Title
Atomic-scale characterization of mature HIV-1 capsid stabilization by inositol hexakisphosphate (IP6)

Permalink
https://escholarship.org/uc/item/3zv15209

Journal
Science Advances, 6(38)

ISSN
2375-2548

Authors
Yu, Alvin
Lee, Elizabeth MY
Jin, Jaehyeok
et al.

Publication Date
2020-09-18

DOI
10.1126/sciadv.abc6465

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed
Atomic-scale characterization of mature HIV-1 capsid stabilization by inositol hexakisphosphate (IP$_6$)

Alvin Yu$^1$, Elizabeth M. Y. Lee$^2$, Jaehyeok Jin$^1$, Gregory A. Voth$^1$

Inositol hexakisphosphates (IP$_6$) are cellular cofactors that promote the assembly of mature capsids of HIV. These negatively charged molecules coordinate an electropositive ring of arginines at the center of pores distributed throughout the capsid surface. Kinetic studies indicate that the binding of IP$_6$ increases the stable lifetimes of the capsid by several orders of magnitude from minutes to hours. Using all-atom molecular dynamics simulations, we uncover the mechanisms that underlie the unusually high stability of mature capsids in complex with IP$_6$. We find that capsid hexamers and pentamers have differential binding modes for IP$_6$. Ligand density calculations show three sites of interaction with IP$_6$ including at a known capsid inhibitor binding pocket. Free energy calculations demonstrate that IP$_6$ preferentially stabilizes pentamers over hexamers to enhance fullerene modes of assembly. These results elucidate the molecular role of IP$_6$ in stabilizing and assembling the retroviral capsid.

INTRODUCTION

The capsids of HIV-1 encase and protect the retroviral genome from natural host defense mechanisms but are transiently stable complexes. These capsids are assembled inside viral particles after proteolytic cleavage of the Gag polyprotein in a large-scale morphological transition of the virus termed maturation (1) and must be disassembled before integration of the virus’ genetic information into the host cell (2). Structurally, immature Gag lattice precursors form incomplete spherical assemblies, composed primarily of hexamers of uncleaved Gag polyprotein (3), while mature HIV-1 capsids are fullerene-like structures, composed of hexamer and pentamer building blocks of a single capsid domain (CA) protein (4). There are exactly 12 pentamers in a mature capsid that is topologically required to enclose the shell and impart curvature onto the CA lattice. Biochemical reconstitutions of CA assemblies show that CA can self-assemble into a wide range of morphologies, including helical tubes and fullerene cones (5). CA lattices composed solely of hexamers, however, form helical tubes rather than fullerene structures. Characterization of capsids with fullerene geometries (6) that resemble those imaged by electron cryo-tomography within intact virions has been notoriously difficult, in part because of the short lifetimes under which capsids remain stable in isolation and the structural heterogeneity of the capsids themselves. Recent x-ray crystal structures demonstrate that pH-gated pores found at the center of hexamer subunits are distributed throughout the capsid surface and contain an electropositive ring of arginines (R18) at the narrowest region of the pore (7, 8). Polyanions, including nucleotides, readily bind to the arginine ring. In particular, inositol hexakisphosphate (IP$_6$) is a cellular polyanion that is abundantly present in the cytosol and enriched in packaged viral particles that bud off from the cell (9). The addition of IP$_6$ to in vitro reconstitutions of the capsid drastically increases the stable lifetimes of the HIV-1 capsid from several minutes to over 10 hours and promotes the assembly of fullerene capsids (9, 10). Viruses that fail to package IP$_6$ do not mature properly and have unstable capsids (11). Previous molecular simulations examined the conformational fluctuations of portions of the immature Gag lattice in the presence and absence of an IP$_6$ ligand and found noticeable distortions of the six-helix bundle in the Gag spacer peptide 1 without IP$_6$ (10). Mutagenesis studies have revealed that removal of electrostatic repulsion between CA domains at the arginine ring (R18A) causes assembly behavior consistent with an increased frequency of pentamer incorporation (5, 12). Despite these advances, there are no studies on whether IP$_6$ interacts with pentamers, and the molecular mechanisms that underlie the increased stability and enhanced assembly of the mature capsid by IP$_6$ remain unclear.

To investigate this, we performed long time scale all-atom (AA) molecular dynamics (MD) simulations of CA proteins with IP$_6$ in bulk solution and analyzed the mechanisms by which IP$_6$ binds to capsid pores. This analysis reveals that CA hexamers and pentamers have unexpectedly different binding modes for IP$_6$, despite the high structural similarity between the capsid oligomers. To identify the protein-ligand interactions made during the IP$_6$ binding process, we calculated three-dimensional (3D) density maps for IP$_6$ that specify the regions the ligand occupies during dynamical simulation. These density maps show two weaker-affinity sites for IP$_6$ and CA, one of which corresponds to a previously discovered capsid inhibitor binding pocket. Free energy calculations of the binding process indicate that IP$_6$ stabilizes pentameric CA states to a greater extent than hexameric CA states. These results suggest that IP$_6$ binding shifts assembly conditions toward an increased population of pentamers to stabilize the fullerene capsid.

RESULTS

CA hexamers and pentamers have differential IP$_6$ binding modes

MD simulations of the complete mature HIV-1 capsid (~100 million atoms, Fig. 1A) (13) are computationally costly. Instead, to investigate how IP$_6$ stabilizes the capsid, we simulated individual hexamer and pentamer components of the capsid [Protein Data Bank (PDB) ID: 3H47, 5HGL, and 5MCY] using unbiased, microsecond time scale AA MD simulations on the special-purpose computer hardware, Anton 2, at the Pittsburgh Supercomputing Center (14). Similar to
prior studies, IP₆ was initially added to the bulk solvent, approximately 10 Å away from any nonwater molecules (15, 16). Four systems each for the CA hexamer and pentamer were prepared with IP₆ in randomized positions in bulk solvent and selected for longer time scale simulations on Anton 2 (see Materials and Methods for a complete description). In aggregate, the AA MD simulations totaled 32 µs for the hexamer and pentamer systems. In all cases, IP₆ was found to bind to the pore of the capsomere.

After exhaustive MD simulation (~4 µs per trajectory) of the CA hexamers, two IP₆ molecules bind to a central arginine (R18) ring that lines the hexameric pore. A single IP₆ ligand occupies the volume ~3 Å above the R18 ring, with phosphate groups coordinated by the functional groups of the R18 side chains, whereas a second IP₆ molecule binds below the arginine ring, also coordinated by the functional groups of the R18 side chains, whereas a second IP₆ molecule binds to the R18 ring at a pore in the central cap of the CA pentamer (PDB ID: 5MCY) after 1.98 µs of AA MD simulations. The bound complex is shown in a top-down and side view. In the MD simulations for (B) and (C), IP₆ molecules were initially placed in bulk solvent, ~10 Å away from the CA domains. The NTD of CA is colored in green, and the CTD of CA is colored in brown. R18 residues are shown in cyan.

**Molecular mechanism of IP₆ binding to CA hexamers and pentamers**

The binding of IP₆ not only enables the in vitro assembly of fulleren capsids from concentrated CA solutions ([CA] > 1 mM) but also markedly increases the lifetime of mature retroviral capsids synthesized in the laboratory from minutes to hours (17). In the simulations, IP₆ exits bulk solvent and can enter the constricted pore from either the N-terminal domain (NTD) or CTD regions of the CA pentamer or hexamer. Snapshots of the binding process taken at various stages of ligand association are shown in Figs. 2 and 3, and the full process is shown in movies S1 and S2.

For the CA pentamer, IP₆ initially diffuses past the β-hairpin loop in the NTD. The side chains of the R18 ring are typically modeled to be planar in x-ray crystal structures and cryo–electron microscopy (cryo-EM) maps with torsional angles of χ₁ = −74°, χ₂ = −156°, χ₃ = 71°, and χ₄ = 84° (Fig. 2, A and B), yet before binding, the pentameric complex shifts asymmetrically, involving a vertical separation of ~3 Å between the CA carbons of R18 on helix H1 of the α and β domain, and the CA domains twist helically upward with β < γ < δ < ε < α along the axis of the pore (Fig. 2, B and C). The pore arginine of the α-CA domain flips away from the plane with rotations about the R18 side chains (χ₁ = −70°, χ₂ = 57°, χ₃ = 60°, and χ₄ = −179°), likely reflecting unfavorable electrostatic interactions between the guanidino groups (Fig. 2C). The arginine side chain coordinates a phosphate group on IP₆ (Fig. 2D), as IP₆ binds to the pocket above the R18 ring. Additional arginine groups from the remaining CA domains contact the uncoordinated phosphate groups of the ligand, as IP₆ binds into the central pore, and the R18 ring relaxes to a planar configuration (Fig. 2, E to H).

In the CA hexamer, the overall complex does not twist asymmetrically, and IP₆ enters the pore initially from the CTD (Fig. 3, A to C). Salt-bridge interactions between K25 and E29 at the base of helix H1 break, and the lysine side chains interact with IP₆ (Fig. 3C). One of the pore arginines of the β-CA domain rotates downward to contact a phosphate group on IP₆ and forms a metastable complex between the protein and ligand, in which two IP₆ phosphate groups are coordinated at both ends by R18 and K25 (Fig. 3, D and E). A second IP₆ molecule enters the pore from the NTD region, anchored by an uncoordinated arginine in the R18 ring (Fig. 3F). Last, interactions between the first IP₆ molecule and K25 break, as both IP₆ ligands relax into the central binding sites above and below the R18 ring (Fig. 3, G and H). Previous simulations have also reported the presence of a salt-bridge contact between K30 and E28 that may stabilize CA pentamers more than hexamers (18). Our simulations were consistent with these results with closer contacts between K30 and E28 in the pentamer than hexamer trajectories (average Ca-Ca separation distances, 12.1 and 14.6 Å, respectively).

Four different initial conditions for the hexamers and pentamers were simulated, including combinations of open and closed CA conformations, containing protonated and deprotonated H12 residues (see table S1). X-ray structures had shown an iris-like opening of the hexamer pore in response to acidic pH conditions, hypothesized to occur by protonation of a critical histidine residue (H12), and a conformational change in the β-hairpin that gates the pore (7). In the simulations, however, IP₆ bound in the same fashion and made similar molecular interaction during binding to the R18 ring for CA complexes containing both protonated and deprotonated H12 and either open or closed conformations. In general, IP₆ bound more quickly to CA domains in open rather than closed conformations, and with protonated H12 residues rather than deprotonated H12 residues. The β-hairpin regions were quite flexible, suggesting that opening or closing of the proposed gate is driven more by small changes in the conformational ensembles adopted by the β-hairpin rather than by abrupt structural changes. IP₆ binding to pentamers...
involved more concerted motions in the CA NTDs compared to hexamers. The amounts of time ligands spent in bulk solvent before binding varied between CA hexamers and pentamers but were of similar time scale in the MD simulations.

Fig. 2. Mechanism of IP₆ binding to CA pentamers. A series of snapshots from part of an MD trajectory show the chemical interactions involved in IP₆ binding to the CA pentamer. Three CA domains are shown in a side view of the pentamer, with helix H1 of the CA domain labeled as α, β, and γ, respectively. Two additional CA domains (δ and ε) are not shown for clarity. The R18 ring and residues (K25-E29) that form a salt-bridge interaction proximal to the CTD are labeled. Time points are labeled in the upper-right corner of each panel. (A) The IP₆ ligand is initially in bulk solvent. (B) Close-up view of the pore region in (A). (C) IP₆ enters the pore past the β-hairpin of CA, and an arginine side chain flips away from the R18 ring. (D) The R18 side chain coordinates IP₆. (E) IP₆ is coordinated to multiple R18 side chains. (F) IP₆ shifts toward the R18 ring. (G) IP₆ reorients in the binding pocket. (H) IP₆ is bound to the central arginine ring with R18 side chains contacting the negatively charged phosphates.

Fig. 3. Mechanism of IP₆ binding to CA hexamers. A series of snapshots from part of an MD trajectory show the chemical interactions involved in IP₆ binding to the CA hexamer. Four CA domains are shown in a side view of the hexamer, with helix H1 of the CA domain labeled as α, β, γ, and δ, respectively. Two additional CA domains (ε and ζ) are not shown for clarity. The R18 ring and residues (K25-E29) that form a salt-bridge interaction proximal to the CTD are labeled. Time points are labeled in the upper-right corner of each panel. (A) IP₆ molecules are initially in bulk solvent (two are shown). (B) Close-up view of the pore region in (A). (C) An IP₆ ligand enters the pore region from the CTD of CA. The K25-E29 salt-bridge interaction breaks, and K25 contacts IP₆. (D) A R18 side chain flips toward the CTD to coordinate IP₆. (E) IP₆ binds at the CTD side of the R18 ring. (F) A second IP₆ ligand enters the pore region, coordinated by an R18 side chain. (G) Multiple R18 side chains contact the second IP₆ ligand. (H) Two IP₆ molecules are bound to the R18 ring. Contacts between IP₆ and K25 are broken, and salt-bridge interactions between K25 and E29 reform.

Ligand density maps reveal alternate binding pockets
Transient metastable interactions between the protein and ligand can contribute substantially to the association process by lowering the energetics of intermediate states along the binding pathway.
Prior simulations of other systems have discovered a wide range of dynamical mechanisms used for binding, which do not involve direct interactions between the ligand and the binding site, including protein conformational changes that open up previously inaccessible binding pockets (19–21), the flipping of ligand configurations between several bound poses (15, 22), and metastable electrostatic interactions that guide ligands into the binding pocket (15, 16, 23–25). To assess the protein–ligand interactions that form during our simulations, we compute the spatial distribution occupied by the ligand’s nonhydrogen atoms or, in other words, a 3D map of the ligand density (Fig. 4).

The ligand density maps reveal three sites of interaction between IP6 and the CA domain pentamer (Fig. 4, A and B). The primary binding site (site 1) has the highest density for the ligand and is positioned at the central pore of the pentamer, stabilized by interactions between IP6 and the R18 ring. Intriguingly, the simulations also showed a second lower-affinity binding site (site 2) at an intramolecular interface between the NTD of one CA domain and the CTD of the adjacent CA domain. At site 2 (fig. S2), IP6 binds between helices H3 and H4 of the CA NTD with several phosphate oxygen atoms forming contacts with the nitrogen atoms of N57, Q67, K70, and R173. Three of the four protein–ligand contacts (N57, Q67, and R173) have also been observed in prior x-ray structures of CA hexamers in complex with PF74, an HIV inhibitor (26). In addition, the CA residues that contribute to the protein–ligand interactions form a binding pocket that is structurally well preserved in atomic models of CA lattices [backbone root mean square deviation (RMSD) < 1.2 Å] derived from crystallography and cryo-EM (6, 26, 27). This secondary site is the binding pocket for drug molecules that target the HIV capsid and inhibit viral replication including BI-2 (28), PF74, and GS-CA1 (29). Although the precise mechanisms of action for these compounds are currently unclear, several studies have suggested that small-molecule binding at this site may overly stabilize the capsid and perturb normal capsid disassembly processes, interfere with CA domain assembly processes that form intact capsids, or compete with nuclear import factor binding to the capsid, which is required for integration of the viral genome into the host cell (29–32).

Significant density is also observed at a third site (site 3) proximal to the CTD and consists of interactions between the flexible CA CTD tail and the negatively charged IP6 ligand. At site 3, IP6 phosphate oxygens form contacts with residues H226, K227, and R229 (fig. S2). There are no known drug molecules that currently bind to the CTD tail, which is a cleavage by-product of the retroviral aspartyl protease (e.g., HIV protease) during virion maturation. The aspartic acid moieties in the catalytic site of HIV protease may, in part, explain the affinity of the CTD tail for negatively charged molecules, although an alternative hypothesis is that the electrostatic properties of the CTD tail allow the CA CTD to form interactions with viral RNA that are necessary for recruiting and containing RNA in the mature capsid.

Ligand density maps for the CA domain hexamer are largely similar to that of the pentamer, containing the same sites and interactions with the IP6 ligand (Fig. 4, C and D). One noticeable difference, however, is that the primary site (site 1) in the hexamer is split into two continuous bodies corresponding to binding sites for IP6 above and below the R18 ring at high contour levels for the density (p = 0.5), whereas the map in the pentamer contains only a single density for site 1 above the R18 ring. This is consistent with our result that hexamers can bind two IP6 molecules, whereas pentamers bind only a single IP6 molecule. The ligand densities for sites 2 and 3 are not entirely either five- or sixfold symmetric around the pentamer or hexamer, respectively, owing to challenges in converging sampling statistics for IP6 and CA interactions at each of the binding sites. Overall, shape complementarity is noticeably higher for the IP6 molecule bound to the pentamer compared to the two bound to the hexamer at site 1 (S5 = 603 Å²; SH = 417, 567 Å²) and for site 2 compared to site 3 (S5 = 696 Å²; S3 = 384 Å²; see also figs. S1 and S2). The binding sites discovered from the density analysis, however, agree well with known interactions at the CA domain, such as interactions with the capsid inhibitors (i.e., BI-2, PF74, and GS-CA1) and HIV protease. Our AA MD simulations did not contain prior information on these interactions, highlighting the potential of these simulations to detect novel binding sites for other molecules.

**Free energy landscapes for IP6 binding**

To quantify the energetics of the binding process, we computed the free energy landscape, i.e., the potential of mean force (PMF), for IP6 binding to CA hexamers and pentamers using umbrella sampling simulations, in which a 1D order parameter was used to describe the transport of IP6 to the R18 ring. The 1D order parameter, ξ, is defined as the projection of the displacement between the center of mass of the IP6 ligand and the center of mass of the backbone atoms of the R18 ring onto the vertical pore axis and is shown for the CA hexamer and CA pentamer (Fig. 5, A and B).

Both PMFs (Fig. 5, C and D) show that IP6 binding alleviates significant electrostatic repulsion between the R18 side chains and
In both hexamer and pentamer PMFs, the CA domain from the NTD region, and conformational states in which IP₆ is systematically for the CA hexamer (B). The NTD is colored in green, whereas the CTD is colored in brown. Residues R18, K25, and K29 are labeled. (C) The 1D PMF for the CA hexamer is shown in blue. The shaded light-blue regions denote the error in the PMF determined by block averaging. (D) The 1D PMF for the CA pentamer is plotted in orange. The shaded light-orange regions denote the error in the PMF determined by block averaging. The pore axis is oriented such that positive values correspond to the CA NTD, whereas negative values correspond to the CA CTD.

contain deep free energy minima proximal to the R18 ring (ξ = 0 Å). The hexamer PMF has a broad free energy basin, reflecting a larger pore volume for IP₆ diffusion, and a global free energy minimum positioned at ξ = 3.7 Å. There is a second local minimum in the hexamer PMF at ξ = −0.8 Å corresponding to the binding site below the R18 ring, which is slightly less stable relative to the global free energy minimum (∆G = 0.6 kcal/mol). This is consistent with electron density maps derived from x-ray crystal structures of CA hexamers in complex with IP₆ that revealed stronger densities for binding sites above the ring than below the ring (9, 10). In contrast, the pentamer PMF has a narrower free energy basin owing to a more restricted and compact helix H1 region below the R18 ring in the CA pentamer as compared to that of the CA hexamer and has a single overall free energy minimum located at ξ = 2.7 Å. Ligand entry from the CTD side of the pentamer is unfavorable compared to entry from the NTD region, and conformational states in which IP₆ is below the R18 ring (ξ < 0 Å) are destabilized (∆G > 1.6 kcal/mol). In both hexamer and pentamer PMFs, the CA domain β-hairpin (ξ ~ 5 to 15 Å) did not form high free energy barriers to ligand entry.

As the ligand exits into bulk solvent (ξ > 25 Å and ξ < −20 Å), both PMFs plateau to a constant value as expected, because the protein and ligand no longer interact. Positive values for the order parameter, ξ, correspond to the binding pathway energetics of IP₆ entering the pore from the exterior of the mature capsid (NTD), whereas negative values of ξ correspond to the energetics of IP₆ entering the pore from the interior of the mature capsid (CTD). The total free energy difference of transporting IP₆ to the R18 ring is greater in the pentamer (∆G = 19.3 kcal/mol) than in the hexamer (∆G = 16.4 kcal/mol), indicating that IP₆ binding stabilizes CA pentamers more than CA hexamers in the fullerene shell. The calculated binding free energies are consistent with that of four to six Arg-phosphate interactions, which have higher stabilities (33–35) than that of Arg-Glu interactions [∆Gₚ = −4 to 5 kcal/mol for a single pair (36, 37)]. Hexamer configurations can form an additional Arg-phosphate interaction compared to pentamer configurations; however, the smaller pentamer pore facilitates closer coordination of IP₆ phosphate groups by R18 side chains. Free volume corrections to account for finite-size effects also lower the effective equilibrium dissociation constants derived from the PMFs by approximately 4 kcal/mol (38). Negative-stained electron microscopy studies demonstrated that removal of the charged arginine ring by an R18A mutation forms CA assemblies with spherical, fullerene, and spiral structures, hypothesized to occur from an increased frequency of pentamer incorporation into the CA hexagonal lattice (5, 39). Although the R18A mutation removes the primary site of CA interaction with IP₆, the mutation also reduces electrostatic repulsion between CA subunits at R18, analogous to the effect of IP₆ binding. Overall, our findings suggest that IP₆ preferentially binds to and stabilizes CA pentamers over CA hexamers as a result of the more compact and closely packed H1 pore helices in the pentamer and a subsequent greater electrostatic repulsion at the R18 ring. Given the higher number of hexamers in the capsid lattice, a large amount of IP₆ in total should be found among the hexamers.

**Fig. 5. The PMF for IP₆ binding.** The order parameter, ξ, is used to describe IP₆ transport to the R18 ring. ξ is defined as the projection of the center-of-mass displacement between IP₆ and R18 residues onto the pore axis and is shown schematically for the CA hexamer (A) and the CA pentamer (B). The NTD is colored in green, whereas the CTD is colored in brown. Residues R18, K25, and K29 are labeled. (C) The 1D PMF for the CA hexamer is shown in blue. The shaded light-blue regions denote the error in the PMF determined by block averaging. (D) The 1D PMF for the CA pentamer is plotted in orange. The shaded light-orange regions denote the error in the PMF determined by block averaging. The pore axis is oriented such that positive values correspond to the CA NTD, whereas negative values correspond to the CA CTD.

**DISCUSSION**

Collectively, our results suggest that CA hexamer and pentamer building blocks of the fullerene capsid bind IP₆ with rather different binding modes. At high IP₆ concentrations (i.e., [IP₆]:[CA] = 1:3), hexamers can bind IP₆ at a secondary site below the R18 ring, whereas at low IP₆ concentrations ([IP₆]:[CA] = 1:6), IP₆ alternates between the two binding sites, although the site above the R18 ring is lower in free energy (∆G = −0.6 kcal/mol). Experimental estimates using radiolabeled IP₆ ligands suggest that ~300 IP₆ molecules are packaged per viral particle, which contains ~1500 CA domains, indicating that the latter case is more likely to occur (7). The molecular pathways by which IP₆ binds involve a conformational twist in the pentamers, which may also be aided by the inclusion of pentamers at highly curved regions of the capsid surface. Binding below the R18 ring in pentamers is less favorable because pentamers have a compact and tighter pore relative to that of hexamers. Ligand density calculations show that interactions at the R18 ring are consistent with the differential binding modes and reveal two additional sites of IP₆ interaction, one of which is a previously known capsid inhibitor binding pocket (site 2) and the other consists of an interaction between the CTD tail and negatively charged molecules that could reflect capsid and RNA interactions (site 3). The two weaker affinity sites are metastable, and similar sites found in AA MD simulations of other systems have been reported to form higher-affinity binding sites for different ligands (22, 40). Radiolabeled ligand binding and mutagenesis experiments could further investigate protein-ligand interactions at these sites.

The computed PMFs demonstrate that IP₆ binding stabilizes individual pentamers more than individual hexamers. Since CA spontaneously assembles into both helical tubes and fullerene cones in vitro (41), and pentameric “defects” (5, 6, 42) are required to produce fullerene cones but are not present in helical tubes, this result indicates that the binding of IP₆ shifts the equilibrium population of CA.
hexamers and pentamers toward increased pentamer formation, thereby promoting the fullerenelike mode of assembly. It is of note to recognize that the pathways discussed characterize the binding of IP$_6$ to preformed mature capsid complexes and not the molecular pathways for the co-assembly of CA and IP$_6$, which may differ. Further simulations that target the co-assembly of CA and IP$_6$ will likely require coarse-grained molecular simulation techniques to overcome the long time scale barriers associated with biomolecular assembly phenomena (43, 44). The computed thermodynamic quantities, however, are state functions that do not change in response to the particular path taken, so the IP$_6$ interactions revealed by the ligand density maps are therefore general. Experimental strategies using engineered disulfide cross-links to stabilize hexamer or pentamer constructs (39, 45) combined with binding assays may also uncover additional thermodynamic quantities relevant to IP$_6$ binding, and AA simulations of large-scale CA lattices bound to IP$_6$ could reveal how IP$_6$ affects global structural properties, such as lattice flexibility.

Since inositol phosphates play important physiological roles in promoting and stabilizing the assembly of the HIV-1 capsid, could the removal of IP$_6$ be involved in disassembly processes? One possible mechanism is that acidic conditions simultaneously trigger protonation of IP$_6$ phosphate groups and H12, thus reducing the affinity of IP$_6$ for the R18 ring and opening of the capsid pores. Although pH has been known to trigger the release of genetic material in viruses by acidification of endosomal vesicles during intracellular transport (46), it is unclear whether HIV capsids are exposed to similarly low pH conditions during infection in either endocytic or membrane fusion pathways (47). In addition, IP$_6$ can chelate divalent metal cations, and a potential mechanism for nucleotide entry into capsid pores blocked by IP$_6$ is for free nucleotides to transfer associated divalent cations to IP$_6$ to unblock the pore.

In this work, we have elucidated the molecular mechanisms responsible for the stabilization of mature capsids by IP$_6$, through extensive AA MD simulations. Additional investigations can further probe how retroviruses exploit environmental cues to initiate replication processes.

### MATERIALS AND METHODS

#### AA models for the CA hexamer, pentamer, and IP$_6$

The initial atomic models for the CA hexamer and pentamer were constructed from the x-ray crystal structures for the CA hexamer in the closed and open conformations (PDB ID: 3H47 and 5HGL, respectively) and the cryo-EM structure for the CA pentamer in the closed conformation (PDB ID: 5MCY). A homology model for the open CA pentamer was constructed on the basis of aligning individual CA domains and transplanting the β-hairpin gate (residues 1 to 16) from the open CA hexamer onto the closed pentamer. Missing amino acid residue backbones were built using MODELLER (48), and missing side chains were built using SCWRL4 (49). The hexamer and pentamer systems, which contained a total of 287,048 and 286,715 atoms, respectively, were solvated and neutralized by adding Na$^+$ and Cl$^-$ ions to the bulk solution until the salt concentration was 150 mM NaCl. Periodic boundary conditions were imposed on an orthorhombic unit cell of approximately 141 Å × 141 Å × 141 Å for both the hexameric and pentameric systems. The AA potential energy function CHARMM36m (50, 51) for proteins and the TIP3P (52) potential energy function for water were used. Electrostatic interactions were computed using the particle mesh Ewald algorithm, and short-range, nonbonded interactions were truncated to 12 Å. Harmonic restraints were applied to the center of masses of two groups of residues (residues 179 to 189 and residues 200 to 207) in α-helices, H9 and H10, at the CTD interface between distinct hexamer and pentamer subunits to maintain the relative position of the CTD interface derived from an atomic model of the complete HIV-1 capsid (6). The relative center-of-mass positions for the two groups of residues were computed after RMSD-aligning individual CA hexamers and pentamers of the capsid and used as the equilibrium positions for the harmonic restraints to mimic the average curvature in the capsid, along with a light force constant of $k = 0.5$ kcal/mol. The solvated system was energy minimized and equilibrated under constant pressure and temperature (NPT) conditions at 1 atm and 310 K with a 2-fs time step. Minimization and equilibration procedures were performed using the AA MD simulation package NAMD 2.13 (53).

#### MD simulations for IP$_6$ binding

IP$_6$ molecules were initially placed at random positions and orientations in bulk solvent, approximately 10 Å away from any nonwater molecules. Fifteen IP$_6$ molecules were used to increase sampling times. The IP$_6$ was parameterized using the CHARMM General Force Field (CGenFF) (54). For each CA domain system, the histidine residues, H12, were completely protonated or deprotonated. Salt concentration was adjusted for the ligand charge to maintain 150 mM NaCl and an electrostatically neutral system. After a 50-ns equilibration run, eight systems consisting of hexamer/pentamer, open/closed, and H12 protonated/deprotonated combinations were prepared for simulation on Anton 2 at the Pittsburgh Supercomputing Center (14). For all production simulations, bond lengths for hydrogen atoms were constrained using the M-SHAKE algorithm (55). An rRESPA integrator was used with a time step of 2 fs (56); long-range electrostatics were computed every 6 fs. Long-range electrostatics were calculated using the k-space Gaussian split Ewald method (57). Short-range interactions, including van der Waals and short-range electrostatics, were truncated at 9 Å. Simulations in the constant NPT ensemble were performed using a chained Nosé-Hoover thermostat at 310 K and isotropic Martyna-Tobias-Klein barostat at 1 atm (58). Each trajectory (four CA hexamers and four CA pentamers) was simulated until no additional binding events were observed. In aggregate, total simulation time was 47 μs for the CA hexamer and pentamer systems (see table S1).

#### Ligand density calculations

The trajectories from production-level simulations were subsampled at 0.12-ns intervals and aggregated into pentamer or hexamer systems. For each frame, the atomic system was aligned by minimizing the RMSD of the protein backbone atoms with respect to the crystallographic structure. Cartesian coordinates for the nonhydrogen atoms of IP$_6$ were recorded, and the density of atomic positions $ρ(7)$ was calculated, using a hard-sphere van der Waals approximation on a discretized grid with a spacing of 0.5 Å × 0.5 Å × 0.5 Å. The 3D ligand density map was subsequently contoured at $ρ = 0.05, 0.11,$ and 0.50 for the CA pentamer systems and $ρ = 0.02, 0.09,$ and 0.05 for the CA hexamer systems.

#### Umbrella sampling

The free energy landscape (PMF) of IP$_6$ binding was computed using an umbrella sampling strategy (59). All systems contained a
single IP₆ molecule, either a CA hexamer or pentamer in the open conformation, and a protonated H12 residue. A 1D order parameter, ξ, is used to describe the relative distance to the R18 ring. ξ specifies the projection of the displacement vector between the center of mass of the IP₆ ligand (r) and center of mass of the backbone atoms of the R18 ring (r₆₁₈₈) onto the z axis [e.g., ξ = (z - r₆₁₈₈)]. The umbrella sampling windows consisted of CA domain complexes and the IP₆ ligand positioned in 0.2-Å increments along ξ. The initial coordinates were obtained using targeted MD simulations in NAMD, using the solvated CA hexamer or pentamer systems. Three hundred sixty umbrella sampling windows were used to compute the PMFs. Harmonic biasing potentials with a force constant of 20 kcal/mol per Å-centered on ξ were used. All other restraints and interactions were computed as described during the equilibration phase. The total simulation time for the CA hexamer is 28.8 µs. The total simulation for the CA pentamer is 28.8 µs. Each PMF, W(ξ), was computed using the weighted histogram analysis method (WHAM) to unbias and recombine the sampled distribution functions from all windows (60, 61). The statistical uncertainty in each PMF was evaluated using the approach of block averaging (62). For each PMF, the time series was divided into five blocks, and WHAM was used to calculate a PMF from the data in each block. The standard deviation of the five PMFs was reported. Using 5 to 15 blocks, all gave qualitatively similar results.

**SUPPLEMENTARY MATERIALS**

Supplemental material for this article is available at http://advances.sciencemag.org/cgi/content/full/6/38/eabc6465/DC1

**REFERENCES AND NOTES**


Acknowledgments: Funding: This work was supported by NIH grant P50 AI150464 for the Center for the Structural Biology of Cellular Host Elements in Egress, Trafficking, and Assembly of HIV. Computational resources were provided by Frontera at the Texas Advanced Computer Center funded by the NSF grant (OAC-1818253). Anton 2 computer time was provided by the Pittsburgh Supercomputing Center (PSC) through grant R01 GM116961 from the NIH. The Anton 2 machine at PSC was made available by D.E. Shaw Research. A.Y. acknowledges support from the National Institute of Allergy and Infectious Diseases of the NIH under grant F32 AI150208. Author contributions: A.Y. and G.A.V. designed research. A.Y. performed research. A.Y. and E.M.Y.L. contributed methods, simulation code, and analytic tools. A.Y., E.M.Y.L., J.J., and G.A.V. wrote the paper. A.Y. and G.A.V. designed research. A.Y. performed research. A.Y. and E.M.Y.L. contributed methods, simulation code, and analytic tools. A.Y., E.M.Y.L., J.J., and G.A.V. wrote the paper. Competing Interests: The authors declare that they have no competing interests. Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 6 May 2020
Accepted 20 July 2020
Published 16 September 2020
10.1126/sciadv.abc6465