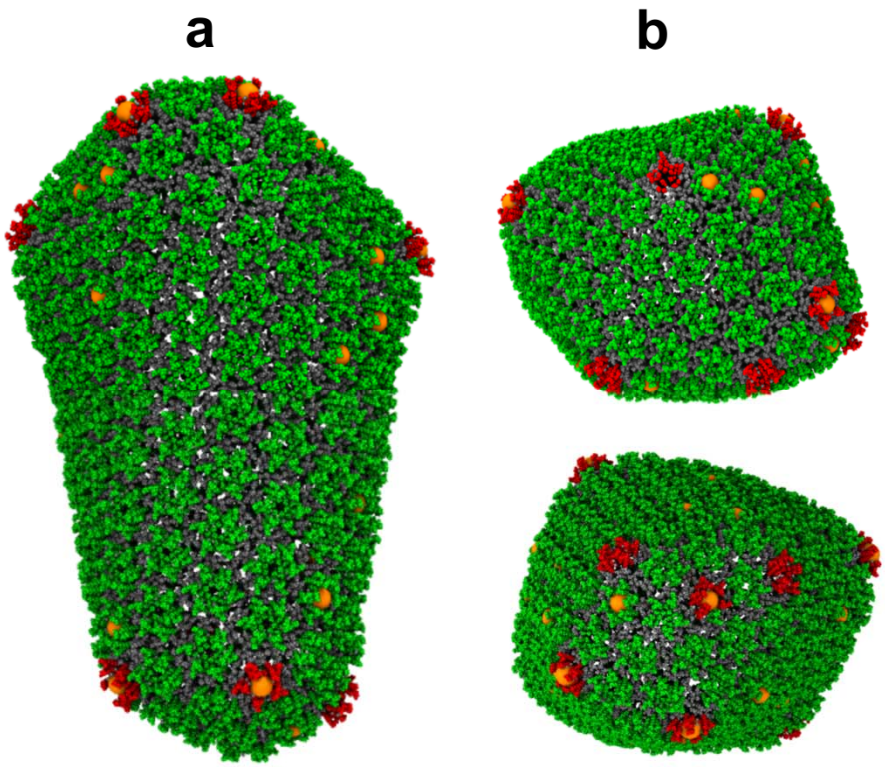

Figure 6. CGMD simulations reveal that IP6 stabilizes the capsid by binding at regions of high curvature. (a) Side view of capsid (color scheme is same as Figure 3). (b) Top and bottom view.

We observed that IP6 prefers binding to the regions of high curvature and are responsible for lattice curvature generation. It should be noted that only 15 to 35 IP6 ions intercalated with the capsid shell during CA assembly in the CGMD runs but IP6 binding continued even after full capsid formation (Fig 4). Recent experimental studies show that the narrow end of the HIV-1 core may not close if the conditions are not favorable.(*41*) However, all capsids generated during our simulations closed completely suggesting that IP6 assists in closure by stabilizing pentamers which accommodate curvature at regions of high stress. This is largely an effect occurring on the kinetic landscape along the capsid formation pathway, while the thermodynamic minimum generally involves the fully formed capsid with significantly more IP6 ions intercalated into its CA lattice.

IP6 stabilization of capsid intermediates determines mature virion morphology

The fullerene core is a defining feature of HIV-1 maturation.(*3*) However, it is not unusual for HIV-1 viral capsids to display morphological diversity.(*21, 25, 40, 42, 43*) Depending on the spacing between pentameric defects, capsid cores adopt a wide distribution of cone angles thereby producing different morphologies. Analogously, the variable distribution in capsid sizes is suggestive of a robust assembly process. Figure 7 presents different morphological outcomes during our simulations of CA self-assembly with IP6. We observed that curvature generation in capsid intermediates was always dictated by stabilization of pentameric defects by IP6. Unusually early incorporation of these pentameric defects resulted in smaller cores (Figure 3 and 7b), whereas more timely pentamer incorporation into the forming extended hexagonal lattice sheet led to appropriate curvature generation for relatively larger mature core formation as shown in Figure 7a,c,e. Although rarer than the regular conical capsid, a pill-shaped core (Fig. 7d), capped by the introduction of six pentameric defects at both ends, was also observed during our simulations. It is noteworthy that our simulations even reproduce this unusually shaped capsid, as well as other structures, seen in cryo-ET studies of actual mature HIV-1 virions.(*25*)

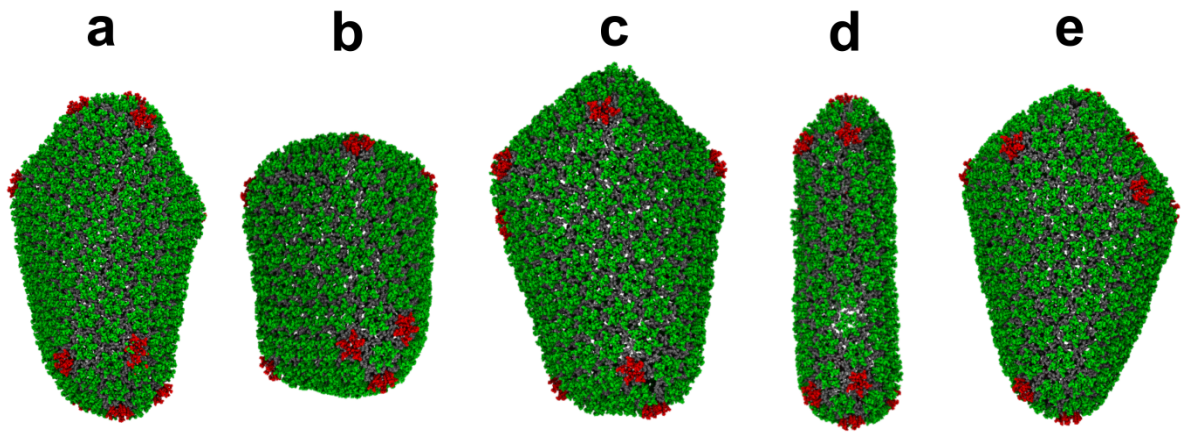

Figure 7. IP6 promotes capsid pleomorphism. (a-e) Snapshots of final structures from five additional CGMD simulations in the presence of IP6 (CA color scheme as in Figure 3). IP6 molecules stabilize lattice curvature inducing the formation of kinetically trapped structural outcomes.

Discussion

We have used CGMD simulations to elucidate the critical role of IP6 during HIV-1 capsid self-assembly. Our CG simulations capture assembly behavior at time and length scales relevant for viral capsid morphogenesis and they generate virion relevant morphologies, thus demonstrating that CG simulations can provide crucial insight into CA assembly kinetics. While the RNA complex is known to promote capsid nucleation,(*26*) it should be noted that the simulations presented here assume that RNA is not essential for capsid formation, which is in agreement with experiments that indicate cone formation can be achieved in the absence of RNA, although we expect RNA to also accelerate CA assembly.(*44*) In the absence of IP6, we find that mature CA assembly is self-regulated through efficient error-correction, leading to tubule formation. This ensures increased contact among CA monomers and removal of transient pentamer defects by local remodeling, suggesting that CA assembly nominally proceeds through a thermodynamically-driven pathway producing near-equilibrium structures. However, depending on the local $[CA]_+$ concentration and crowding, lattice growth can be too rapid for error correction, leading to conical structures with stable pentamers embedded within the lattice as seen occasionally in our simulations and select *in vitro* experiments.(*22*) This explains the primary formation of cylinders and the rarer occurrence of cone-shaped mature cores in the absence of IP6.

The morphological transition from an immature spherical Gag lattice to the mature core within HIV-1 virions is not well understood but there are two competing models for this

transformation: the disassembly/reassembly model(*45*) and the displacive mechanism.(*46, 47*) Evidence in support of the disassembly/reassembly model includes several structural studies demonstrating distinct lattice spacings and different protomer-protomer contacts between the immature and mature shells.(*5, 48*) Alternatively, the transition of CA-SP1 mutants with cleavage defects into mature cores supports the displacive model.(*47*) Recent *in vitro* experiments and phenomenological computer simulations indicate a sequential combination of both displacive and disassembly/reassembly pathways during capsid maturation(*49*) where the growth of nucleated CA proceeds through initial formation of the broad capsid followed by capsid closure at the narrow end.(*50, 51*) Our simulations explore the basis of IP6 mediated capsid growth and closure, which is relevant for both the sequential displacive and disassembly/reassembly pathways and CA assembly under *in vitro* conditions. It has been unclear why viral capsids assume conical morphologies with pentameric defects over more stable cylinders. However, our results show that n the absence of IP6 pentamers are unstable and eventually get annealed into hexamers, as evidenced by fluctuation in pentamer population in our simulations (Fig. 2b). By contrast, in the presence of IP6 a combined effect of accelerated assembly and early pentamer stabilization by IP6 allows for the conical structure to form with the incorporation of exactly 12 stable pentamers. However, we note that, over our simulation timescales, not all pentamers bind to IP6, indicating that only a few pentamers stabilized by IP6 are sufficient to facilitate curvature generation with subsequent pentamer formation and hence the closed capsid formation. This result also explains CA tubule formation under low concentrations of IP6 *in vitro*.(*31*) Cone formation may be difficult to accomplish at low concentrations of IP6 due to the stochastic nature of pentamer trapping by IP6 intercalation. It should be noted, however, that our simulations predict additional IP6 intercalation even after the full capsid formation. Our simulations also suggest a delicate balance between curvature generation and lattice growth during CA self-assembly. Therefore, depending on the extent of capsid growth achieved before the inclusion of pentamers necessary for capsid closure, mature capsids may contain 1000-1500 CA monomers, explaining the heterogeneity in capsid shape and size.(*21*)

Interestingly, a single CA protein is capable of wide conformational flexibility, as evident by the range of morphologies with varied curvature in expressed and purified capsids.(*24, 52*) Modeling studies of the HIV-1 capsid revealed that pentameric declination hinges upon flexibility in the CTD-CTD dimer interface,(*15*) whereas the NTD-CTD interface is believed to modulate variable lattice curvature in both hexamers and pentamers by limiting the movement of CA-CTD with respect to CA-NTD.(*42*) Nonetheless, the free energy penalty for sharp lattice curvature

formation, as seen in some of non-canonical capsid morphologies, can be prohibitive as evident from the ubiquitous appearance of CA cylinders (tubules) in the absence of IP6.

In our simulations, IP6 ions stabilize capsid intermediates, thereby enabling the formation of high curvature regions and relatedly promoting non-canonical core formation. We thus speculate that small molecules that target IP6 binding sites may preclude HIV-1 viral infectivity by inducing early capsid closure through rapid incorporation of pentamer defects or by over-stabilizing the capsid.(*38*) Considering the hierarchical nature of capsid assembly, our results underscore the importance of the IP6 binding site as a target of opportunity for the development of potential maturation inhibitors and can provide valuable insight into that effort. It is of note that the need for efficient and authentic assembly imparts considerable genetic fragility on HIV-1 CA.(*53*) However, the current CG model is derived from experimental phenomenology which is incapable of providing detailed chemically-specific insights or interactions and, therefore, is unable to explore CA sequence mutation related variation in assembly properties. Further simulations that probe the basis of assembly behavior of CA variants will likely require incorporation of all-atom statistics recapitulating the configuration space of protein mutation. Additionally, the CG interactions between the ring of R18 and IP6 can be further modified using, e.g., coarse-graining technique such as relative entropy minimization(*54*) to approximate the configurational dependence of the potential of mean force by parameters that accurately reproduce structural features at all-atom resolution. The recruitment of nucleotides into the positively charged cavity of R18 is also essential for reverse transcription. Since IP6 plays an important role in stabilizing the ring of R18, additional CG simulations that target the role of IP6 in nucleotide recruitment and import at the virion level will be valuable in an effort to understand the conformity of the capsid pore as a potential drug target.

Materials & Methods

CG model details. The NTD and CTD of a CA CG monomer were represented as independent elastic network models (ENMs) containing Cα atoms from CA α–helices, while a weak ENM maintained inter-domain connections. The CG NTD was generated from PDB 3H4E(*55*) using average NTD Cα positions as the location of CG particles. The potential energy of the bonds (U_{bond}) was given by:

$$U_{bond} = K(r - r_0)^2 \qquad \text{Eq. 1}$$

where r is the separation distance, and r_0 is the computed average distance (full data in Table S2).

To preserve the local helix structure, an elastic network connected CG particles within an α–helix with harmonic spring constant K = 10 kcal/mol/Å^2 while the relative positions of α–helices were preserved by connecting each CG particle with its nearest neighbor from every other helix with harmonic spring constant K = 10 kcal/mol/Å^2. The complete CA CG ENM network is summarized in Table S1. CA CTD was constructed from chain A of PDB 2KOD(*56*) using the same method and the two CTDs in the CA dimer were connected with an ENM (K = 10 kcal/mol/Å^2). The CA protein assumes an assembly competent (or active) and assembly incompetent (or inactive) state depending on its conformational flexibility controlled by the flexible linker region. Therefore, an additional weak ENM (K = 0.01 kcal/mol/Å^2) was employed to connect CA NTD and CTD. It should be noted that the intent of the model is to maintain the internal shape of the CA dimer to study its self-assembly while reproduction of its detailed structure is not intended here. We used an inert molecular crowding agent to represent the mass and excluded volume of Ficoll70.(*40*) Ficoll70 was modeled using two independent spheres of CG sites connected by ENMs (K = 10 kcal/mol/Å^2), each consisting of 42 CG beads and ≈ 5.1 nm diameter. Ficoll70 CG beads had a default excluded volume separation of 1 nm, an effective radii of 0.5 nm. To capture the molecular shape of CA, CG monomers were overlapped with PDB structures 3H4E(*55*) and 3P0A(*15*) with excluded volume separations adjusted to maintain the relative separation between CG particles. A soft cosine excluded volume potential with an effective radius was used to prevent unphysical overlap between CG beads:

$$\mathrm{U}_{excl}(r_{jk}) = A\left[1 + \cos\left(\frac{\pi r_{jk}}{r_{cut}}\right)\right], r_{jk} < r_{cut} \qquad \text{Eq. 2}$$

Here, r_{jk} is the distance between CG particle centers with $r_{cut} = 20\text{Å}$ the onset of excluded volume repulsion for specific CG sites *j* and *k*. The value of *A* is 10 kcal mol^{-1}Å^{-1} in all cases.

The CG model also incorporated well-conserved and important CA protein/protein interfaces that were previously determined by centroid analysis of experimental structures 3H4E and 3P0A.(*15, 55*) “Ghost” CG beads were embedded into the CA CG structure at these selected centroid locations (residue 51 and 57 to the NTD, residue 63 and 204 to the CTD)) to function as binding regions in assembly and interacted only with non-ghost CG particles of adjacent monomers. The interaction between the “ghost/non-ghost” particles were represented by attractive interactions which can be enabled and disabled, generating assembly competent ($[CA]_+$) and assembly incompetent ($[CA]_-$) CA populations. The binding potential was given by a double-Gaussian:

$$\mathrm{U}_{bind}(r_{jk}) = \left[-A\exp\left[-Br_{jk}{}^2\right] + C\exp\left[-Dr_{jk}{}^2\right]\right], r_{jk} < r_{cut} \qquad \text{Eq. 3}$$

Where r_{jk} is the separation distance between CG particles *j* and *k* with $r_{cut} = 30$ Å. The parameters $A = 1.1$ kcal mol^{-1}, $B = 0.1$ Å^{-2}, $C = 2.0$ kcal mol^{-1}, and $D = 0.01$ Å^{-2} describe the depth and width of the long-range and short-range Gaussian interactions, respectively. Full details of the CG models for CA and the inert crowding agent are described in ref 36. Aditionally, we used virion relevant assembly conditions and an explicit IP6 to produce mature-style capsids over the simulation length examined. We generated a single site CG model for IP6 using a linear center-of-mass mapping. Interactions between CA and IP6 were represented by an attractive interaction, with the locations of energy minima parameterized using the CG model of the full capsid. The CA self-assembly process is sensitive to the specifics of the local environment. In addition to that, accurate representation of the weak binding effects between proteins and the dynamics of IP6 binding with the growing lattice in virion-relevant conditions renders CG parameterization challenging. Therefore, consistent with the CG CA model, we modeled the attractive interaction between the R18 residue of CA and IP6 using a Gaussian interaction.

$$\mathrm{U}_{attr}(r_{jk}) = -\frac{H}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(r_{jk}-r_0)^2}{2\sigma^2}\right] \qquad \text{Eq. 4}$$

Here $r_0 = 8.3$ Å is the location of the energy minima determined from the average distance between the vertex and the centroid of the R18 ring within hexamers and pentamers of a pre-assembled CG capsid (Fig. S2). The parameter *H* = 2.8 kcal Å mol^{-1} and $\sigma = 0.4$ Å describe the depth and width of the Gaussian interaction, respectively, and were determined by extensive CGMD simulations of IP6 binding to a pre-assembled capsid (Fig. S2). For simplicity, we used parameters to identify the minimum required CG binding energy for at least 50% IP6 occupancy within R18 rings throughout the pre-assembled capsid after 2×10^8 CGMD timesteps.. The mass of each CG particle in CA, inert crowder and IP6 were set to 10 Da, 22.6 Da and 33 Da (≈ 20 fold reduction) respectively to access longer time-scales in the CGMD simulations. All simulations were prepared using Moltemplate.(*57*)

CGMD simulations. For the CA self-assembly simulations, 1232 monomers, 1700 inert crowders were randomly distributed throughout a 80×80×80 nm^3 cubic box with periodic boundaries; four replicas were simulated. Another distinct system using a 90×90×90 nm^3 cubic box with periodic boundaries and with 1760 CA and 2496 inert crowders was prepared for comparing the assembly rates of CA; five replicas were simulated. For the CA and IP6 co-assembly simulations, a random distribution of 1760 CA, 600 IP6 molecules and 2496 inert

crowders were prepared in a 90×90×90 nm^3 cubic box with periodic boundaries; six replicas were simulated. All CGMD simulations were performed using LAMMPS 21Jul2020.(*58*) All simulations were performed in a constant NVT ensemble using a Langevin thermostat(*59*) at 300 K with a damping period of 100 ps. The correlation times of solution-state CA inter-domain motion is 2-5 ns. With CG timestep τ = 10 fs, 5×10^5 τ corresponds to 5 ns CG time. Therefore, $[CA]_+/[CA]_-$ the UCG state switching was attempted every 5×10^5 τ For all simulations, a swarm of 50 parallel independent trajectories were propagated for 10^7 τ initially to expedite the identification of a nucleated cluster. The trajectory yielding a nucleated cluster (size < 12 CA monomers) was used to initialize the production runs. Simulation snapshots were saved every 10^7 τ.

Data analysis and visualization. Graph analysis of assembled lattices was performed using the Python package NetworkX 2.1 (http://networkx.github.io/) using two proximity criteria. Inter-hexameric contacts were defined by the CTD dimer interface, i.e., when the distance between CG types 102/102 were less than 2.5 nm. Intra-hexameric contacts were defined by the distance between CA R18, i.e., when CG types 2/2 were less than 1.4 nm. Time series profiles of extracted clusters were generated using Python package Matplotlib.(*60*) Visualization of extracted clusters were generated using VMD 1.9.3.(*61*)

References

1. W. I. Sundquist, H. G. Kräusslich, HIV-1 assembly, budding, and maturation. *Cold Spring Harb Perspect Med* **2**, a006924 (2012).
2. B. K. Ganser-Pornillos, M. Yeager, W. I. Sundquist, The structural biology of HIV assembly. *Curr. Opin. Struct. Biol.* **18**, 203-217 (2008).
3. E. O. Freed, HIV-1 assembly, release and maturation. *Nat. Rev. Microbiol.* **13**, 484-496 (2015).
4. 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).
5. J. A. Briggs, H. G. Kräusslich, The molecular architecture of HIV. *J. Mol. Biol.* **410**, 491-500 (2011).
6. V. B. Shah, J. Shi, D. R. Hout, I. Oztop, L. Krishnan, J. Ahn, M. S. Shotwell, A. Engelman, C. Aiken, The host proteins transportin SR2/TNPO3 and cyclophilin A exert opposing effects on HIV-1 uncoating. *J. Virol.* **87**, 422-432 (2013).
7. E. Toccafondi, D. Lener, M. Negroni, HIV-1 Capsid Core: A Bullet to the Heart of the Target Cell. *Front. Microbiol.* **12**, (2021).
8. C. Liu, J. R. Perilla, J. Ning, M. Lu, G. Hou, R. Ramalho, B. A. Himes, G. Zhao, G. J. Bedwell, I.-J. Byeon, J. Ahn, A. M. Gronenborn, P. E. Prevelige, I. Rousso, C. Aiken, T. Polenova, K.

Schulten, P. Zhang, Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site. *Nat. Commun.* **7**, 10714 (2016).

9. A. J. Price, A. J. Fletcher, T. Schaller, T. Elliott, K. Lee, V. N. KewalRamani, J. W. Chin, G. J. Towers, L. C. James, CPSF6 Defines a Conserved Capsid Interface that Modulates HIV-1 Replication. *PLOS Pathog.* **8**, e1002896 (2012).
10. T. Dorfman, A. Bukovsky, A. Ohagen, S. Höglund, H. G. Göttlinger, Functional domains of the capsid protein of human immunodeficiency virus type 1. *J. Virol.* **68**, 8180-8187 (1994).
11. F. Mammano, A. Ohagen, S. Höglund, H. G. Göttlinger, Role of the major homology region of human immunodeficiency virus type 1 in virion morphogenesis. *J. Virol.* **68**, 4927-4936 (1994).
12. U. K. von Schwedler, K. M. Stray, J. E. Garrus, W. I. Sundquist, Functional surfaces of the human immunodeficiency virus type 1 capsid protein. *J. Virol.* **77**, 5439-5450 (2003).
13. J. R. Perilla, A. M. Gronenborn, Molecular Architecture of the Retroviral Capsid. *Trends Biochem. Sci.* **41**, 410-420 (2016).
14. L. Deshmukh, C. D. Schwieters, A. Grishaev, R. Ghirlando, J. L. Baber, G. M. Clore, Structure and Dynamics of Full-Length HIV-1 Capsid Protein in Solution. *J.Am.Chem.Soc.* **135**, 16133-16147 (2013).
15. O. Pornillos, B. K. Ganser-Pornillos, M. Yeager, Atomic-level modelling of the HIV capsid. *Nature* **469**, 424-427 (2011).
16. J. A. Briggs, J. D. Riches, B. Glass, V. Bartonova, G. Zanetti, H. G. Krausslich, Structure and assembly of immature HIV. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11090-11095 (2009).
17. S. D. Fuller, T. Wilk, B. E. Gowen, H. G. Kräusslich, V. M. Vogt, Cryo-electron microscopy reveals ordered domains in the immature HIV-1 particle. *Curr. Biol.* **7**, 729-738 (1997).
18. L. A. Carlson, J. A. Briggs, B. Glass, J. D. Riches, M. N. Simon, M. C. Johnson, B. Muller, K. Grunewald, H. G. Krausslich, Three-dimensional analysis of budding sites and released virus suggests a revised model for HIV-1 morphogenesis. *Cell Host Microbe* **4**, 592-599 (2008).
19. J. A. Briggs, M. N. Simon, I. Gross, H. G. Krausslich, S. D. Fuller, V. M. Vogt, M. C. Johnson, The stoichiometry of Gag protein in HIV-1. *Nat. Struct. Mol. Biol.* **11**, 672-675 (2004).
20. I. Kucharska, P. Ding, K. K. Zadrozny, R. A. Dick, M. F. Summers, B. K. Ganser-Pornillos, O. Pornillos, Biochemical Reconstitution of HIV-1 Assembly and Maturation. *J. Virol.* **94**, (2020).
21. L. S. Ehrlich, B. E. Agresta, C. A. Carter, Assembly of recombinant human immunodeficiency virus type 1 capsid protein in vitro. *J. Virol.* **66**, 4874-4883 (1992).
22. B. K. Ganser-Pornillos, U. K. von Schwedler, K. M. Stray, C. Aiken, W. I. Sundquist, Assembly properties of the human immunodeficiency virus type 1 CA protein. *J. Virol.* **78**, 2545-2552 (2004).
23. G. Zhao, J. R. Perilla, E. L. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A. M. Gronenborn, K. Schulten, C. Aiken, P. Zhang, Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **497**, 643-646 (2013).
24. S. Li, C. P. Hill, W. I. Sundquist, J. T. Finch, Image reconstructions of helical assemblies of the HIV-1 CA protein. *Nature* **407**, 409-413 (2000).

25. 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).
26. J. A. Briggs, T. Wilk, R. Welker, H. G. Kräusslich, S. D. Fuller, Structural organization of authentic, mature HIV-1 virions and cores. *Embo. j.* **22**, 1707-1715 (2003).
27. D. A. Jacques, W. A. McEwan, L. Hilditch, A. J. Price, G. J. Towers, L. C. James, HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis. *Nature* **536**, 349-353 (2016).
28. R. A. Dick, D. L. Mallery, V. M. Vogt, L. C. James, IP6 Regulation of HIV Capsid Assembly, Stability, and Uncoating. *Viruses* **10**, (2018).
29. D. L. Mallery, K. M. R. Faysal, A. Kleinpeter, M. S. C. Wilson, M. Vaysburd, A. J. Fletcher, M. Novikova, T. Böcking, E. O. Freed, A. Saiardi, L. C. James, Cellular IP6 Levels Limit HIV Production while Viruses that Cannot Efficiently Package IP6 Are Attenuated for Infection and Replication. *Cell Rep.* **29**, 3983-3996.e3984 (2019).
30. S. Campbell, R. J. Fisher, E. M. Towler, S. Fox, H. J. Issaq, T. Wolfe, L. R. Phillips, A. Rein, Modulation of HIV-like particle assembly in vitro by inositol phosphates. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10875-10879 (2001).
31. R. A. Dick, K. K. Zadrozny, C. Xu, F. K. M. Schur, T. D. Lyddon, C. L. Ricana, J. M. Wagner, J. R. Perilla, B. K. Ganser-Pornillos, M. C. Johnson, O. Pornillos, V. M. Vogt, Inositol phosphates are assembly co-factors for HIV-1. *Nature* **560**, 509-512 (2018).
32. 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**, (2020).
33. J. R. Perilla, K. Schulten, Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations. *Nat. Commun.* **8**, 15959 (2017).
34. A. Yu, E. M. Y. Lee, J. A. G. Briggs, B. K. Ganser-Pornillos, O. Pornillos, G. A. Voth, Strain and rupture of HIV-1 capsids during uncoating. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2117781119 (2022).
35. A. J. Pak, G. A. Voth, Advances in coarse-grained modeling of macromolecular complexes. *Curr. Opin. Struct. Biol.* **52**, 119-126 (2018).
36. J. M. A. Grime, J. F. Dama, B. K. Ganser-Pornillos, C. L. Woodward, G. J. Jensen, M. Yeager, G. A. Voth, Coarse-grained simulation reveals key features of HIV-1 capsid self-assembly. *Nat. Commun.* **7**, 11568 (2016).
37. J. M. Grime, G. A. Voth, Early stages of the HIV-1 capsid protein lattice formation. *Biophys. J.* **103**, 1774-1783 (2012).
38. A. J. Pak, J. M. A. Grime, A. Yu, G. A. Voth, Off-Pathway Assembly: A Broad-Spectrum Mechanism of Action for Drugs That Undermine Controlled HIV-1 Viral Capsid Formation. *J.Am.Chem.Soc.* **141**, 10214-10224 (2019).
39. D. L. Mallery, A. B. Kleinpeter, N. Renner, K. M. R. Faysal, M. Novikova, L. Kiss, M. S. C. Wilson, B. Ahsan, Z. Ke, J. A. G. Briggs, A. Saiardi, T. Böcking, E. O. Freed, L. C. James, A stable immature lattice packages IP(6) for HIV capsid maturation. *Sci. adv.* **7**, eabe4716 (2021).
40. 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).

41. Z. Yu, M. J. Dobro, C. L. Woodward, A. Levandovsky, C. M. Danielson, V. Sandrin, J. Shi, C. Aiken, R. Zandi, T. J. Hope, G. J. Jensen, Unclosed HIV-1 capsids suggest a curled sheet model of assembly. *J. Mol. Biol.* **425**, 112-123 (2013).
42. B. K. Ganser-Pornillos, M. Yeager, O. Pornillos, Assembly and architecture of HIV. *Adv. Exp. Med. Biol.* **726**, 441-465 (2012).
43. M. Grättinger, H. Hohenberg, D. Thomas, T. Wilk, B. Müller, H. G. Kräusslich, In vitro assembly properties of wild-type and cyclophilin-binding defective human immunodeficiency virus capsid proteins in the presence and absence of cyclophilin A. *Virology* **257**, 247-260 (1999).
44. S. Mattei, A. Flemming, M. Anders-Össwein, H.-G. Kräusslich, J. A. G. Briggs, B. Müller, W. I. Sundquist, RNA and Nucleocapsid Are Dispensable for Mature HIV-1 Capsid Assembly. *J. Virol.* **89**, 9739-9747 (2015).
45. P. W. Keller, R. K. Huang, M. R. England, K. Waki, N. Cheng, J. B. Heymann, R. C. Craven, E. O. Freed, A. C. Steven, A two-pronged structural analysis of retroviral maturation indicates that core formation proceeds by a disassembly-reassembly pathway rather than a displacive transition. *J. Virol.* **87**, 13655-13664 (2013).
46. G. A. Frank, K. Narayan, J. W. Bess, G. Q. Del Prete, X. Wu, A. Moran, L. M. Hartnell, L. A. Earl, J. D. Lifson, S. Subramaniam, Maturation of the HIV-1 core by a non-diffusional phase transition. *Nat. Commun.* **6**, 5854 (2015).
47. X. Meng, G. Zhao, E. Yufenyuy, D. Ke, J. Ning, M. DeLucia, J. Ahn, A. M. Gronenborn, C. Aiken, P. Zhang, Protease Cleavage Leads to Formation of Mature Trimer Interface in HIV-1 Capsid. *PLOS Pathog.* **8**, e1002886 (2012).
48. T. A. Bharat, N. E. Davey, P. Ulbrich, J. D. Riches, A. de Marco, M. Rumlova, C. Sachse, T. Ruml, J. A. Briggs, Structure of the immature retroviral capsid at 8 Å resolution by cryo-electron microscopy. *Nature* **487**, 385-389 (2012).
49. J. Ning, G. Erdemci-Tandogan, E. L. Yufenyuy, J. Wagner, B. A. Himes, G. Zhao, C. Aiken, R. Zandi, P. Zhang, In vitro protease cleavage and computer simulations reveal the HIV-1 capsid maturation pathway. *Nat. Commun.* **7**, 13689 (2016).
50. C. L. Woodward, S. N. Cheng, G. J. Jensen, Electron cryotomography studies of maturing HIV-1 particles reveal the assembly pathway of the viral core. *J. virol.* **89**, 1267-1277 (2015).
51. J. Liu, F. Sadre-Marandi, S. Tavener, C. Chen, Curvature Concentrations on the HIV-1 Capsid. *Mol. Based Math. Biol.* **3**, (2015).
52. T. Thames, A. J. Bryer, X. Qiao, J. Jeon, R. Weed, K. Janicki, B. Hu, P. L. Gor'kov, I. Hung, Z. Gan, J. R. Perilla, B. Chen, Curvature of the Retroviral Capsid Assembly Is Modulated by a Molecular Switch. *J. Phys. Chem. Lett.* **12**, 7768-7776 (2021).
53. S. J. Rihn, S. J. Wilson, N. J. Loman, M. Alim, S. E. Bakker, D. Bhella, R. J. Gifford, F. J. Rixon, P. D. Bieniasz, Extreme Genetic Fragility of the HIV-1 Capsid. *PLOS Pathog.* **9**, e1003461 (2013).
54. M. S. Shell, The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *Chem. Phys.* **129**, 144108 (2008).
55. O. Pornillos, B. K. Ganser-Pornillos, B. N. Kelly, Y. Hua, F. G. Whitby, C. D. Stout, W. I. Sundquist, C. P. Hill, M. Yeager, X-ray structures of the hexameric building block of the HIV capsid. *Cell* **137**, 1282-1292 (2009).

56. I. J. Byeon, X. Meng, J. Jung, G. Zhao, R. Yang, J. Ahn, J. Shi, J. Concel, C. Aiken, P. Zhang, A. M. Gronenborn, Structural convergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. *Cell* **139**, 780-790 (2009).
57. A. I. Jewett, Z. Zhuang, J.-E. Shea, Moltemplate a Coarse-Grained Model Assembly Tool. *Biophys. J.* **104**, 169a (2013).
58. S. Plimpton, Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **117**, 1-19 (1995).
59. E. Vanden-Eijnden, G. Ciccotti, Second-order integrators for Langevin equations with holonomic constraints. *Chem. Phys. Lett.* **429**, 310-316 (2006).
60. J. D. Hunter, Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90-95 (2007).
61. W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33-38, 27-38 (1996).
62. D. Stanzione, J. West, R. T. Evans, T. Minyard, O. Ghattas, D. K. Panda, "Frontera: The Evolution of Leadership Computing at the National Science Foundation" in *Practice and Experience in Advanced Research Computing* (Association for Computing Machinery, 2020), pp. 106–111.
63. J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. R. Scott, N. Wilkins-Diehr, XSEDE: Accelerating Scientific Discovery. *Comput. Sci. Eng.* **16**, 62-74 (2014).

Acknowledgements

The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC resources that have contributed to the research results reported within this paper.(*62*) This research also used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation grant number ACI-1548562.(*63*)

Funding

This research was supported by the National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health under grant R01 AI154092. A.J.P gratefully acknowledges support from the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under grant F32 AI150477.

Author contributions: M.G., A.J.P and G.A.V. designed research. M.G. and A.J.P performed research. M.G. and A.J.P contributed methods, simulation code, and analytic tools. M.G., A.J.P., and G.A.V. analyzed data. M.G., A.J.P., and G.A.V. wrote the paper.

Competing interests: None

Data and materials availability:

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All relevant data are available from the authors on request.