

Inositol Hexakisphosphate (IP6) Accelerates Immature HIV-1 Gag Protein Assembly toward Kinetically Trapped Morphologies

Alexander J. Pak, Manish Gupta, Mark Yeager, and Gregory A. Voth*

Cite This: *J. Am. Chem. Soc.* 2022, 144, 10417–10428

Read Online

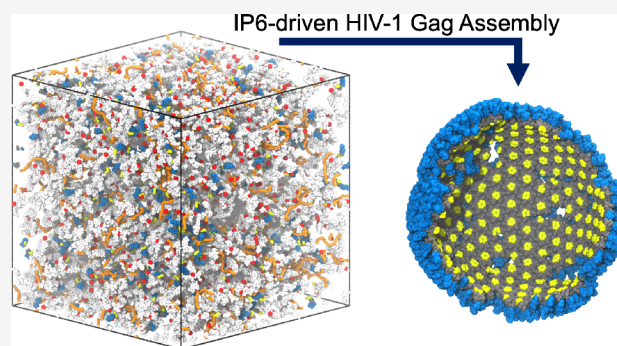
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: During the late stages of the HIV-1 lifecycle, immature virions are produced by the concerted activity of Gag polyproteins, primarily mediated by the capsid (CA) and spacer peptide 1 (SP1) domains, which assemble into a spherical lattice, package viral genomic RNA, and deform the plasma membrane. Recently, inositol hexakisphosphate (IP6) has been identified as an essential assembly cofactor that efficiently produces both immature virions *in vivo* and immature virus-like particles *in vitro*. To date, however, several distinct mechanistic roles for IP6 have been proposed on the basis of independent functional, structural, and kinetic studies. In this work, we investigate the molecular influence of IP6 on the structural outcomes and dynamics of CA/SP1 assembly using coarse-grained (CG) molecular dynamics (MD) simulations and free energy calculations. Here, we derive a bottom-up, low-resolution, and implicit-solvent CG model of CA/SP1 and IP6, and simulate their assembly under conditions that emulate both *in vitro* and *in vivo* systems. Our analysis identifies IP6 as an assembly accelerant that promotes curvature generation and fissure-like defects throughout the lattice. Our findings suggest that IP6 induces kinetically trapped immature morphologies, which may be physiologically important for later stages of viral morphogenesis and potentially useful for virus-like particle technologies.



INTRODUCTION

During the late stages of human immunodeficiency virus type 1 (HIV-1) replication, an immature particle is assembled at the plasma membrane interface and released,^{1–3} this latter step is aided by the endosomal sorting complex required for transport (ESCRT) machinery.^{1,4} The capsid (CA) and spacer peptide 1 (SP1) domains of the structural Gag polyprotein are responsible for coordinating Gag oligomerization into the immature lattice,^{5–7} which is both hexagonal and spherical (and therefore, incomplete).^{7–10} The matrix (MA) and nucleocapsid (NC) domains mediate interactions with the plasma membrane and viral genomic RNA, respectively.^{1,11–13} The molecular mechanisms that regulate Gag assembly are potential targets for antiretroviral therapies, as similarly demonstrated by capsid inhibitor drugs that perturb capsid assembly after maturation.^{14–16}

The CA domain of Gag consists of two globular domains: the N-terminal (CA_{NTD}) domain consists of 7 α -helices (helix 1–7), and the C-terminal (CA_{CTD}) domain consists of 4 α -helices (helix 8–11).^{17–19} On the basis of atomistic structures resolved at 3.9 and 3.27 Å resolutions, using cryo-electron tomography (cryo-ET)²⁰ and X-ray crystallography,²¹ respectively, several critical inter- and intrahexameric contacts have been identified. A trimeric contact formed by helix 2 stabilizes interhexameric interactions between CA_{NTD} while a dimeric contact formed by helix 1 stabilizes intrahexameric interactions

between CA_{NTD}.^{20,22} Interhexameric interactions are further mediated by the CA_{CTD} dimeric contact formed by helix 9, which is partially preserved in the mature CA configuration.^{22–25} Within the CA_{CTD} domain, the major homology region (MHR; residues I285–L304) loop and GVGG (residues G352–V355) β -turn are two structural motifs that contain important interactions, likely required to nucleate α -helical folding of the CA_{CTD}/SP1 junction (residues P356–T371).²¹ This helical junction forms a six-helix bundle (6HB) that is stabilized by hydrophobic knobs-in-holes interactions and represents an important intrahexameric interaction that is minimally necessary for immature assembly.^{21,26} Mutations within the aforementioned hexameric contacts have been shown to abrogate immature assembly.^{6,21,26,27}

Beyond the protein–protein interactions described above, immature particle assembly requires other cellular constituents. For instance, viral RNA and, to a lesser extent, generic oligonucleotides serve as catalysts for protein multimerization

Received: March 8, 2022

Published: June 6, 2022



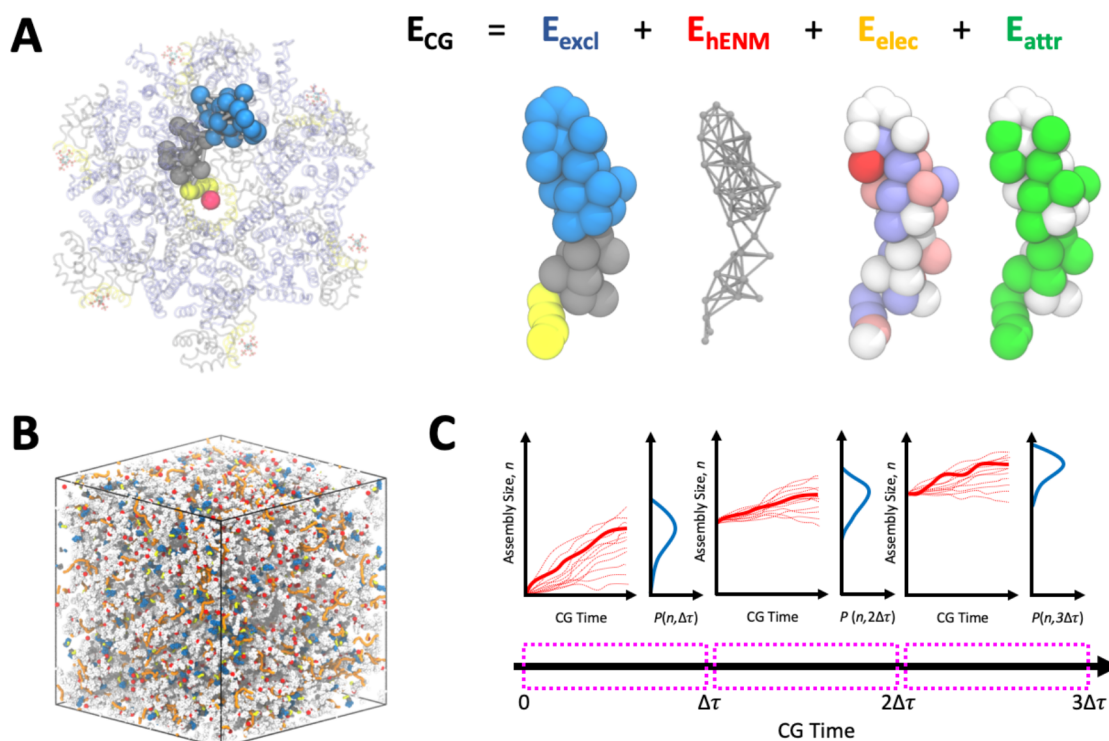


Figure 1. Schematic of coarse-grained (CG) HIV-1 CA/SP1 model and CGMD simulation. (A) Overlay of a CG CA/SP1 monomer (solid blue, gray, and yellow spheres) and IP6 monomer (red sphere) over an AA CA/SP1 18-mer (transparent blue, gray, and yellow tubes) bound to 7 IP6 molecules (cyan, purple, and red sticks); blue, gray, and yellow represent the NTD, CTD, and SP1 domains, respectively. The additive interaction elements of the CG Hamiltonian are depicted to the right; excluded volume (E_{excl}), heteroelastic network model (E_{hENM}), screened Yukawa electrostatics (E_{elec}), and protein–protein associative contacts (E_{attr}). (B) A representative snapshot of an initial simulation configuration; active CA/SP1 (blue, gray, and yellow beads), inactive CA/SP1 (white beads), and oligonucleotide (orange beads) molecules are randomly distributed. (C) Schematic of the parallel sampling technique used to propagate assembly during CGMD simulations. For each CGMD simulation of length $\Delta\tau$, multiple CGMD trajectories were spawned (thin red lines) from the previously selected trajectory. From the probability distribution of resultant assembly sizes ($P(n)$), a trajectory that corresponded to the maximum $P(n)$ was selected to spawn the next generation of trajectories. All selected trajectories (thick red lines) were appended to create the final trajectory.

by promoting colocalization;^{11,28–30} in the former case, this is primarily mediated by interactions between the NC domain and the RNA Ψ packaging signal.^{31,32} The plasma membrane serves as a scaffold to promote protein assembly through dimensional reduction,^{11,33,34} which is directed by myristoyl insertion and interactions between the MA domain and phosphatidylinositol 4,5-bisphosphate (PIP₂) lipids,^{35–37} and ultimately evolves into the lipid envelope of the released viral particle. *In vitro* assembly studies have previously identified short oligonucleotides and inositol hexakisphosphate (IP6) as two constituents minimally necessary for immature virus-like particle (VLPs) assembly.^{30,38,39} In mammalian cells, IP6 is present in the cytoplasm at 10–40 μM concentrations.^{40,41} Recently, combined biochemical and structural studies have shown that IP6 is an essential cofactor that promotes immature particle production around a stoichiometric ratio of one IP6 molecule per immature hexamer.^{42,43} Within the immature hexamer, the negatively charged IP6 molecule binds between two six-membered lysine rings, K290 and K359, which line the interior pore of the hexamer and are positioned directly above the 6HB.⁴²

While it is evident that the presence of IP6 is essential, the precise role of IP6 during immature virus assembly remains unclear. For example, is the role of IP6 to stabilize the 6HB by neutralizing the two lysine rings along the central pore? Previous *in vitro* studies have shown that CA/SP1 morphology is sensitive to solution pH or salt (e.g., NaCl or KCl)

concentrations, which modulate electrostatic interactions,^{42,44} under high salt or low pH conditions, CA/SP1 proteins assemble into mature tubes while, under low salt or high pH conditions, these same proteins assemble into spherical VLPs.^{42,44,45} Intriguingly, alanine mutations of the lysine rings tended to increase *in vitro* VLP formation⁴⁶ while reducing *in vivo* particle production,⁴³ both with respect to wild-type (WT) virus. Alternatively, is the role of IP6 to influence assembly kinetics? For example, a recent study by Kucharska et al. suggests that Gag encodes all of the necessary interactions for assembly and that IP6 and viral RNA act as assembly rate modulators.⁴⁷ Finally, it is possible that IP6 serves both roles in a synergistic fashion.

In this work, we use coarse-grained (CG) molecular dynamics (MD) simulations to investigate the molecular influence of IP6 on CA/SP1 assembly. To do so, we derive a bottom-up, low-resolution, and implicit-solvent CG model of CA/SP1 and IP6 and implement a MD trajectory-based enhanced sampling approach. By quantifying 3D and 2D assembly behavior under three different buffer conditions (150 mM monovalent salt, 150 mM salt + IP6, and 400 mM salt), as well as free energy calculations, we show that IP6 primarily acts as an assembly accelerant. Immature lattices formed through the influence of IP6 assemble at faster rates compared to without, thereby promoting curvature generation and fissure-like defects throughout the lattice. Our analysis suggests that IP6 induces kinetically trapped immature morphologies, which

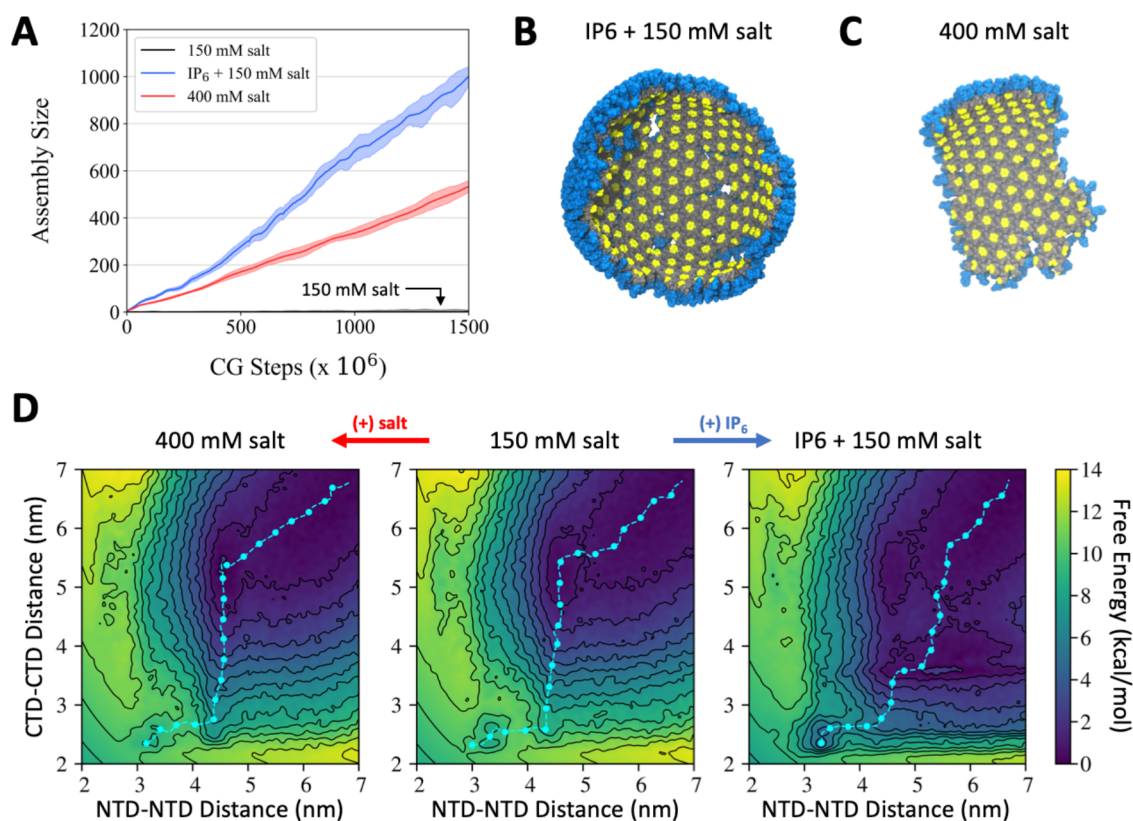


Figure 2. CGMD simulations of “*in vitro*” HIV-1 CA/SP1 assembly. (A) Time-series profiles of assembled protein cluster size as a function of CG time step under varying salt and cofactor conditions. (B) Snapshot of assembled protein under the presence of IP₆ and 150 mM salt; blue, gray, and yellow balls represent the NTD, CTD, and SP1 domains of CA/SP1 while IP₆ and RNA are not depicted for clarity. (C) Snapshot of assembled protein under the presence of 400 mM salt; same color scheme as (B). (D) Comparison of free energy profiles for the association of two CA/SP1 monomers projected along their interprotein CTD and NTD distances under varying salt and cofactor conditions. The cyan points and line trace the minimum free energy path.

furthermore, may be physiologically important for later stages of viral morphogenesis or leveraged for efficient production of VLPs.

RESULTS

Coarse-Grained Modeling and Simulation. We first derived “bottom-up” implicit-solvent CG models for CA/SP1 (35 sites) and for IP₆ (1 site), which were parametrized from AAMD trajectories (depicted in Figure 1A). The two reference AAMD simulations were of a CA/SP1 18-mer, i.e. a hexamer surrounded by six partial hexamers, and the same 18-mer complexed with seven IP₆ molecules positioned between the K290 and K359 rings in each hexamer pore; the first trajectory was used to derive the CA/SP1 CG model while the second trajectory was used to derive CG interactions between CA/SP1 and IP₆. The CG Hamiltonian consists of four additive terms (see Figure 1A) that respectively represent the excluded volume, flexibility, electrostatics, and short-range attraction of each molecule. Electrostatics are represented by a screened Yukawa potential; monovalent salt concentrations are implicitly modeled by adjusting the Debye length. Complete details on the mapping and parametrization procedure are provided in Methods.

Using the aforementioned CG models, we performed two types of CGMD assembly simulations. We first simulated CA/SP1 assembly under conditions that emulate prior *in vitro* studies. Here, we included CA/SP1 with 50-nt oligonucleotide (represented as a linear polymer) randomly dispersed within a

cubic box (see Figure 1B) at (i) 150 mM monovalent salt, (ii) 400 mM monovalent salt, and (iii) 150 mM monovalent salt and IP₆ conditions. We next simulated CA/SP1 under simplified *in vivo* conditions, i.e., at a plasma membrane interface. We used the same salt and IP₆ conditions and additionally included a 9000-nt oligonucleotide and a CG lipid membrane using a generic model with 30 $k_B T$ bending rigidity.

For both sets of simulations, we implemented a constant protein concentration algorithm and a path sampling algorithm. In the former, we intermittently identified CA/SP1 that have oligomerized with a hexameric CA/SP1 seed, which served as a nucleator. All remaining CA/SP1 were randomly partitioned into “active” and “inactive” monomers, similar to the ultra-coarse-grained model strategy used in prior mature CA assembly simulations.^{16,48} CG monomers in the “active” state maintained their short-range attractions while those in the “inactive” state had no short-range attractions; the monomer states were stochastically adjusted throughout the simulations such that the concentration of free “active” monomer was maintained. Assuming an excess of initial CA/SP1 is provided, this strategy ensures that a constant concentration of CA/SP1 is available for assembly. To efficiently sample CA/SP1 assembly into larger aggregates, we discretized the assembly process into sequential segments of length $\Delta\tau_{CG}$ CG time steps (see Figure 1C). For each segment, 10 parallel CGMD simulations following randomized Langevin dynamics were propagated. Cluster sizes were computed at the final configuration of each parallel run, and

the probability distribution of cluster sizes ($P(n)$) was approximated as a log-normal distribution. The trajectory yielding a cluster closest to the mean of $P(n)$ was used to initialize the next set of parallel trajectories; the selected trajectory from each chunk was concatenated to create the final assembly trajectory, a strategy conceptually similar to parallel cascade MD.⁴⁹ Depending upon the system conditions, we have found that CA/SP1 assembly ranges between strongly disfavored and strongly favored (see Figure S1 of Supporting Information). When CA/SP1 assembly is weakly favored, our strategy reduces the explicit number of necessary trajectories by extracting an effective trajectory that represents the mean kinetics predicted by the underlying CG model. Complete details on the CGMD simulations are provided in Methods while additional discussion on the path sampling algorithm is presented in the Supporting Information.

To quantify free energies of monomer–monomer association, we performed 2D well-tempered metadynamics (WTMetaD) simulations.^{50,51} Each system contained two monomers that were biased against two collective variables (CVs): the distances between the two CA_{CTD} centroids and the two CA_{NTD} centroids; when included, IP6 was restrained to be proximal to the K290 and K359 residues of one monomer. To remove volumetric entropic effects, the free energies were shifted by the Jacobian correction factor $k_B T \ln R^2$.^{52,53} All reported free energy surfaces (FESs) were averaged over three independent replicas after alignment to the free energy at large distances. Additional details on the WTMetaD simulations are provided in Methods.

Assembly into Virus-like Particles. *In vitro* experiments have shown that the presence of 500 μM IP6 induces CA/SP1 (also 500 μM) assembly into spherical VLPs at pH 6 and physiological salt conditions, while the absence of IP6 induces amorphous protein aggregates. Under high salt conditions, CA/SP1 instead assembles into tubules. Mutating K290 and K359 to alanine was previously shown to yield VLPs, suggesting that electrostatic repulsion of the lysine ring inhibits assembly, which may be recovered by salt-induced electrostatic screening. Here, we use CGMD simulations to investigate the differences between IP6-induced and salt-screening-induced CA/SP1 assembly under similar *in vitro* conditions.

We simulated 850 μM CA/SP1 assembly under three different conditions: (i) 150 mM monovalent salt (i.e., low salt), (ii) 400 mM monovalent salt (i.e., high salt), and (iii) 150 mM monovalent salt with 425 μM IP6. In all three cases, 50-nt RNA-representing polymers were also added to the system. The CG oligonucleotides represent nonspecific oligonucleotides that were shown to be necessary for productive assembly in prior experiments.^{42,47} Similarly, our CG models fail to undergo assembly under these three conditions in the absence of oligonucleotides. Our results are summarized in Figure 2. The time-series profiles in Figure 2A show that while CA/SP1 fails to assemble at low-salt conditions, both high-salt and low-salt with IP6 conditions lead to productive protein assembly. We note that, under our simulated salt and IP6 concentrations, the presence of IP6 (elevated salt) yields a lattice consisting of 999 ± 89 (534 ± 27) CA/SP1 after $1500 \times 10^6 \tau_{\text{CG}}$, suggesting that IP6 accelerates CA/SP1 by nearly 2-fold compared to high-salt conditions. Beyond accelerated kinetics, the presence of IP6 appears to shift the morphological outcome of CA/SP1 assembly. As depicted in Figure 2B, CA/SP1 assembles into a spherical VLP under low-salt with IP6 conditions. Here, the

particle consists of fissure defects that are characteristic of immature HIV-1 virions as seen in prior cryo-ET lattice maps.^{3,8,54} Alternatively, CA/SP1 assembles into a curved and contiguous hexameric lattice under high-salt conditions, as evident from Figure 2C. Here, each protomer appears to maximize its interhexameric and intrahexameric contacts, suggesting that the assembled lattice is representative of an energetically minimized state. In contrast, the VLP formed under the presence of IP6 is likely a kinetically trapped state that results from enhanced assembly kinetics. Prior assembly simulations have shown that similar defects emerge when protomer association energetics are enhanced,¹¹ suggesting that IP6 interactions induce a comparable effect.

To quantify thermodynamic differences between protomer association under the aforementioned conditions, we performed WTMetaD simulations. We compare our computed FESs in Figure 2D, which are projected onto the distances between CTD-CTD and NTD-NTD centers-of-mass. The low-salt FES serves as a baseline to characterize how the underlying FES is altered under high-salt and low-salt with IP6 conditions; recall that 850 μM CA/SP1 does not spontaneously assemble under these conditions (see Figure 2A). All three cases exhibit qualitatively similar minimum free energy paths (MFEPs), in which association begins at the CTD-CTD interface, which is followed by additional association at both CTD-CTD and NTD-NTD interfaces. However, there are several key differences to note. For example, CTD-CTD association is facilitated under high-salt conditions, as demonstrated by the broadening of the FES toward smaller CTD-CTD distances when the NTD-NTD distance is below 4.5 nm. The shapes of the free energy minimum where the CTD-CTD (NTD-NTD) distance is around 2.5 (3.3) nm are also distinct. This local minimum is comparatively shallow in the absence of IP6, whereas the minimum is both deepened and broadened when IP6 is present. With IP6 present, the free energy barriers for CA/SP1 association and dissociation are 5.7 ± 0.6 and 1.5 ± 0.8 kcal/mol, respectively. The association barrier is notably lower than that of both low-salt (10.6 ± 0.1 kcal/mol) and high-salt (9.6 ± 0.2 kcal/mol) conditions, while the dissociation barriers (1.9 ± 0.3 and 0.8 ± 0.4 kcal/mol for low-salt and high-salt, respectively) are comparable across all three cases. In other words, rather than stabilizing the bound state, i.e., increasing the dissociation barrier, IP6 appears to promote CA/SP1 association by reducing the association barrier, in part by promoting colocalization up to CTD-CTD and NTD-NTD distances of 3.5 and 4.5 nm, respectively. We should note, however, that the above calculations only pertain to monomer–monomer association, while oligomer–monomer association may include additional collective effects. Nonetheless, these calculations serve as a minimal basis to identify and compare the impact of electrostatic screening and IP6 interactions on CA/SP1 assembly.

Immature Lattice Assembly at the Membrane. In our prior CGMD simulations of CA/SP1 assembly on a membrane, we had found that initial protein assembly first required a perturbation or “puncta” to the local membrane curvature¹¹ as the assembling immature Gag without it could not sufficiently drive membrane bending toward a budding event. Interestingly, prior experimental studies have shown that the number of virions released from cells is reduced by around 60–90% when IP6-producing enzymes are knocked out without loss of integrity to the produced virion.⁴³ We therefore hypothesize that the effect of IP6 on the assembling Gag may

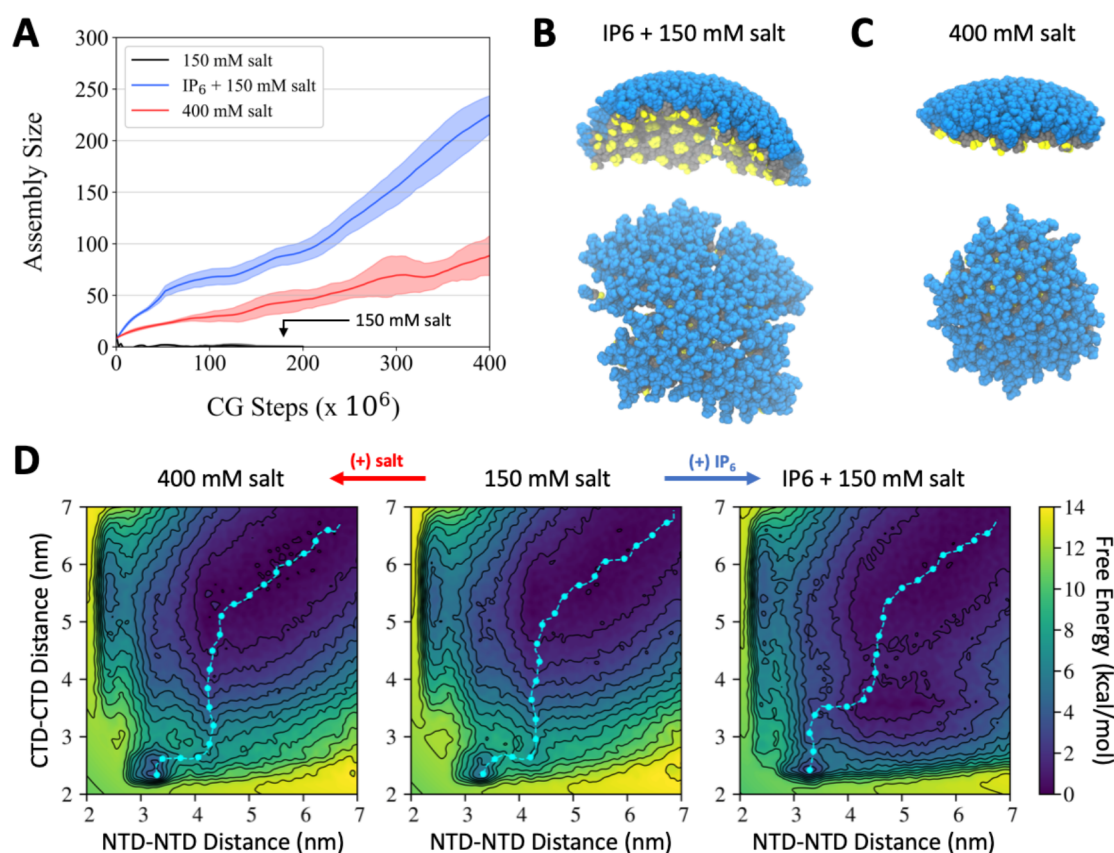


Figure 3. CGMD simulations of HIV-1 CA/SP1 assembly at a membrane interface. (A) Time-series profiles of assembled protein cluster size as a function of CG time step under varying salt and cofactor conditions. (B) Snapshot of assembled protein under the presence of IP₆ and 150 mM salt; blue, gray, and yellow balls represent the NTD, CTD, and SP1 domains of CA/SP1 while IP₆, RNA, and lipids are not depicted for clarity. (C) Snapshot of assembled protein under the presence of 400 mM salt; same color scheme as (B). (D) Comparison of free energy profiles for the association of two CA/SP1 monomers projected along their interprotein CTD and NTD distances under varying salt and cofactor conditions. The cyan points and line trace the minimum free energy path.

induce the necessary membrane curvature generation (and therefore, force) necessary for productive immature lattice assembly and budding.

We thus simulated CA/SP1 assembly with RNA at a lipid membrane interface (30 $k_B T$ bending rigidity) under the same three conditions as above: (i) 150 mM monovalent salt (i.e., low salt), (ii) 400 mM monovalent salt (i.e., high salt), and (iii) 150 mM monovalent salt with IP₆ (using a 1:1 CA/SP1:IP₆ stoichiometric ratio). Here, a 9000-nt RNA-representing polymer was added to the system to represent a nonspecific oligonucleotide with a length comparable to the viral genome. Similar to our prior CG study,¹¹ the inclusion of RNA was necessary to promote assembly. To reduce the lag time arising from the nucleation event, we ran short simulations until a small hexamer was formed, which we used to seed the rest of the lattice. We maintained a constant surface coverage of 2380 #/ μm^2 for CA/SP1 protein.

We summarize assembly time-series statistics in Figure 3A. At our simulated protein concentrations, we find that CA/SP1 is unable to assemble at 150 mM salt; as discussed later, assembly under these salt concentrations is possible with elevated protein expression levels. In contrast, both 400 mM salt and 150 mM salt with IP₆ conditions result in immature lattice growth. However, in the case of 150 mM salt with IP₆, it is notable that the lattice grows to one containing 220 ± 15 protomers after 400×10^6 CG time steps, while the 400 mM salt case grows to one containing 90 ± 20 protomers within

the same simulation time. In other words, the presence of IP₆ at 150 mM salt accelerates assembly nearly 2.5-fold compared to that of elevated salt at 400 mM concentrations (i.e., increased electrostatic screening). As seen in Figure 3B, accelerated assembly kinetics under the presence of IP₆ also appears to incorporate fissure-like defects throughout the lattice, which correspondingly results in greater curvature generation. In comparison, as seen in Figure 3C, the lattice assembled under 400 mM salt remains uniform, contiguous, and flatter in comparison to the IP₆-containing lattice. Although these assembly rates are slower than those computed for the *in vitro* system, we note that the two systems should not be directly compared due to differences in effective protein concentration and dimensional reduction.

In Figure 3D, we quantify the 2D FES for CA/SP1 association when diffusion is restricted to the lateral xy plane. Similar to our prior analysis under *in vitro* conditions, we investigate the changes to the FES under high-salt and low-salt with IP₆ conditions in comparison to low-salt conditions, under which extended assembly is not observed. Under high-salt conditions, the shape of the close contact free energy minimum located when the NTD-NTD and CTD-CTD distances are around 3.2 and 2.4 nm, respectively, is qualitatively similar to that of low-salt conditions. Interestingly, the free energy barriers for dissociation under low-salt and high-salt conditions are comparable (3.2 ± 0.6 and 4.1 ± 0.4 kcal/mol, respectively) while the barriers for association are

slightly lowered in the latter case (7.9 ± 0.4 and 7.1 ± 0.3 kcal/mol, respectively, with $p = 0.04$ according to the unpaired t test). Although the position of the free energy minimum is conserved under low-salt with IP6 conditions, the shape of this minimum is widened to favor an alternative CTD-CTD association path, while the barriers for dissociation and association are 2.0 ± 0.6 and 4.1 ± 0.4 kcal/mol, respectively. Hence, while the presence of IP6 yields a comparable dissociation barrier to that of the low-salt case, the association barrier is notably reduced by 3.8 ± 0.5 kcal/mol. It is also interesting to note that the association barriers across the three cases are consistently lower than that of the above *in vitro* system, while the dissociation barriers are consistently larger. Both effects are likely due to dimensional reduction, which effectively reduces the number of accessible transition state configurations.

To further quantify CA/SP1 assembly kinetics, we performed several shorter simulations (10×10^6 CGMD time steps) at varying protein surface coverage levels. While our previous simulations predict that CA/SP1 does not assemble under low-salt conditions at $2380 \text{ \#}/\mu\text{m}^2$ protein surface coverage, assembly may proceed at elevated concentrations. In particular, we depict the mean CA/SP1 assembly rate as a function of protein surface coverage and assembly conditions in Figure 4A. At low surface coverage (≤ 2500

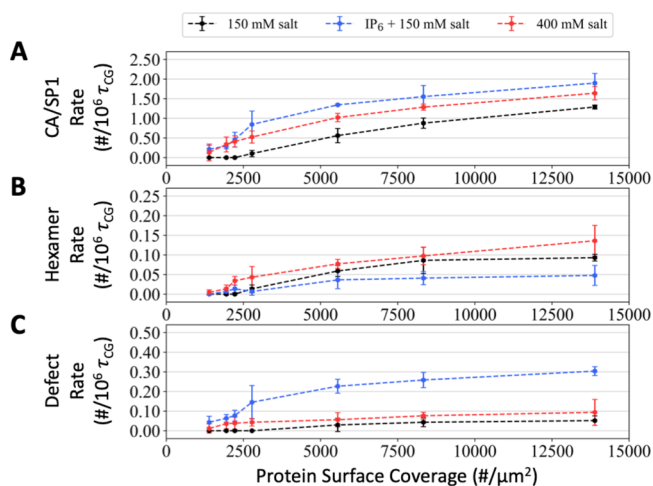


Figure 4. CA/SP1 assembly statistics under varying protein surface coverage densities. (A) Mean assembly rate (# CA/SP1 per 10^6 CG time step) describing the growth of the immature lattice under varying salt and cofactor conditions. (B, C) The mean formation rate (per 10^6 CG time step) for (B) hexamers and (C) defects (i.e., incomplete hexamer) within the immature lattice.

μm^{-2}), both high-salt and low-salt with IP6 conditions promote assembly while low-salt conditions do not. At high surface coverage ($\geq 2500 \mu\text{m}^{-2}$), all three conditions result in productive assembly, with the largest rates achieved under low-salt with IP6 conditions, followed by high-salt and low-salt conditions. Assuming power law rate kinetics, we find that all three assembly conditions follow sublinear scaling with respect to protein surface coverage, in which low-salt, high-salt, and low-salt with IP6 conditions have scaling law orders of 0.63, 0.66, and 0.74, respectively. At this point, it is not clear why sublinear reaction orders emerge. The impact of RNA as a substrate for CA/SP1 colocalization or the assembly of CA/SP1 into small intermediates prior to lattice growth,^{55–58}

including the basic trimer-of-dimer unit that has been implicated for both mature and immature lattice assembly,^{11,48,59} are two factors that are likely to contribute. In the latter case, a sublinear reaction order may indicate that competing intermediates, that is, intermediates that are not productive for lattice assembly, increasingly form at elevated protein concentrations. Nonetheless, it is clear that the addition of IP6 and increased electrostatic screening due to increased salt consistently increase assembly rates compared to physiological salt conditions; low-salt with IP6 and high-salt conditions represent a 3-fold and 2-fold increase, respectively, in effective protein concentration compared to that of low-salt conditions.

We further assess the morphological characteristics of the assembled lattices under varying protein surface coverages and assembly conditions. In Figure 4B and Figure 4C, we depict the rate of formation for complete and defective (i.e., incomplete) hexamers, respectively, within the assembled lattice. We find that high-salt conditions consistently yield the largest rate for hexamer growth while low-salt with IP6 conditions yield the lowest. Intriguingly, both high-salt and low-salt conditions accumulate defective hexamers at rates comparable to that of their hexamer growth rates, although reduced by around 30–50%. The inclusion of IP6, on the other hand, dramatically increases the rate of defect growth by 6-fold compared to the rate of hexamer growth. It is therefore evident that while defects are incorporated into the lattice under all three assembly conditions, and more so at elevated protein concentrations, the rate of defect incorporation is substantially increased due to the presence of IP6.

Defects induced by IP6 are due to kinetic trapping.

To test whether the defects induced by IP6 are kinetically trapped or not, we extracted lattices grown under the influence of IP6 and continued CA/SP1 assembly after IP6 had been removed. Here, we expect kinetically trapped defects to anneal (i.e., heal) once IP6 is removed. We selected two early time-points of lattice assembly from our prior “*in vitro*” CA/SP1 simulations with both 150 mM salt and IP6 present (Figure 2): $\tau = 25 \times 10^6$ and 250×10^6 CG time steps. In both cases, we integrated 10 trajectories with IP6 and 10 trajectories without IP6, each for 75×10^6 CG time steps.

Figure 5 depicts the assembly profiles from all 40 trajectories. As expected, the trajectories in which IP6 was present universally tended toward increasing lattice sizes. It is interesting to note that local remodeling occurs to a certain degree in these cases, as evident by the fluctuations in the assembly profiles. However, as seen in the CA_{CTD} snapshots, nonhexameric defects persist along the lattice edges. In Figure 5A, these defects tended to be pentameric while, in Figure 5B, both pentameric and heptameric defects were observed. By comparison, the trajectories in which IP6 was absent tended toward decreasing lattice size, especially in the larger lattice case, before continuing to grow. The CA_{CTD} snapshots reveal that the lattice edges are largely hexameric. Some partial hexamers persist along the edges yet remain open, and we would expect free CA/SP1 monomers to complete these partial hexamers given time. Most importantly, the heptameric defects in the larger of the two initial lattices (Figure 5B) have been annealed after the removal of IP6. These results demonstrate that the defects induced by IP6 are primarily due to kinetic trapping. Furthermore, our simulations suggest that IP6 can be used to nucleate CA/SP1 assembly but may be

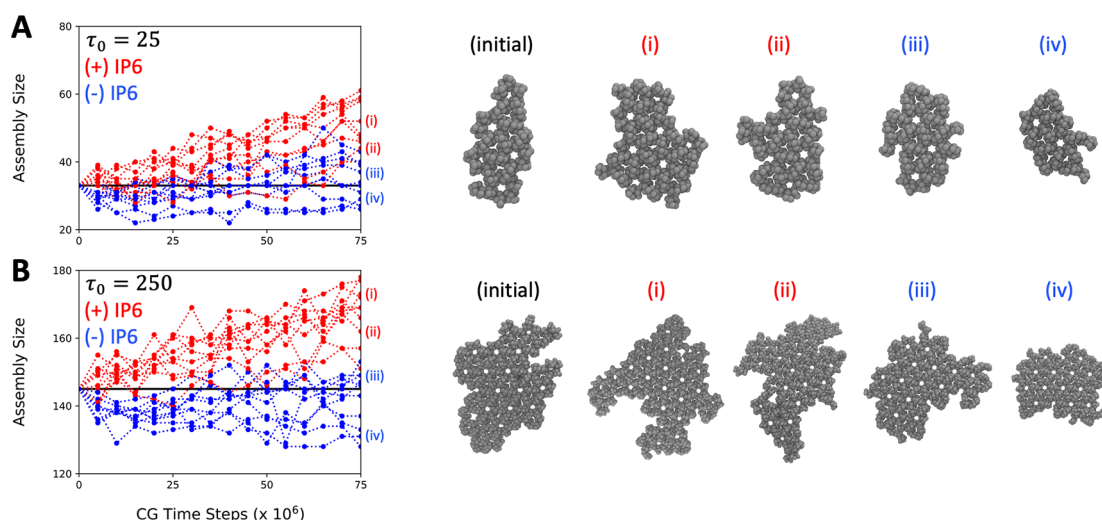


Figure 5. Defects in CA/SP1 + IP6 assemblies anneal when IP6 is removed. Time-series profiles of assembled protein cluster sizes when initialized from “in vitro” CA/SP1 + IP6 morphologies from (A) $\tau = 25 \times 10^6$ and (B) $\tau = 250 \times 10^6$ time steps. The red and blue lines indicate trajectories where IP6 is maintained or removed, respectively, at the beginning of the simulation. The horizontal black line indicates the size of the initial protein cluster. The snapshots to the left correspond to the assembled end-point lattice for each indicated trajectory. Here, only the CTD domains (gray spheres) are shown for clarity.

dispensable for assembly propagation and completion, as also recently shown by Kucharska et al.⁴⁷

DISCUSSION

We have developed a bottom-up (i.e., derived from atomistic simulation data), low-resolution, and implicit-solvent CG model for HIV-1 CA/SP1 and IP6 in order to study their assembly behavior on length- and time-scales relevant to the formation of viral particles. Despite the simplicity of the model, our CGMD simulations recapitulate key experimental observations, including the selective induction of spherical VLPs under the presence of IP6^{42,46} that are morphologically consistent with immature lattices observed *in vivo*.^{3,8} Furthermore, we have investigated the sensitivity of CA/SP1 assembly kinetics and structural outcomes to salt conditions and the presence of IP6.

Thermodynamically driven self-assembly often favors a self-regulated nucleation and growth mechanism that ensures error correction and tends to consist of weak multivalent interactions between specific protomer–protomer interfaces resulting in the formation of structures that maximize these contacts.^{60,61} Assembly can also proceed along kinetically trapped pathways that may yield morphologies that differ from thermodynamically preferred structures. Our simulations suggest that Gag tubule formation, as seen under select *in vitro* conditions⁴⁴ or with single point mutations,⁶² is slow and proceeds through disassembly/reassembly, a known error correction mechanism,⁶³ whereas spherical particle formation is kinetically driven, in which lattice defects represent kinetically trapped states. In the latter case, the lattice grows too quickly for error correction by local remodeling, producing defects that subsequently facilitate curvature throughout the lattice. Our simulations suggest that IP6 is an instigator of kinetically trapped defects, which are also amenable to annealing if IP6 is removed and the defects are present along the lattice edges. Thus, we propose that tubules are a near-equilibrium assembly product while VLPs are a kinetically trapped product. Relatedly, we find that IP6 promotes kinetically trapped morphologies under conditions where

CA/SP1 would not otherwise assemble by lowering the free energy barrier for association rather than stabilizing the associated state, which accelerates CA/SP1 assembly. We should note that while our free energy analysis can qualitatively explain the observed differences between CG assembly outcomes, the quantitative relationship between our CG free energies and true free energies is currently unknown for this model; establishing such a relationship is known as the issue of representability.^{64,65}

The robust creation of immature lattice defects by IP6 may serve additional purposes beyond simple curvature generation. The ability to transform spherical particles into cone-shaped cores via Gag proteolysis is an intriguing property of HIV-1 maturation.¹ During this process, viral protease (PR) cleaves Gag precursors at five different sites, in which cleavage of the site within the CA/SP1 6HB is the rate-limiting step.^{1,66,67} Our prior cryo-ET and MD study showed that lattice defects facilitate uncoiling of the CA/SP1 partial helical bundle, which is crucial for protease access to the CA/SP1 cleavage site.⁵⁴ By inducing fissure-like defects via a kinetically driven process, IP6 may facilitate the emergence of cleavage initiation sites during maturation.

Recent studies of Gag multimerization at the plasma membrane interface suggest that the extent of membrane deformation,¹¹ and relatedly lipid phase segregation,⁶⁸ is coupled to the dynamics of Gag assembly. On the basis of our membrane-bound simulations, we propose that the free energy penalty for lattice growth due to membrane deformation resistance is mitigated by the IP6-induced lattice defects, which serve to relieve strain from lattice curvature. In addition, our free energy calculations indicate that IP6 lowers the free energy of the protomer–protomer bound state compared to the case when IP6 is absent, suggesting that IP6 lowers the free energy of formation for the lattice. Both factors may explain why IP6 appears to be essential for efficient virion production,⁴³ although additional work is necessary to explain why other inositol phosphate molecules, such as inositol pentakisphosphate (IP5), and related polyanions are not as effective.

In the absence of IP6, CA/SP1 constructs have been shown to assemble into mature tubes *in vitro*^{42,46} while our simulations produce relatively flatter and contiguous immature lattices. This observation underscores a limitation of the present CG model, which was trained using immature state statistics and, therefore, restricted to the immature state. It is possible that non-native contacts in the immature state, which are not encoded by the CG model, are participants in the assembly process described in this work, e.g., perhaps reducing the effective rate or propensity for immature assembly. In particular, we note that the mature state, which requires a structural transition from the immature state that involves the uncoiling of the CA/SP1 helical junction and a hinge-like motion such that CA_{NTD} forms the primary hexameric pore, involves native contacts that differ from the immature state.^{24,25,67,69} Incorporating this structural transition into future CG models is likely to be nontrivial, but may be aided by recent methodological efforts to recapitulate conformational transitions within CG space.⁷⁰ Given that the presence of IP6 shifts mature CA assembly outcomes toward conical cores instead of tubules,⁴² we speculate that mature CA assembles into tubules following a near-equilibrium pathway while the presence of IP6 induces kinetically driven assembly into mature cores. Nonetheless, the present work highlights the utility and importance of kinetically driven assembly in the context of immature HIV-1 viral particle production, although additional studies are needed to understand the connection between the fissure-like defects, the kinetics of PR activity throughout the immature lattice, and mature CA assembly during the final stages of maturation.

In conclusion, our findings demonstrate the critical role of IP6 as both an assembly accelerant and morphological modulator of the immature lattice during the late stages of HIV-1 viral morphogenesis. Beyond HIV-1, IP6 has also been implicated as an assembly cofactor in other retroviruses, such as Rous sarcoma virus and equine infectious anemia virus.^{71,72} It would be interesting to investigate the utility of IP6 and its derivatives for the efficient production of broad VLPs that have been repurposed for biomedical technologies, including vaccine development and biologics delivery.^{73–76} Our results will hopefully stimulate future experimental studies as well as the use and development of even more sophisticated CG models.

METHODS

All-Atom (AA) MD Simulations. We used the 18-mer atomic model of CA/SP1 from ref 20 (PDB 5L93) as our initial protein structure. Seven IP6 molecules (each phosphate was deprotonated) were placed between the seven K290/K359 double lysine rings using Monte Carlo insertion. The protein–cofactor complex was solvated by 163 010 water molecules, 646 Na⁺ ions, and 508 Cl[−] ions (i.e., 150 mM NaCl) in a dodecahedron box with 1.5 nm of space between the edges of the protein and the box.

All simulations used the CHARMM36m force field⁷⁷ and were performed using GROMACS 2016.⁷⁸ Minimization was performed using steepest descent until the maximum force was reduced to 500 kJ/mol/nm. Then, equilibration was performed in several phases. First, 10 ns were integrated in the constant NVT ensemble at 310 K using the stochastic velocity rescaling thermostat⁷⁹ with a damping time of 0.1 ps and a time step of 2 fs. During this phase, the heavy atoms of the protein were harmonically restrained with a force constant of 1000 kJ/mol/nm². An additional 10 ns were integrated in the constant NPT ensemble using the Nose–Hoover chain thermostat⁸⁰ at 310 K (2 ps damping time) and the Parrinello–Rahman barostat⁸¹ at 1 bar (10 ps damping time) and a time step of 2 fs; the

restraints on the protein were maintained. Finally, an additional 530 ns were integrated in the constant NVT ensemble using the Nose–Hoover chain thermostat⁸⁰ at 310 K (2 ps damping time) and a time step of 2 fs. Throughout this procedure, H-containing bonds were constrained using the LINCS algorithm.⁸² Protein and cofactor configurations were gathered every 200 ps. The same process was repeated to generate a second trajectory without IP6 molecules. The final 500 ns of each trajectory were used for CG model generation.

CG Model Generation. Using the generated AAMD trajectories described above, we derived a bottom-up model of CA/SP1 and IP6. To map the CA/SP1 protomer, we used the Essential Dynamics Coarse-Graining (EDCG) method⁸³ to identify residue groupings that maximized overlap of the CG model motions with the AA collective motions. We specified 35 CG sites to balance computational efficiency and least-squared error in the represented principal component subspace. We mapped the IP6 molecule into a single CG site using a linear center-of-mass mapping.

After CG mapping, effective CG interactions (E_{CG}) were determined using the following four terms:

$$E_{CGG}(\mathbf{R}) = E_{excl}(\mathbf{R}) + E_{hENM}(\mathbf{R}) + E_{elec}(\mathbf{R}) + E_{attr}(\mathbf{R})$$

where E_{excl} represents excluded volume interactions, E_{hENM} represents intraprotein interactions, E_{elec} represents screened electrostatic interactions, E_{attr} represents attractive interprotein interactions, and \mathbf{R} represents configuration space of the CG coordinates. The heteroelastic network model (hENM) method⁸⁴ with bond energies $k_i (r_i - r_{0,i})^2$ where k_i is the spring constant of a particular effective harmonic bond i and $r_{0,i}$ is the equilibrium bond length for that bond, was used to represent E_{hENM} . These parameters were optimized using the hENM method with a cutoff distance of 2 nm. For E_{excl} a soft cosine potential, $A \left[1 + \cos\left(\frac{\pi r}{r_c}\right) \right]$, was used, where $A = 25$ kcal/mol and r_c is the onset for excluded volume determined from pair correlation functions. For E_{elec} a Yukawa potential,⁸⁵ $\frac{q_i q_j}{4\pi\epsilon_0\epsilon_r\sigma_{ij}} \exp(-\kappa r_{ij})$, was used, where q_i is the aggregate charge of CG site i , $\kappa = 1.274$ or 0.481 nm^{−1} is the inverse Debye length for 150 and 400 mM NaCl, respectively, and ϵ_r is the effective dielectric constant of the protein environment, approximated as 17.5.⁸⁶

Our CG model initially assumes that all close contacts between protomers contribute to protein association through two-body attractive interactions, i.e., E_{attr} . We identified relevant CG site pairs on the basis of their pair correlation functions. If the first peak of the pair correlation function was located below 1.75 nm, and the standard deviation was less than 0.15 nm, that pair of CG sites was selected. We represented E_{attr} as a pairwise Gaussian potential, $\frac{H}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(r_{ij}-r_0)^2}{2\sigma^2}\right]$, where r_0 and σ are the mean and standard deviation determined by a fit to the first peak of the pair correlation function between CG sites i and j through least-squares regression. The constant H was optimized for each pair interaction using relative entropy minimization (REM).⁸⁷ During this process, some of the parameters exhibited diverging behavior or yielded repulsive interactions. Interactions with diverging behavior were removed from the CG model while the remaining parameters were reoptimized, although only minor changes were observed.

We used the same CG RNA polymer model as in our prior CG study of HIV-1 Gag assembly.¹¹ The interaction energies between RNA and the C-terminus of CA/SP1 were represented by E_{scjp} a modified soft-core Lennard–Jones model:⁸⁸

$$E_{scjp} = 4\epsilon\lambda^n \left[\frac{1}{\left(\alpha_{LJ}(1-\lambda)^2 + \left(\frac{r_{ij}}{\sigma}\right)^6\right)^2} - \frac{1}{\alpha_{LJ}(1-\lambda)^2 + \left(\frac{r_{ij}}{\sigma}\right)^6} \right]$$

where $n = 2$, $\alpha_{LJ} = 0.5$, $\lambda = 0.5$, $\epsilon = 1.82$ kcal/mol, and $\sigma = 16$ Å. Here, the minimum interaction energy necessary for RNA binding to the C-terminus of the CA/SP1 protein was identified and used. We also

used the same CG lipid model as in our prior CG study of the SARS-CoV-2 virion.⁸⁹ The interaction energies between the CG lipid headgroup and the top of CA/SP1 (CG types 11 to 13) were also represented by E_{scij} where $n = 2$, $\alpha_{IJ} = 0.5$, $\lambda = 0.6$, $\epsilon = 2.75$ kcal/mol, and $\sigma = 15$ Å. Here, the minimum interaction energy necessary for membrane binding to the N-terminus of the CA/SP1 protein was identified and used.

CG Assembly Simulations. All simulations were prepared using PACKMOL and Moltemplate.^{90,91} For the *in vitro* simulations, 3012 CA/SP1 monomers, 200 50-nt RNA polymers, and 1500 IP6 molecules (when relevant) were randomly distributed throughout a $180 \times 180 \times 180$ nm³ cubic box; two replicas were simulated. For the membrane simulations, a CG lipid bilayer was prepared using a grid distribution of 152 352 lipids (76 176 per monolayer) in a 450×450 nm² lateral (*xy*) domain with an initial spacing of 4 nm between the two monolayer head groups in the *z* direction. An additional 4090 CA/SP1 monomers were uniformly distributed along a 2D grid 1.5 nm below the CG membrane followed by the 9000-nt RNA polymer (3 nm below the layer of CA/SP1) and a random distribution of 4096 IP6 molecules (3 nm below the RNA polymer). Due to computational cost, only one replica was simulated.

All CGMD simulations were performed using LAMMPS 18Jun2019.⁹² Conjugate gradient energy minimization was performed on each system until the change in force was less than 10^{-6} . All systems were initially equilibrated using a Langevin thermostat⁹³ at 310 K (5 ps damping constant) over 5×10^6 τ_{CG} (with $\tau_{CG} = 50$ fs) with attractive protein–protein interactions turned off to remove bias due to initial protein distributions. For the membrane simulations, a Berendsen barostat⁹⁴ at 0 atm (10 ps damping constant, applied over the lateral *xy* dimension) was also applied; during this process, the *xy* simulation domain expanded to 454×454 nm², which remained fixed for the remainder of the simulation.

We then integrated all CG molecules using a Langevin thermostat at 310 K (50 ps damping constant). We simultaneously integrated 10 trajectories over 50×10^6 τ_{CG} with $\tau_{CG} = 50$ fs; during these simulations, we maintained an 85% and 15% active population of CA/SP1 which were randomly selected every 0.5×10^6 τ_{CG} for the *in vitro* and membrane-bound simulations, respectively; active CA/SP1 had their E_{attr} interactions turned on while inactive CA/SP1 had these interactions turned off. Trajectory snapshots were saved every 5×10^6 τ_{CG} . At the end of each set of 10 CGMD simulations, the sizes of the assembled CA/SP1 cluster in each trajectory was computed and their distribution fit to a log-normal function. The variance of the distribution was used to compute the standard error of the mean cluster size. The trajectory closest to the mean cluster size was used to initiate the next set of 10 simultaneous trajectories.

CG Metadynamics Simulations. Free energy profiles were computed using the well-tempered metadynamics method⁵¹ via LAMMPS interfaced with the PLUMED 2.4 plugin.⁹⁵ We used the same CGMD conditions as mentioned above. To emulate membrane-bound effects, we used two planar indenter forces ($F_{indent}(z) = K(z - z_0)^2$) to restrict the *z*-diffusion of the NTD domain; here, $K = 1$ kcal/mol/Å². Gaussian biases with a height of 0.1 kcal/mol and width of 0.05 nm were deposited every 1000 τ_{CG} using a bias factor of $2 k_B T$. The biases were deposited along two collective variables (CVs): (i) the distance between the centers-of-mass of the CTD and (ii) that of the NTD. A harmonic restraint was applied with a force constant of 1000 kcal/mol/Å to prevent the two CVs from exceeding 8 nm.

CG Analysis and Visualization. Assembled clusters were identified using two proximity criteria. Interhexameric contacts were defined by the CTD dimer interface, i.e. when the distance between CG types 25/26 were less than 1.8 nm. Intrahexameric contacts were defined by the top of the CA/SP1 6HB, i.e. when the distance between CG type 32 was less than 1.8 nm. For each trajectory snapshot, an adjacency graph was constructed and extracted using the NetworkX 2.1 (<http://networkx.github.io/>) Python package. Here, each node of the adjacency graph represented a single monomer and the number of nodes in the largest adjacency graph was used to represent cluster sizes. Extracted clusters were visualized using VMD 1.9.3.⁹⁶

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/jacs.2c02568>.

Information and plots regarding distributions of cluster size differences and additional information on the path sampling algorithm. (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Gregory A. Voth – Department of Chemistry, Chicago Center for Theoretical Chemistry, Institute for Biophysical Dynamics, and James Franck Institute, The University of Chicago, Chicago, Illinois 60637, United States; orcid.org/0000-0002-3267-6748; Email: gavoth@uchicago.edu

Authors

Alexander J. Pak – Department of Chemistry, Chicago Center for Theoretical Chemistry, Institute for Biophysical Dynamics, and James Franck Institute, The University of Chicago, Chicago, Illinois 60637, United States; Present Address: Department of Chemical and Biological Engineering, Colorado School of Mines, Golden, CO 80401, United States

Manish Gupta – Department of Chemistry, Chicago Center for Theoretical Chemistry, Institute for Biophysical Dynamics, and James Franck Institute, The University of Chicago, Chicago, Illinois 60637, United States

Mark Yeager – Department of Molecular Physiology and Biological Physics, Cardiovascular Research Center, and Department of Medicine, Division of Cardiovascular Medicine, University of Virginia School of Medicine, Charlottesville, Virginia 22908, United States; Center for Membrane Biology, University of Virginia School of Medicine, Charlottesville, Virginia 22908, United States, United States; Present Address: Frost Institute for Chemistry and Molecular Science, University of Miami, Coral Gables, FL 33124, United States

Complete contact information is available at: <https://pubs.acs.org/10.1021/jacs.2c02568>

Notes

The authors declare no competing financial interest. Data Availability: Model parameters, simulation input files, and custom code are available on Github at: https://github.com/uchicago-voth/MSCG-models/tree/master/HIV_CASPI. The input files and data generated from this study are accessible from Zenodo: <https://zenodo.org/record/6335601>.

■ ACKNOWLEDGMENTS

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. The authors also acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC and visualization resources that have contributed to the research results reported within this paper.⁹⁷ This work also used the Extreme Science and Engineering Discovery Environ-

ment (XSEDE), which is supported by the National Science Foundation Grant Number ACI-1548562.⁹⁸

REFERENCES

- (1) Sundquist, W. I.; Kräusslich, H. G. HIV-1 assembly, budding, and maturation. *Cold Spring Harb Perspect Med.* **2012**, *2* (7), a006924.
- (2) Pornillos, O.; Ganser-Pornillos, B. K. Maturation of retroviruses. *Curr. Opin Virol* **2019**, *36*, 47–55.
- (3) Briggs, J. A.; Riches, J. D.; Glass, B.; Bartonova, V.; Zanetti, G.; Krausslich, H. G. Structure and assembly of immature HIV. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (27), 11090–11095.
- (4) Popova, E.; Popov, S.; Gottlinger, H. G. Human immunodeficiency virus type 1 nucleocapsid p1 confers ESCRT pathway dependence. *J. Virol* **2010**, *84* (13), 6590–6597.
- (5) Datta, S. A. K.; Clark, P. K.; Fan, L.; Ma, B.; Harvin, D. P.; Sowder, R. C.; Nussinov, R.; Wang, Y.-X.; Rein, A.; Sundquist, W. I. Dimerization of the SP1 Region of HIV-1 Gag Induces a Helical Conformation and Association into Helical Bundles: Implications for Particle Assembly. *Journal of Virology* **2016**, *90* (4), 1773–1787.
- (6) Datta, S. A.; Temeselew, L. G.; Crist, R. M.; Soheilian, F.; Kamata, A.; Mirro, J.; Harvin, D.; Nagashima, K.; Cachau, R. E.; Rein, A. On the role of the SP1 domain in HIV-1 particle assembly: a molecular switch? *J. Virol* **2011**, *85* (9), 4111–4121.
- (7) Wright, E. R.; Schooler, J. B.; Ding, H. J.; Kieffer, C.; Fillmore, C.; Sundquist, W. I.; Jensen, G. J. Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells. *EMBO J.* **2007**, *26* (8), 2218–2226.
- (8) Carlson, L. A.; de Marco, A.; Oberwinkler, H.; Habermann, A.; Briggs, J. A.; Krausslich, H. G.; Grunewald, K. Cryo electron tomography of native HIV-1 budding sites. *PLoS Pathog* **2010**, *6* (11), No. e1001173.
- (9) Johnson, D. S.; Bleck, M.; Simon, S. M. Timing of ESCRT-III protein recruitment and membrane scission during HIV-1 assembly. *Elife* **2018**, *7*, DOI: 10.7554/eLife.36221
- (10) Carlson, L. A.; Briggs, J. A.; Glass, B.; Riches, J. D.; Simon, M. N.; Johnson, M. C.; Muller, B.; Grunewald, K.; Krausslich, H. G. Three-dimensional analysis of budding sites and released virus suggests a revised model for HIV-1 morphogenesis. *Cell Host Microbe* **2008**, *4* (6), 592–599.
- (11) Pak, A. J.; Grime, J. M. A.; Sengupta, P.; Chen, A. K.; Durumeric, A. E. P.; Srivastava, A.; Yeager, M.; Briggs, J. A. G.; Lippincott-Schwartz, J.; Voth, G. A. Immature HIV-1 lattice assembly dynamics are regulated by scaffolding from nucleic acid and the plasma membrane. *Proc. Natl. Acad. Sci.* **2017**, *114* (47), e10056–E10065.
- (12) Muriaux, D.; Darlix, J. L. Properties and functions of the nucleocapsid protein in virus assembly. *RNA Biol.* **2010**, *7* (6), 744–753.
- (13) Yang, Y.; Qu, N.; Tan, J.; Rushdi, M. N.; Krueger, C. J.; Chen, A. K. Roles of Gag-RNA interactions in HIV-1 virus assembly deciphered by single-molecule localization microscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (26), 6721–6726.
- (14) Shi, J.; Zhou, J.; Shah, V. B.; Aiken, C.; Whitby, K. Small-molecule inhibition of human immunodeficiency virus type 1 infection by virus capsid destabilization. *J. Virol* **2011**, *85* (1), 542–549.
- (15) Bester, S. M.; Wei, G.; Zhao, H.; Adu-Ampratwum, D.; Iqbal, N.; Courouble, V. V.; Francis, A. C.; Annamalai, A. S.; Singh, P. K.; Shkriabai, N.; et al. Structural and mechanistic bases for a potent HIV-1 capsid inhibitor. *Science* **2020**, *370* (6514), 360–364.
- (16) Pak, A. J.; Grime, J. M. A.; Yu, A.; Voth, G. A. Off-Pathway Assembly: A Broad-Spectrum Mechanism of Action for Drugs That Undermine Controlled HIV-1 Viral Capsid Formation. *J. Am. Chem. Soc.* **2019**, *141* (26), 10214–10224.
- (17) Du, S.; Betts, L.; Yang, R.; Shi, H.; Concel, J.; Ahn, J.; Aiken, C.; Zhang, P.; Yeh, J. I. Structure of the HIV-1 full-length capsid protein in a conformationally trapped unassembled state induced by small-molecule binding. *J. Mol. Biol.* **2011**, *406* (3), 371–386.
- (18) Momany, C.; Kovari, L. C.; Prongay, A. J.; Keller, W.; Gitti, R. K.; Lee, B. M.; Gorbalenya, A. E.; Tong, L.; McClure, J.; Ehrlich, L. S.; et al. Crystal structure of dimeric HIV-1 capsid protein. *Nat. Struct. Biol.* **1996**, *3* (9), 763–770.
- (19) Gamble, T. R.; Yoo, S.; Vajdos, F. F.; von Schwedler, U. K.; Worthylake, D. K.; Wang, H.; McCutcheon, J. P.; Sundquist, W. I.; Hill, C. P. Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science* **1997**, *278* (5339), 849–853.
- (20) Schur, F. K. M.; Obr, M.; Hagen, W. J. H.; Wan, W.; Jakobi, A. J.; Kirkpatrick, J. M.; Sachse, C.; Kräusslich, H.-G.; Briggs, J. A. G. An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science* **2016**, *353*, 506–508.
- (21) Wagner, J. M.; Zdrozny, K. K.; Chrustowicz, J.; Purdy, M. D.; Yeager, M.; Ganser-Pornillos, B. K.; Pornillos, O. Crystal structure of an HIV assembly and maturation switch. *Elife* **2016**, *5*, DOI: 10.7554/eLife.17063.
- (22) Schur, F. K.; Hagen, W. J.; Rumlová, M.; Ruml, T.; Müller, B.; Kräusslich, H. G.; Briggs, J. A. Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution. *Nature* **2015**, *517* (7535), 505–508.
- (23) Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; et al. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **2013**, *497* (7451), 643–646.
- (24) Gres, A. T.; Kirby, K. A.; KewalRamani, V. N.; Tanner, J. J.; Pornillos, O.; Sarafianos, S. G. STRUCTURAL VIROLOGY. X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability. *Science* **2015**, *349* (6243), 99–103.
- (25) Pornillos, O.; Ganser-Pornillos, B. K.; Kelly, B. N.; Hua, Y.; Whitby, F. G.; Stout, C. D.; Sundquist, W. I.; Hill, C. P.; Yeager, M. X-ray structures of the hexameric building block of the HIV capsid. *Cell* **2009**, *137* (7), 1282–1292.
- (26) Accola, M. A.; Strack, B.; Göttlinger, H. G. Efficient particle production by minimal Gag constructs which retain the carboxy-terminal domain of human immunodeficiency virus type 1 capsid-p2 and a late assembly domain. *J. Virol* **2000**, *74* (12), 5395–5402.
- (27) Guo, X.; Roldan, A.; Hu, J.; Wainberg, M. A.; Liang, C. Mutation of the SP1 sequence impairs both multimerization and membrane-binding activities of human immunodeficiency virus type 1 Gag. *J. Virol* **2005**, *79* (3), 1803–1812.
- (28) Gross, I.; Hohenberg, H.; Krausslich, H. G. In vitro assembly properties of purified bacterially expressed capsid proteins of human immunodeficiency virus. *Eur. J. Biochem.* **1997**, *249* (2), 592–600.
- (29) Khorchid, A.; Halwani, R.; Wainberg, M. A.; Kleiman, L. Role of RNA in facilitating Gag/Gag-Pol interaction. *J. Virol* **2002**, *76* (8), 4131–4137.
- (30) Muriaux, D.; Mirro, J.; Harvin, D.; Rein, A. RNA is a structural element in retrovirus particles. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (9), 5246–5251.
- (31) Webb, J. A.; Jones, C. P.; Parent, L. J.; Rouzina, I.; Musier-Forsyth, K. Distinct binding interactions of HIV-1 Gag to Psi and non-Psi RNAs: implications for viral genomic RNA packaging. *RNA* **2013**, *19* (8), 1078–1088.
- (32) Ding, P.; Kharytonchyk, S.; Waller, A.; Mbaekwe, U.; Basappa, S.; Kuo, N.; Frank, H. M.; Quasney, C.; Kidane, A.; Swanson, C.; et al. Identification of the initial nucleocapsid recognition element in the HIV-1 RNA packaging signal. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (30), 17737–17746.
- (33) Dick, R. A.; Vogt, V. M. Membrane interaction of retroviral Gag proteins. *Front Microbiol* **2014**, *5*, 187.
- (34) Yandrapalli, N.; Muriaux, D.; Favard, C. Lipid domains in HIV-1 assembly. *Front Microbiol* **2014**, *5*, 220.
- (35) Monje-Galvan, V.; Voth, G. A. Binding mechanism of the matrix domain of HIV-1 gag on lipid membranes. *Elife* **2020**, *9*, DOI: 10.7554/eLife.58621.
- (36) Vlach, J.; Saad, J. S. Structural and molecular determinants of HIV-1 Gag binding to the plasma membrane. *Front Microbiol* **2015**, *6*, 232.

- (37) Ono, A.; Ablan, S. D.; Lockett, S. J.; Nagashima, K.; Freed, E. O. Phosphatidylinositol (4,5) bisphosphate regulates HIV-1 gag targeting to the plasma membrane. *P Natl. Acad. Sci. USA* **2004**, *101* (41), 14889–14894.
- (38) Campbell, S.; Fisher, R. J.; Towler, E. M.; Fox, S.; Issaq, H. J.; Wolfe, T.; Phillips, L. R.; Rein, A. Modulation of HIV-like particle assembly in vitro by inositol phosphates. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (19), 10875–10879.
- (39) Campbell, S.; Vogt, V. M. Self-Assembly in-Vitro of Purified Ca-Nc Proteins from Rous-Sarcoma Virus and Human-Immunodeficiency-Virus Type-1. *Journal of Virology* **1995**, *69* (10), 6487–6497.
- (40) Letcher, A. J.; Schell, M. J.; Irvine, R. F. Do mammals make all their own inositol hexakisphosphate? *Biochem. J.* **2008**, *416* (2), 263–270.
- (41) Barker, C. J.; Wright, J.; Hughes, P. J.; Kirk, C. J.; Michell, R. H. Complex changes in cellular inositol phosphate complement accompany transit through the cell cycle. *Biochem. J.* **2004**, *380*, 465–473.
- (42) Dick, R. A.; Zadrozny, K. K.; Xu, C.; Schur, F. K. M.; Lyddon, T. D.; Ricana, C. L.; Wagner, J. M.; Perilla, J. R.; Ganser-Pornillos, B. K.; Johnson, M. C.; et al. Inositol phosphates are assembly co-factors for HIV-1. *Nature* **2018**, *560* (7719), 509–512.
- (43) Mallery, D. L.; Faysal, K. M. R.; Kleinpeter, A.; Wilson, M. S. C.; Vaysburd, M.; Fletcher, A. J.; Novikova, M.; Böcking, T.; Freed, E. O.; Saiardi, A.; et al. Cellular IP6 Levels Limit HIV Production while Viruses that Cannot Efficiently Package IP6 Are Attenuated for Infection and Replication. *Cell Reports* **2019**, *29* (12), 3983–3996.
- (44) Gross, I.; Hohenberg, H.; Wilk, T.; Wieggers, K.; Grattinger, M.; Muller, B.; Fuller, S.; Krausslich, H. G. A conformational switch controlling HIV-1 morphogenesis. *EMBO J.* **2000**, *19* (1), 103–113.
- (45) von Schwedler, U. K.; Stemmler, T. L.; Klishko, V. Y.; Li, S.; Albertine, K. H.; Davis, D. R.; Sundquist, W. I. Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly. *EMBO J.* **1998**, *17* (6), 1555–1568.
- (46) Dostalkova, A.; Kaufman, F.; Krizova, I.; Vokata, B.; Ruml, T.; Rumlova, M. In Vitro Quantification of the Effects of IP6 and Other Small Polyanions on Immature HIV-1 Particle Assembly and Core Stability. *J. Virol* **2020**, *94* (20), DOI: 10.1128/JVI.00991-20.
- (47) Kucharska, I.; Ding, P.; Zadrozny, K. K.; Dick, R. A.; Summers, M. F.; Ganser-Pornillos, B. K.; Pornillos, O. Biochemical Reconstitution of HIV-1 Assembly and Maturation. *J. Virol* **2020**, *94* (5), DOI: 10.1128/JVI.01844-19.
- (48) Grime, J. M. A.; Dama, J. F.; Ganser-Pornillos, B. K.; Woodward, C. L.; Jensen, G. J.; Yeager, M.; Voth, G. A. Coarse-grained simulation reveals key features of HIV-1 capsid self-assembly. *Nat. Commun.* **2016**, *7*, 11568.
- (49) Harada, R.; Kitao, A. Parallel Cascade Selection Molecular Dynamics (PaCS-MD) to generate conformational transition pathway. *J. Chem. Phys.* **2013**, *139* (3), 035103.
- (50) Dama, J. F.; Parrinello, M.; Voth, G. A. Well-tempered metadynamics converges asymptotically. *Phys. Rev. Lett.* **2014**, *112* (24), 240602.
- (51) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100* (2), 020603.
- (52) Borech, S.; Karplus, M. The Jacobian factor in free energy simulations. *J. Chem. Phys.* **1996**, *105* (12), 5145–5154.
- (53) Sprik, M.; Ciccotti, G. Free energy from constrained molecular dynamics. *J. Chem. Phys.* **1998**, *109* (18), 7737–7744.
- (54) Tan, A.; Pak, A. J.; Morado, D. R.; Voth, G. A.; Briggs, J. A. G. Immature HIV-1 assembles from Gag dimers leaving partial hexamers at lattice edges as potential substrates for proteolytic maturation. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118* (3), No. e2020054118.
- (55) Morozov, A. Y.; Bruinsma, R. F.; Rudnick, J. Assembly of viruses and the pseudo-law of mass action. *J. Chem. Phys.* **2009**, *131* (15), 155101.
- (56) Zlotnick, A. Distinguishing reversible from irreversible virus capsid assembly. *J. Mol. Biol.* **2007**, *366* (1), 14–18.
- (57) Bruinsma, R. F.; Gelbart, W. M.; Reguera, D.; Rudnick, J.; Zandi, R. Viral self-assembly as a thermodynamic process. *Phys. Rev. Lett.* **2003**, *90* (24), DOI: 10.1103/PhysRevLett.90.248101.
- (58) Hagan, M. F.; Chandler, D. Dynamic pathways for viral capsid assembly. *Biophys. J.* **2006**, *91* (1), 42–54.
- (59) Tsiang, M.; Niedziela-Majka, A.; Hung, M.; Jin, D. B.; Hu, E.; Yant, S.; Samuel, D.; Liu, X. H.; Sakowicz, R. A Trimer of Dimers Is the Basic Building Block for Human Immunodeficiency Virus-1 Capsid Assembly. *Biochemistry-U S* **2012**, *51* (22), 4416–4428.
- (60) Williams, R. J.; Smith, A. M.; Collins, R.; Hodson, N.; Das, A. K.; Ulijn, R. V. Enzyme-assisted self-assembly under thermodynamic control. *Nat. Nanotechnol* **2009**, *4* (1), 19–24.
- (61) Whitelam, S.; Jack, R. L. The statistical mechanics of dynamic pathways to self-assembly. *Annu. Rev. Phys. Chem.* **2015**, *66*, 143–163.
- (62) Bharat, T. A. M.; Menendez, L. R. C.; Hagen, W. J. H.; Lux, V.; Igonet, S.; Schorb, M.; Schur, F. K. M.; Krausslich, H. G.; Briggs, J. A. G. Cryo-electron microscopy of tubular arrays of HIV-1 Gag resolves structures essential for immature virus assembly. *P Natl. Acad. Sci. USA* **2014**, *111* (22), 8233–8238.
- (63) Keller, P. W.; Huang, R. K.; England, M. R.; Waki, K.; Cheng, N.; Heymann, J. B.; Craven, R. C.; Freed, E. O.; Steven, A. C. 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* **2013**, *87* (24), 13655–13664.
- (64) Wagner, J. W.; Dama, J. F.; Durumeric, A. E. P.; Voth, G. A. On the representability problem and the physical meaning of coarse-grained models. *J. Chem. Phys.* **2016**, *145* (4), 044108.
- (65) Jin, J.; Pak, A. J.; Voth, G. A. Understanding Missing Entropy in Coarse-Grained Systems: Addressing Issues of Representability and Transferability. *J. Phys. Chem. Lett.* **2019**, *10* (16), 4549.
- (66) Pettit, S. C.; Lindquist, J. N.; Kaplan, A. H.; Swanstrom, R. Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-Pol precursor are cleaved by the viral protease at different rates. *Retrovirology* **2005**, *2*, 66.
- (67) Mattei, S.; Tan, A.; Glass, B.; Muller, B.; Krausslich, H. G.; Briggs, J. A. G. High-resolution structures of HIV-1 Gag cleavage mutants determine structural switch for virus maturation. *P Natl. Acad. Sci. USA* **2018**, *115* (40), e9401–E9410.
- (68) Sengupta, P.; Seo, A. Y.; Pasolli, H. A.; Song, Y. E.; Johnson, M.; Lippincott-Schwartz, J. A lipid-based partitioning mechanism for selective incorporation of proteins into membranes of HIV particles. *Nat. Cell Biol.* **2019**, *21* (4), 452.
- (69) Wang, M. Z.; Quinn, C. M.; Perilla, J. R.; Zhang, H. L.; Shirra, R.; Hou, G. J.; Byeon, I. J.; Suiter, C. L.; Ablan, S.; Urano, E. et al. Quenching protein dynamics interferes with HIV capsid maturation. *Nat. Commun.* **2017**, *8*, DOI: 10.1038/s41467-017-01856-y.
- (70) Sharp, M. E.; Vazquez, F. X.; Wagner, J. W.; Dannenhoffer-Lafage, T.; Voth, G. A. Multiconfigurational Coarse-Grained Molecular Dynamics. *J. Chem. Theory Comput.* **2019**, *15* (5), 3306–3315.
- (71) Dick, R. A.; Xu, C. Y.; Morado, D. R.; Kravchuk, V.; Ricana, C. L.; Lyddon, T. D.; Broad, A. M.; Feathers, J. R.; Johnson, M. C.; Vogt, V. M. Structures of immature EIAV Gag lattices reveal a conserved role for IP6 in lentivirus assembly. *Plos Pathogens* **2020**, *16* (1), e1008277.
- (72) Obr, M.; Ricana, C. L.; Nikulin, N.; Feathers, J. P. R.; Klanschnig, M.; Thader, A.; Johnson, M. C.; Vogt, V. M.; Schur, F. K. M.; Dick, R. A. Structure of the mature Rous sarcoma virus lattice reveals a role for IP6 in the formation of the capsid hexamer. *Nat. Commun.* **2021**, *12* (1), DOI: 10.1038/s41467-021-23506-0.
- (73) Ding, X. W.; Liu, D.; Booth, G.; Gao, W.; Lu, Y. Virus-Like Particle Engineering: From Rational Design to Versatile Applications. *Biotechnol J.* **2018**, *13* (5), 1700324.
- (74) Frietze, K. M.; Peabody, D. S.; Chackerian, B. Engineering virus-like particles as vaccine platforms. *Current Opinion in Virology* **2016**, *18*, 44–49.
- (75) Hill, B. D.; Zak, A.; Khera, E.; Wen, F. Engineering Virus-like Particles for Antigen and Drug Delivery. *Curr. Protein Pept Sc* **2018**, *19* (1), 112–127.

- (76) Banskota, S.; Raguram, A.; Suh, S.; Du, S. W.; Davis, J. R.; Choi, E. H.; Wang, X.; Nielsen, S. C.; Newby, G. A.; Randolph, P. B.; et al. Engineered virus-like particles for efficient in vivo delivery of therapeutic proteins. *Cell* **2022**, *185* (2), 250–265.
- (77) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71.
- (78) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.
- (79) Bussi, G.; Zykova-Timan, T.; Parrinello, M. Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *J. Chem. Phys.* **2009**, *130* (7), 074101.
- (80) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé-Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **1992**, *97* (4), 2635–2643.
- (81) Parrinello, M.; Rahman, A. Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Phys. Rev. Lett.* **1980**, *45* (14), 1196–1199.
- (82) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463–1472.
- (83) Zhang, Z.; Lu, L.; Noid, W. G.; Krishna, V.; Pfandner, J.; Voth, G. A. A Systematic Methodology for Defining Coarse-Grained Sites in Large Biomolecules. *Biophys. J.* **2008**, *95* (11), 5073–5083.
- (84) Lyman, E.; Pfandner, J.; Voth, G. A. Systematic Multiscale Parameterization of Heterogeneous Elastic Network Models of Proteins. *Biophys. J.* **2008**, *95* (9), 4183–4192.
- (85) Yukawa, H. On the Interaction of Elementary Particles. I. *Proceedings of the Physico-Mathematical Society of Japan. 3rd Series* **1935**, *17*, 48–57.
- (86) Li, L.; Li, C.; Zhang, Z.; Alexov, E. On the Dielectric “Constant” of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi. *J. Chem. Theory Comput.* **2013**, *9* (4), 2126–2136.
- (87) Shell, M. S. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *J. Chem. Phys.* **2008**, *129* (14), 144108.
- (88) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994**, *222*, 529–539.
- (89) Yu, A.; Pak, A. J.; He, P.; Monje-Galvan, V.; Casalino, L.; Gaieb, Z.; Dommer, A. C.; Amaro, R. E.; Voth, G. A. A multiscale coarse-grained model of the SARS-CoV-2 virion. *Biophys. J.* **2021**, *120* (6), 1097–1104.
- (90) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30* (13), 2157–2164.
- (91) Jewett, A. I.; Zhuang, Z.; Shea, J.-E. Moltemplate a Coarse-Grained Model Assembly Tool. *Biophys. J.* **2013**, *104* (2), 169a.
- (92) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117* (1), 1–19.
- (93) Vanden-Eijnden, E.; Ciccotti, G. Second-order integrators for Langevin equations with holonomic constraints. *Chem. Phys. Lett.* **2006**, *429* (1), 310–316.
- (94) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690.
- (95) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185* (2), 604–613.
- (96) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph* **1996**, *14* (1), 33–38 27–38.
- (97) Stanzione, D.; West, J.; Evans, R. T.; Minyard, T.; Ghattas, O.; Panda, D. K. Frontera: The Evolution of Leadership Computing at the National Science Foundation. In *Practice and Experience in Advanced Research Computing (PEARC '20)*, July 26, 2020, 2020; pp 106–111. DOI: 10.1145/3311790.3396656.
- (98) Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gathier, K.; Grimshaw, A.; Hazlewood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D.; et al. XSEDE: Accelerating Scientific Discovery. *Computing in Science & Engineering* **2014**, *16* (5), 62–74.