


Structure of the mature Rous sarcoma virus lattice reveals a role for IP6 in the formation of the capsid hexamer

Martin Obr¹, Clifton L. Ricana ², Nadia Nikulin³, Jon-Philip R. Feathers³, Marco Klanschnig¹, Andreas Thader¹, Marc C. Johnson², Volker M. Vogt³, Florian K. M. Schur ¹✉ & Robert A. Dick³✉

Inositol hexakisphosphate (IP6) is an assembly cofactor for HIV-1. We report here that IP6 is also used for assembly of Rous sarcoma virus (RSV), a retrovirus from a different genus. IP6 is ~100-fold more potent at promoting RSV mature capsid protein (CA) assembly than observed for HIV-1 and removal of IP6 in cells reduces infectivity by 100-fold. Here, visualized by cryo-electron tomography and subtomogram averaging, mature capsid-like particles show an IP6-like density in the CA hexamer, coordinated by rings of six lysines and six arginines. Phosphate and IP6 have opposing effects on CA in vitro assembly, inducing formation of T = 1 icosahedrons and tubes, respectively, implying that phosphate promotes pentamer and IP6 hexamer formation. Subtomogram averaging and classification optimized for analysis of pleomorphic retrovirus particles reveal that the heterogeneity of mature RSV CA polyhedrons results from an unexpected, intrinsic CA hexamer flexibility. In contrast, the CA pentamer forms rigid units organizing the local architecture. These different features of hexamers and pentamers determine the structural mechanism to form CA polyhedrons of variable shape in mature RSV particles.

¹Institute of Science and Technology (IST) Austria, Klosterneuburg, Austria. ²Department of Molecular Microbiology and Immunology, University of Missouri, Columbia, MO, USA. ³Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA. ✉email: florian.schur@ist.ac.at; rad82@cornell.edu

Late in the retrovirus life cycle, the structural protein Gag assembles into an immature lattice at the inner leaflet of the cellular plasma membrane. As the virus buds from the cell, the viral protease cleaves Gag into several fragments, including the capsid protein (CA) that then assembles into the mature capsid shell (core), rendering the virus particle infectious. Experiments based on X-ray crystallography, nuclear magnetic resonance spectroscopy, and helical reconstruction have provided information about the structure of CA in isolation and in highly regular tubular arrays^{1–5}. In stark contrast, authentic retrovirus particles are pleomorphic, lacking size, and shape uniformity, rendering these techniques not applicable. Cryo-electron tomography (cryo-ET) with subtomogram averaging has proven highly effective in addressing the structure of pleomorphic immature and mature retroviral lattices, optimally resulting in ~4 Å structures^{6–9}. An emerging generalization from all of these experiments is that both the immature Gag lattice and the mature CA lattice are based on hexamers of the protein subunits, which exploit different CA interfaces¹⁰. In most retroviruses, the mature hexameric lattice is closed by incorporation of exactly 12 pentamers, and complete core closure is generally thought to be important for infectivity of retrovirus particles¹¹. However, not all retroviral cores form completely closed shells¹². The CA portion of all retroviral Gag proteins folds into two separate domains, the N-terminal domain (CA_{NTD}) and C-terminal domain (CA_{CTD}). In mature lattices, CA_{NTD} interactions form only intra-hexamer and intra-pentamer interfaces, while CA_{CTD} interactions are also involved in formation of inter-hexamer and inter-pentamer interfaces.

For viruses in the *Lentivirus* genus, HIV-1 and SIV are highly sensitive to inositol hexakisphosphate (IP6) for immature viral lattice formation and virus release in cells and in vitro, respectively, and HIV-2, FIV, BIV, and EIAV (immunodeficiency viruses from human, simian, feline, bovine, and equine species, respectively) are sensitive to IP6 for the formation of the immature Gag lattice in vitro^{13–15}. IP6 is bound to and thereby stabilizes the Gag lattice by interacting with two rings of six lysine residues at the CA–SP hexamer interface. For HIV-1, the mature CA lattice also is influenced by IP6, which binds to a ring of six arginine residues in the N-terminal domain of the CA hexamer. In cells, IP6 is known to be critically important for HIV-1 and SIV replication, since genetic manipulation of cells to remove IP6 dramatically reduces the release of infectious virus particles¹³. In vitro assembly of Gag and CA protein into immature virus-like particles (VLPs) and mature capsid-like particles (CLPs) has facilitated study of the lattice structure of many retroviruses¹⁶. In this methodology, the viral proteins are purified after expression in *Escherichia coli*, incubated under defined conditions of pH, ionic strength, temperature, and in some cases with an oligonucleotide or other additions, and then observed by negative stain electron microscopy. The structures of these VLPs and CLPs have been shown to accurately mimic those of bona fide virus particles^{7,17}.

Using in vitro assembled particles, we here describe the role of IP6 in the mature structure of Rous sarcoma virus (RSV), an alpharetrovirus that has been used widely as a model system for HIV-1 and for studying aspects of cancer biology. Our results show that, like lentiviruses, RSV is sensitive to IP6 in vitro and in cells, where in the mature CA lattice IP6 binds in a pore formed in the center of the CA_{NTD} hexamer. Moreover, our subtomogram averaging analysis of pleomorphic polyhedral CASPNC CLPs reveals that a remarkable, previously unobserved flexibility of mature CA hexamers accommodates the wide range of curvatures intrinsic to mature RSV particles, and that CA pentamers represent rigid building blocks determining local lattice geometry.

Results

IP6 is an RSV assembly cofactor. Like all Gag proteins, RSV Gag comprises multiple domains (Fig. 1a), but as shown by the well-studied truncated protein called GagΔMBDΔPR, only the p10, CA, SP, and NC domains are required for the efficient immature virus particle assembly in vitro^{18,19}. In immature assembly in lentiviruses, conserved lysine residues in the major homology region (MHR) and CA–SP1 helix bind IP6 (ref. 15; Fig. 1a). These residues, however, are not present at the corresponding positions in RSV Gag. Consistent with this lack of sequence conservation, RSV GagΔMBDΔPR supports efficient in vitro assembly of immature VLPs in the absence of IP6 (Fig. 1b). At a physiological 50 μM concentration of IP6, assembly was not affected. However, at higher concentrations, IP6 led to the decreased or abrogated assembly. As described below, this inhibition of immature assembly may be due to the strong effect of IP6 on mature assembly.

RSV CASPNC, a construct similar to GagΔMBDΔPR, but missing the p10 domain that is known to be critical for immature lattice formation, assembled into mature long narrow tubes, consistent with previously published results²⁰. The addition of IP6 to this protein resulted in a dramatic change in assembly morphology, from tubes to (nonuniform) polyhedrons that resemble mature cores in the infectious virus particles²¹ (Fig. 1c). These polyhedral CLPs appeared first in significant numbers at 20 μM IP6, and then dominated the population of CLPs as the IP6 concentration was raised to 100 μM. Consistent with this effect of IP6 on mature assembly, RSV CA has lysine and arginine residues in helix 1 of CA, located near the site of the HIV-1 R18 residue that interacts with IP6 at the mature CA_{NTD} hexamer pore (Fig. 1a).

Previously, we showed that the depletion of IP6 from HEK293FT cells by CRISPR-Cas9 knock-out of inositol-pentakisphosphate 2-kinase (IPPK), the enzyme that converts IP5 to IP6, caused a ~20-fold reduction in the production of infectious HIV-1 particles. In the case of RSV, infectious RSV production was reduced ~100-fold from HEK293FT IPPK-KO compared to WT without the need for further genetic manipulation (Fig. 1d). These results show that IP6 plays an important role for infectious particle production of RSV, but do not clarify at what step or steps IP6 acts.

To delineate how the removal of IP6 impacts RSV, we performed western blot analyses of whole cell lysates, as well as the supernatants to measure relative levels of viral protein produced and virus particles released (Fig. 1e and Supplementary Fig. 1). Quantification showed that Gag synthesis was reduced by ~3-fold in HEK293T IPPK-KO cells compared with WT HEK293T cells (Fig. 1e, left). Virus release was reduced by ~50-fold in IPPK-KO cells (Fig. 1e, right). We interpret these results to mean that RSV may be partially dependent on IP6 for Gag translation in cells, but the major effect of IP6 removal is to reduce the assembly and release of virus particles from cells. These results are the first to indicate IP6 as an assembly cofactor of a retrovirus outside of the *Lentivirus* genus.

Structure determination of the mature RSV lattice. We sought to structurally characterize the mature RSV CA lattice to better understand how IP6 influences mature assembly. We acquired cryo-ET data of CASPNC CLPs assembled in the presence of 100 μM IP6. The reconstructed tomograms of this sample showed abundant regular tubes and pleomorphic polyhedral CLPs. (Fig. 2a and Supplementary Movie 1). Conventional subtomogram averaging, as described previously¹⁴, of tubular CLPs with regular diameter and shape resulted in a reconstruction of the CA hexamer at 4.3 Å resolution (Fig. 2b, left, and

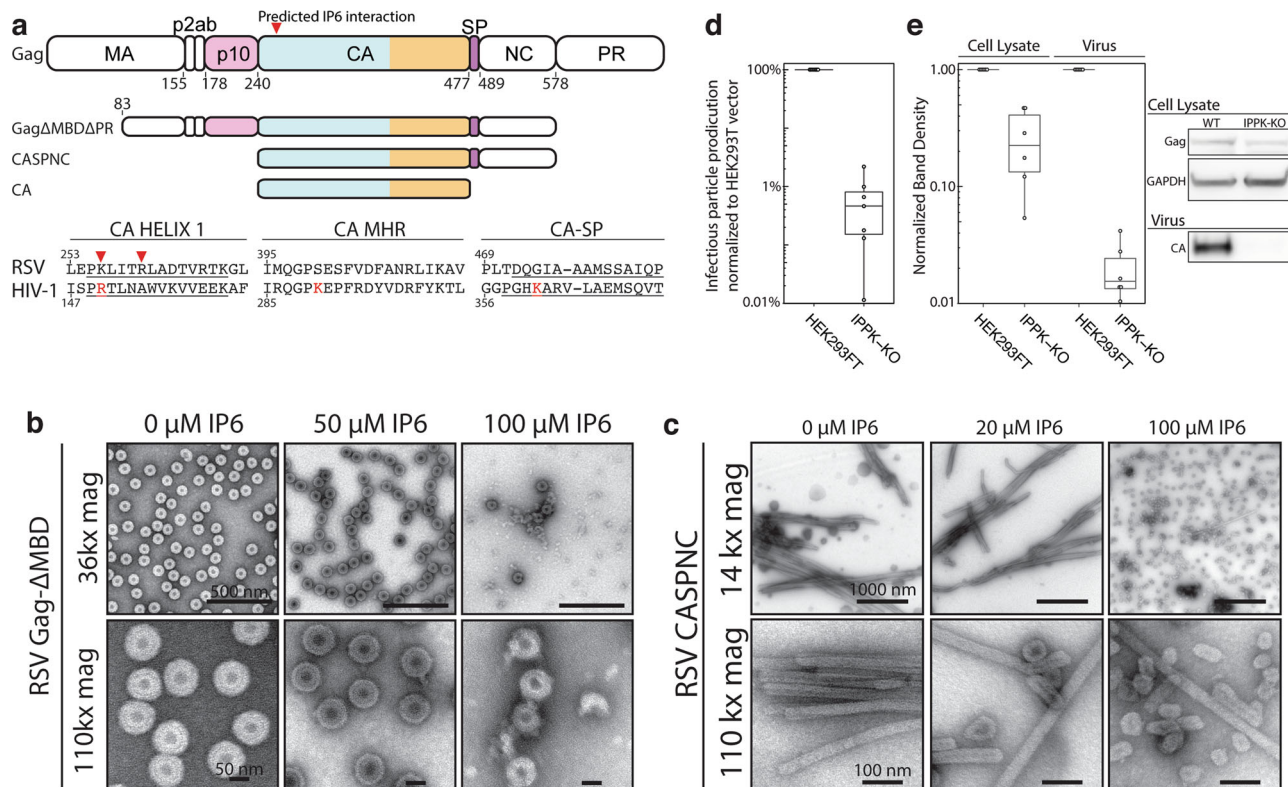


Fig. 1 Effect of IP6 on RSV in vitro and in cells. **a** RSV Gag and truncation proteins used for assembly and structure determination. Alignment of RSV and HIV-1 amino acid sequences. Known HIV-1 IP6-interacting residues shown in red and predicted RSV IP6-interacting residues indicated (red arrow). **b** TEM images of immature RSV Gag- Δ MBD assembly without and with IP6 at low and high magnification. **c** TEM images of mature RSV CASPNC assembly without and with IP6 at low and high magnification. **d** RSV infectious virus particle production in wild type (HEK293T) and IPPK-KO cell lines. **e** (left) Western blot normalized band intensity, and (right) representative western blots of cell lysate and GAPDH control (top) and virus fractions (bottom). Uncropped western blots are shown in Supplementary Fig. 1. Box plots: center lines show the medians, box limits indicate the 25th and 75th percentile; whiskers extend to minimum and maximum values; data points are plotted as circles. Source data are provided with this paper.

Supplementary Figs. 2a and 3a). The EM density map revealed a similar arrangement of the CA_{NTD} and CA_{CTD}, as previously reported for other mature retroviral CLPs (Supplementary Movie 2). Given the inherent twofold symmetry of the tubes, CA monomers within tubes are present in three symmetry-independent copies. In the center of the CA_{NTD} hexamer, an additional density is present in the identical position, which has previously been reported as an IP6-binding site in mature HIV-1 (ref. 15). No ordered density for NC and also no ordered density for the β -hairpin, which has been observed in other retroviral mature CAs, were observed. At the resolution of 4.3 Å, we were able to refine a model of RSV CA²² into the EM density of the mature hexameric CA assembly. At this resolution, we refrain from unambiguously interpreting side chain rotamer positions. As reported for other mature retrovirus CLPs, intra-hexameric interactions are established via CA_{NTD}-CA_{NTD}, as well as CA_{NTD}-CA_{CTD} interfaces, while the inter-hexameric interactions are formed solely via CA_{CTD}-CA_{CTD} interactions maintaining contacts across twofold and threefold interfaces^{1,3,9,22}. Notably, an IP6 molecule could be fitted into the density in the CA_{NTD} hexamer center in an upright position, where it is coordinated by two concentric rings of positively charged residues K17 and R21 (Fig. 2d and Supplementary Movie 2).

Structural determinants of pleomorphic mature RSV CA lattices. To further understand how RSV CA is able to generate both regular tubular and pleomorphic polyhedral CLPs, and in particular how the different inter- and intra-hexameric CA interfaces allow for the observed flexibility, we also performed

subtomogram averaging on polyhedral CLPs. Most important, we determined the structure of the mature CA pentamer in these polyhedrons to define the interactions between hexameric and pentameric RSV CA assembly states.

The highly variable size and shape of RSV CA polyhedral CLPs, ranging from small icosahedrons and polyhedrons to several micrometer long tubes (Fig. 2a), resulted in a variable and limited local order of the CA lattice. A conventional subtomogram averaging approach (Supplementary Figs. 2a and 5a) assumes a high degree of rigidity and symmetry of the CA oligomeric units (i.e., hexamers and pentamers) and an extended local order of the CA lattice, which implies that the variations in the local curvature come primarily from varying interfaces between CA hexamers or pentamers. However, as conventional subtomogram averaging of polyhedral RSV CASPNC CLPs only yielded a structure of the CA hexamer with limited resolution, despite a significant number of asymmetric units, we concluded that these described geometric conditions were not met in RSV polyhedral CLPs, and that hexamers and their surrounding show significant flexibility and variability.

Hence, in order to appropriately address the plasticity of the polyhedron CASPNC lattice and any potential conformational flexibility within individual CA hexamers, which could be associated with their different local environments (such as a varying curvature), we devised a subtomogram averaging approach that uses classification and alignment guided by local geometry and contextual information (Supplementary Figs. 2 and 4, and see also the “Methods” section). In order to classify hexamer and/or pentamer (unit) pairs according to their

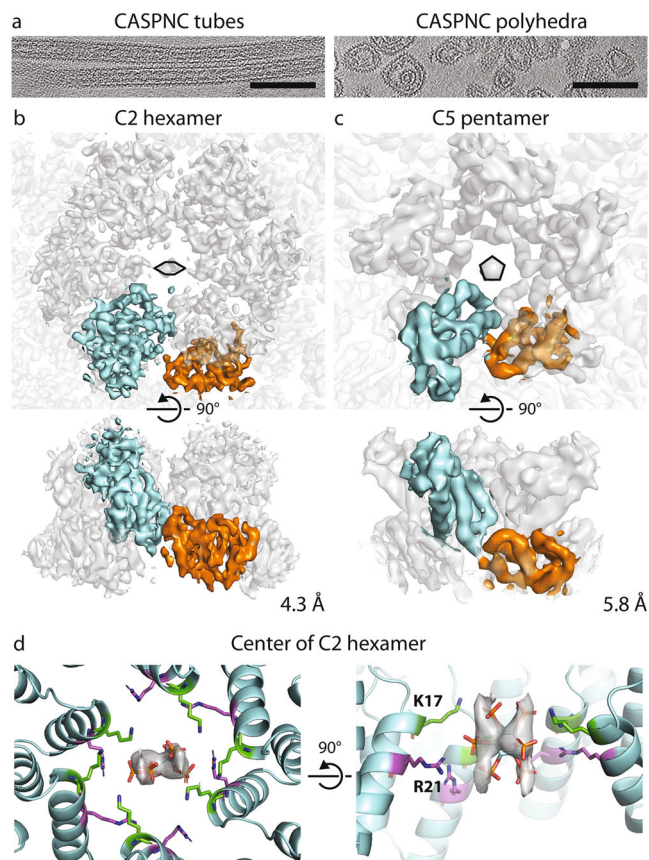


Fig. 2 Cryo-electron tomography of RSV CASPNC CLPs assembled in the presence of IP6. **a** Sum of ten computational slices through a gaussian-filtered tomogram containing CASPNC tubes and CASPNC polyhedrons. Scale bar is 100 nm. The slices are representative of the 49 tomogram dataset. **b, c** Isosurface representation of the subtomogram average of a C2-symmetric CA hexamer from the CASPNC tubes (**b**) and a C5-symmetric pentamer from CASPNC polyhedrons (**c**). In both cases, the CA_{NTD} and CA_{CTD} of one CA monomer are colored cyan and orange, respectively. The C2 and C5 symmetry of the hexamer and pentamer are annotated by a distorted schematic hexamer and a pentamer, respectively. The resolution of the individual structures is annotated. See corresponding FSC curves in Supplementary Fig. 3a. **d** Model of the IP6 binding site in the CA-NTD hexamer pore formed by amino acids K17 (green) and R21 (purple), as seen from the outside of the CLP (left) and rotated by 90° (right). The side chains are shown with their most likely rotamer position in agreement with the EM density.

geometrical context, we grouped them using two criteria that we found to influence the geometry of the mature CA lattice: (1) the presence and location of one or more pentamers in the vicinity of a hexamer (Supplementary Fig. 4a), and (2) the orientation of a pair of hexamers with respect to the local curvature (Supplementary Fig. 4b). According to these criteria all hexamer and/or pentamer pairs were classified into different classes belonging to six general groups (Supplementary Fig. 4c).

For determining the structure of the asymmetric parts of lattice, we implemented an alignment strategy respecting the asymmetry of each unit given by its local environment. Specifically, instead of treating hexamers and pentamers within the RSV CA lattice as isolated symmetric units and using a single alignment against a common reference, the alignment of each hexamer or pentamer in the mature lattice is decomposed into several independent alignments (Fig. 3a, b, Supplementary Fig. 5b, and see “Methods” section for further details). In brief, each hexamer and pentamer is

aligned multiple times (i.e., six times in case for a hexamer that is having six neighbors), each time using a reference specific to its local context, and hence geometry. The consensus of these separate alignments then determines the position and orientation of the respective oligomeric unit (Supplementary Fig. 5b). In the final step, an additional alignment centered in the middle of the unit pair (i.e., the dimer interface) was performed, in order to optimize the structures of different classes with respect to the interfaces between the unit pairs (Supplementary Fig. 2b).

This approach yielded 21 distinct structures at resolutions of 6–9.1 Å of RSV hexamer/pentamer, hexamer/hexamer, and pentamer/pentamer pairs that exist in mature RSV CA lattice (Fig. 3b, Supplementary Fig. 3b, and Supplementary Movie 3) and also yielded a structure of the CA pentamer at 5.8 Å (Fig. 2c and Supplementary Fig. 3a). For further analysis, we rigid-body fitted individual CA domains into the EM density maps of hexamer/pentamer unit pairs. These structures and models indeed confirmed that the mature polyhedron CA lattice consists of structurally different, intrinsically flexible CA hexamers that strongly deviate from regular sixfold symmetry. In other words, the hexamers need to accommodate varying interactions and curvatures to form the observed highly pleomorphic polyhedrons (Fig. 3a, b). Hexamers that are separated from pentamers by two or more hexamers, are essentially identical to those found in regular tubes. Pentamers, on the other hand, organize rigid vertices, where the lattice is curved in different directions. The hexamers that are adjacent to pentamers are deformed to allow bridging of the tube-like areas and the pentamer vertices (Fig. 3a, b).

Further analysis of the different hexamers forming the mature CA lattice in polyhedral CLPs revealed that despite the deformability of the CA hexamer, local symmetries exist in the lattice. Aside from the C2-symmetric hexamers that are also observed in regular tubes (and are surrounded by only hexamers, Fig. 2b, Fig. 3b group V), a fraction of hexamers in the polyhedrons were surrounded by either two or three pentamers in a symmetrical manner and hence followed C2 symmetry and C3 symmetry, respectively (Supplementary Fig. 6a, b). As the structure of the tube-like hexamer with C2 symmetry was already solved at high resolution from regular tubes, it was not further characterized. In contrast, the structures of the other two-symmetric hexamers highlight the structurally dominant influence of the pentamer, where it actively defines the conformation of its adjacent hexamers.

Specifically, the two CA monomers facing a pentamer adopt an arrangement similar to those two CA monomers in the pentamer, causing an increased spatial separation between the CA_{NTD}'s of the hexamer (Supplementary Fig. 6a–d and Supplementary Movie 4). This observation was further confirmed via the structure determination of a small set of $T=3$ icosahedral particles solved by single-particle cryo-EM (Supplementary Figs. 3c and 6e). In $T=3$ icosahedrons each hexamer is surrounded by three pentamers, which causes the described threefold-symmetric distortion of the CA hexamer. The similarity of the C3 hexamer from the $T=3$ icosahedra (formed by CA) to the C3 hexamer from the polyhedrons (formed by CASPNC and IP6; Supplementary Fig. 6f, g) suggests that the structural deviation from C6 symmetry is not due to the presence of either the IP6 molecule, nor the absence of the β -hairpin, nor the SPNC region.

The finding of the strong plasticity of the mature RSV CA hexamer led us to compare the variability of the intra-hexamer and intra-pentamer CA_{NTD}–CA_{NTD} and CA_{NTD}–CA_{CTD} interfaces, as well as the dimeric and trimeric CA_{CTD} interfaces in our structures using pairwise RMSD analysis. (Fig. 3c, d). In agreement with the observed deformability of the CA_{NTD} ring

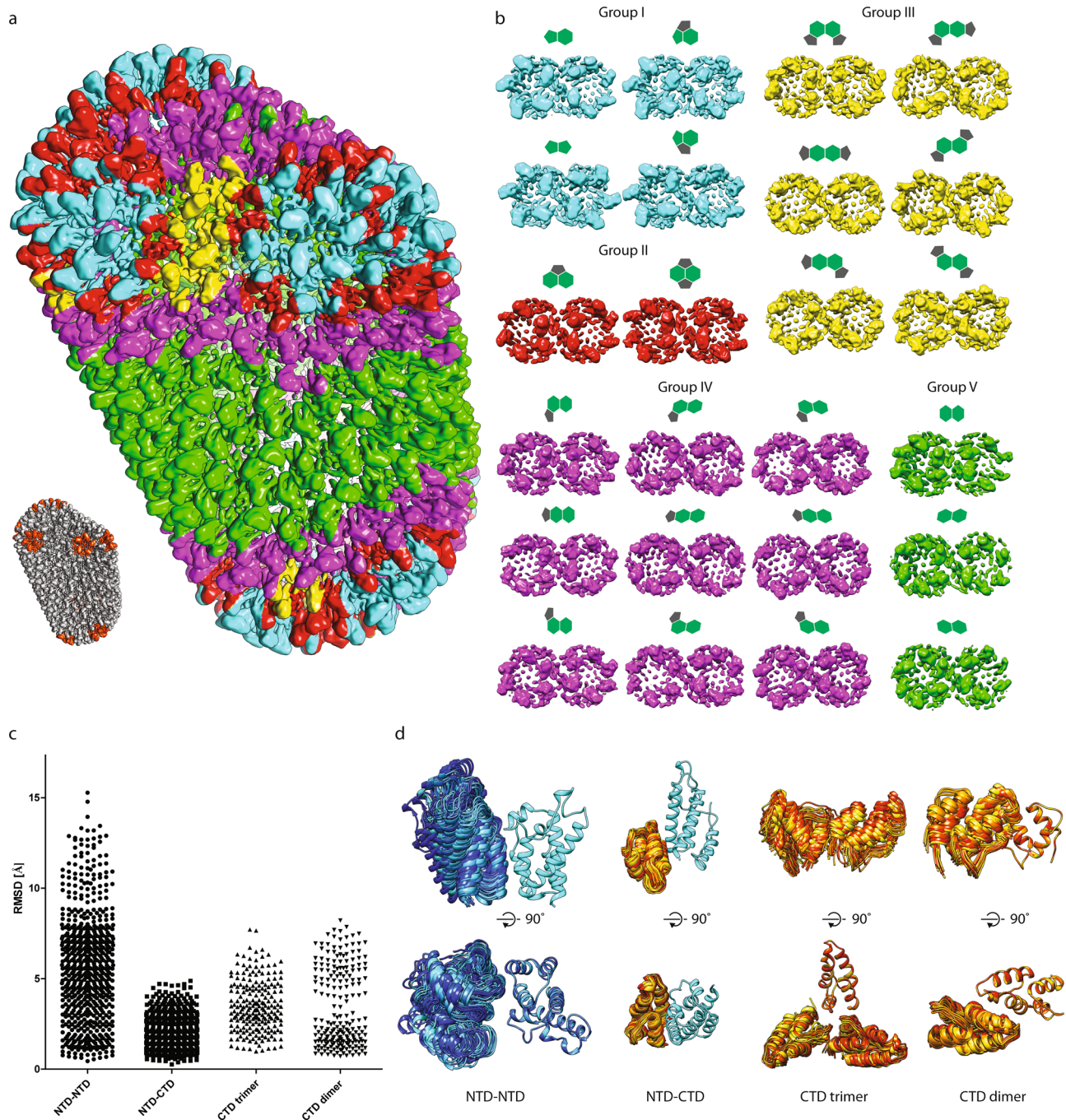


Fig. 3 Building blocks of the mature CASPNC lattice. **a** A composite model of a representative RSV CASPNC CLP, generated by placing respective unit pair averages (see groups in **b**) into their positions derived from subtomogram alignment. The models shown in **a** and in Supplementary Movie 3 are identical. **b** Final averages of unit pairs filtered to 8 Å. Groups are identical to their schematic depiction in Supplementary Fig. 4. Coloring of the groups corresponds with **a**. **c** Pairwise RMSD measurements between rigid-body fitted models derived from the maps of the different classes shown in **b**. In case of $CA_{NTD}-CA_{NTD}$, $CA_{NTD}-CA_{CTD}$, and CA_{CTD} dimer interfaces, one CA domain is fixed and RMSD is measured for the other CA domain across the respective interface. For the CA_{CTD} trimer interface one CA domain is fixed and RMSD is measured for the other two CA_{CTD} 's forming the trimer interface. **d** Superimposition of the models showing the flexibility of CA domains involved in the different interfaces. Color code: cyan to blue— CA_{NTD} ; yellow to red— CA_{CTD} . Source data are provided with this paper.

in hexamers, $CA_{NTD}-CA_{NTD}$ interfaces displayed the largest variation. In contrast, $CA_{NTD}-CA_{CTD}$ interactions are locking the two domains in a quaternary arrangement that does not deviate extensively between the different classes. Two populations of CA_{CTD} dimer interfaces (termed here as state A and B) were observed (Fig. 3d, Supplementary Fig. 7, and Supplementary Movie 5). State A was found in all dimer interfaces involving

pentamers (with higher curvature) and low curvature hexamer-hexamer interfaces, whereas state B was found only in hexamer-hexamer interfaces with high curvature. Specifically, comparison of high-resolution structures of $T=1$ particles (see below) and one hexamer-hexamer interface from regular tubes revealed striking similarity (Supplementary Fig. 7a and Supplementary Movie 5). The changes of the dimer interface in the two

states had impact also on the CA_{CTD} trimer interface, and induced its deformation from C3 symmetry. (Fig. 3c, d).

This interface analysis explains how the rigid CA_{CTD}-CA_{NTD} interface acts as a structural link to propagate the changes in the CA_{CTD} dimer interface toward the hexameric CA_{NTD} ring to distort the weak interactions across the CA_{NTD}-CA_{NTD} interface.

IP6 shifts phosphate-induced CA CLP assembly from pentameric to hexameric. The addition of IP6 to CASPNC assembly reactions caused a dramatic shift from tubular to polyhedral CLPs. Since polyhedrons contain both hexamers and pentamers, while tubes contain only hexamers, a possible interpretation of this result is that IP6 promotes pentamer formation. An alternative interpretation is that IP6 promotes initiation of assembly, which ultimately results in rapid depletion of CA. This rapid depletion of CA might favor polyhedron assembly. To shed light on these possibilities, we performed in vitro assembly of CA protein with IP6 and/or NaPO₄. High concentrations of NaPO₄ are known to induce RSV CA to assemble predominantly into $T=1$ icosahedral CLPs, which consist of only 12 pentamers. If IP6 promotes pentamer formation this small molecule would be expected also to induce assembly of icosahedral CLPs. On the other hand, if IP6 promotes hexamer formation, we predicted that it would promote tubular CLPs. The results of this experiment showed that in absence of IP6 or NaPO₄, CA appeared in negative stain EM images as small aggregates and unassembled protein (Fig. 4a), but in the presence of 100 μM IP6, CA assembled into multilayered tubes. Raising the level of IP6 to 500 μM led to a transition to single-layered tubes (Fig. 4a). Higher concentrations of IP6 resulted in protein aggregation. We interpret this result to mean that IP6 favors the formation of CA hexamers over CA pentamers.

High phosphate concentrations were previously predicted to affect assembly in vitro by shielding the charges of the K17 and R21 of the CA pentamer^{23,24}, allowing these side chains to come into close proximity in the CA pentamer. In the presence of 200 μM NaPO₄, CA assembled into some multilayered tubes (CA hexamers), but also considerable aggregates (Fig. 4b). In presence of 300 mM NaPO₄, CA formed predominantly short tubes with curved ends, which we interpret to reflect the presence of CA pentamers. At 400 mM NaPO₄, $T=1$ icosahedral CLPs were observed mixed with blunt-ended tubes. The number of $T=1$ CLPs increased as NaPO₄ was increased, and at a concentration of 700 mM nearly all of the particles were small icosahedral CLPs. If phosphate favors pentamer formation and IP6 favors hexamer formation, we speculated that the addition of IP6 to assembly reactions that normally result in icosahedrons could shift the resulting CLPs from predominantly icosahedral to predominantly blunt tubes. While at 500 mM NaPO₄ CA formed predominantly small icosahedral CLPs, the addition of 2 μM IP6 resulted in an observable shift to fewer icosahedrons and more polyhedrons and blunt-ended tube CLPs (Fig. 4c). At 20 and 100 μM IP6 the majority of CLPs were polyhedrons. At 200 μM IP6, the majority of CA assembled into blunt-ended tubes. Based on these results, we conclude that IP6 promotes assembly by interacting primarily at the CA hexamer interface.

Based on the in vitro assembly and structural data, IP6 and NaPO₄ likely promote assembly via interactions with K17 and/or R21 amino acid side chains. To verify this, CASPNC protein with either K17A or R21A mutations was purified and tested for assembly. Compared to WT (Fig. 1c) CASPNC K17A formed fewer tubes, and many small polyhedral CLPs in the absence of IP6 (Supplementary Fig. 8a). The addition of IP6 decreased the number of polyhedrons, with assembly favoring tubes. CASPNC R21A protein was assembly incompetent both without and with

IP6 (Supplementary Fig. 8a). We also tested the effect of the K17A mutation on CA assembly. In the absence of IP6 or NaPO₄, CA K17A formed some aggregates and thin-walled assemblies (Supplementary Fig. 8b). Assembly in the presence of IP6 resulted in the formation of small (~10 nm) cross-structures, which we interpret to be a regular CA complex that does not support mature lattice formation. Assembly reactions performed in the presence of NaPO₄ resulted in narrow tubes (~10 nm in diameter) and the same cross-like structures. Both K17A and R21A mutations abrogated immature assembly of the GagΔMBDΔPR protein (Supplementary Fig. 8c), likely due to the mutation disrupting an interface critical to the immature lattice. Taken together, we interpret these results to mean that the charge of K17 and its neutralization (by IP6 or NaPO₄) are critical factors regulating RSV assembly.

CA pentamer structure from the RSV $T=1$ icosahedron. The observation that within pleomorphic polyhedrons the CA pentamer constitutes the rigid building block prompted us to investigate if the pentamers within $T=1$ icosahedrons and polyhedrons are structurally identical.

Hence, we acquired cryo-EM micrographs of $T=1$ icosahedral CLPs and subjected them to single-particle analysis, yielding a 3.1 Å resolution reconstruction (Fig. 5a-c). At this resolution refinement of RSV CA into the EM density was unambiguously possible (Fig. 5d). We compared the model of $T=1$ pentamers with the model of the pentamer derived from the polyhedrons. In fact, the two models are virtually superimposable, showing an RMSD of their C-alpha atoms of 0.74 Å (Fig. 5e). This finding further underscores the observations that the RSV CA pentamer (1) is very rigid due to the tighter packing of the CA domains in the pentamer, (2) always promotes the same local geometry, and (3) is independent in arrangement of the size and shape of the CLP.

Discussion

IP6 is an assembly cofactor for HIV-1 and other lentiviruses^{13-15,25}. Here, we have demonstrated that IP6 has a potent effect on the assembly and morphology of the mature RSV CA lattice, and that depletion of IP6 from cells inhibits virus release and spread. This is the first study to demonstrate that IP6 is an assembly cofactor for a virus not in the *Lentivirus* genus. In addition, we have determined the structure of the mature RSV CA hexamer and pentamer at the highest resolution to date, using cryo-ET and cryo-EM, respectively, to reveal structural mechanisms underlying RSV mature core formation.

The effect of IP6 on mature RSV assembly is unambiguous. The interaction of IP6 with K17 and R21 is similar to the interaction of IP6 with R18 in HIV-1 (ref. 15). For HIV-1, the IP6 molecule sits above the top of the pore at residue R18 (but can also be found below the R18 residue, effectively in the pore), whereas for RSV IP6 is found in the pore interacting with residues K17 and R21. Most retroviruses have an arginine or lysine residue in the first helix of CA, corresponding to K17 and R18 of RSV and HIV-1, respectively. The charged pore of the mature hexamer is thought to be used for dNTP import, with IP6 acting there to overcome the repulsive charge of closely positioned arginine and/or lysine residues during assembly, as proposed by Cardone et al. and Keller et al.^{23,24}. The only orthoretroviruses that do not have basic residues in helix 1 belong to the *Gammaretrovirus* genus, which includes murine leukemia virus (MLV). From a recent cryo-ET structure of the mature MLV lattice, it appears that unlike lentiviruses and alpharetroviruses, gammaretroviruses do not form a closed capsid shell¹², thus possibly allowing dNTP import through gaps in the lattice. If this

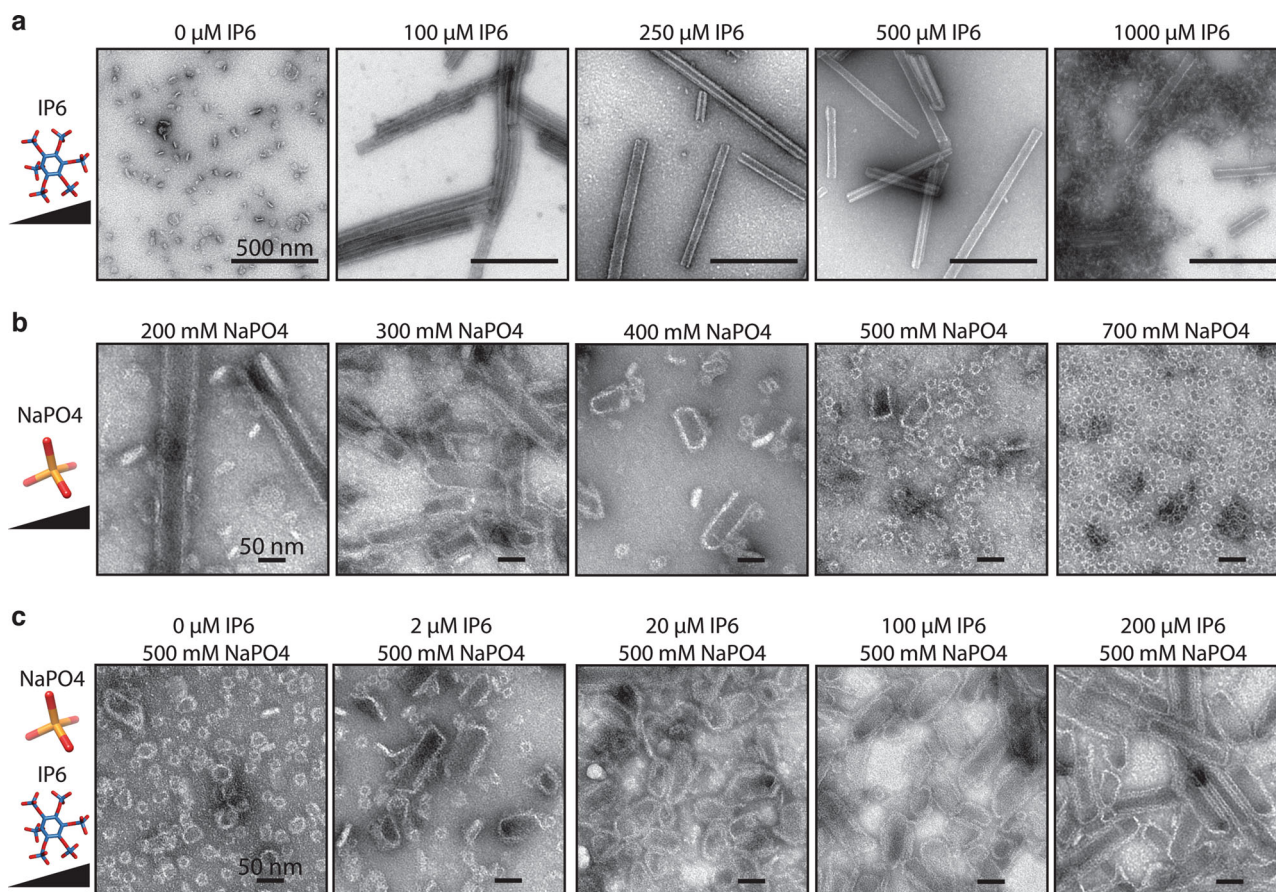


Fig. 4 Effect of NaPO₄ and IP6 on in vitro CA assembly. **a** Example TEM images of CA assembly reactions with increasing concentrations of IP6. **b** Example TEM images of CA assembly reactions with increasing concentrations of NaPO₄. **c** Example TEM images of CA assembly with fixed 500 μM NaPO₄ and increasing concentration of IP6.

model is correct, IP6 would be important for the assembly of retroviruses from many different genera.

While IP6 significantly enhances the immature assembly of lentiviruses^{13–15}, we did not observe a similar effect for RSV. The lentiviral immature IP6-binding site consists of two rings of six lysine residues each. For HIV-1, this is K290 (located in the MHR) and K359 (located at the C-terminal end of CA, in the first turn of the helix that forms the six-helix bundle). RSV does not have basic residues located at or near its MHR or the putative six-helix bundle. Nevertheless, we found that depletion of IP6 from cells reduces virus particle release. Thus, IP6 either directly influences immature assembly via a Gag–IP6 interaction, or indirectly influences assembly or release via some yet to be determined mechanism.

For HIV-1, IP6 binding to the immature interaction site at K290 and K359 in the Gag lattice ensures packaging of IP6 into the budding virus particle. Following maturation, IP6 thus is already present to promote mature assembly via interactions with the mature hexamer at R18. If RSV does not directly require IP6 for immature assembly, how is IP6 incorporated into the virus particle? Sensitivity to IP6 during mature assembly was 100–1000-fold greater for RSV than for HIV. Perhaps a small amount of IP6 is taken into the budding RSV particle non-specifically, and this amount is sufficient for the formation of the mature lattice. Alternatively, perhaps in RSV proteolytic maturation of Gag occurs during budding, allowing IP6 to be incorporated by the growing mature CA lattice before the viral membrane has been sealed. The genome structure of RSV is

unusual in that the viral protease is part of the Gag polyprotein, and possibly allowing PR to become active earlier than for other retroviruses.

Retroviruses from different genera adopt varying quaternary CA_{NTD} arrangements in the immature Gag lattice¹⁰. In contrast, no significant deviations have so far been reported for quaternary CA arrangements in mature retroviruses. The most notable differences instead were observed on the global architecture of the mature CA lattices, with HIV-1 forming cone-shaped cores with a defined incorporation of pentamers²⁶ and MLV forming multilayered, spiral structures with irregular pentamer distribution¹². In authentic RSV virions mature CA cores form highly pleomorphic structures, either having continuous “spherical” curvature, or tubular and fullerene-like cone shapes²¹. This remarkable core polymorphism is recapitulated by our mature RSV CASPNC and CA in vitro assembly, which exhibits a range of differently sized CLPs. The exact mechanism of how mature retroviral CA assemblies are able to form architectures with such a large variety of different curvatures and molecular interactions, is not yet entirely understood. Recent studies analyzing mature HIV-1 CA tubes suggested that the plasticity of the CA hexamer, specifically its deviation from sixfold symmetry, defines curvature adaptation in mature cores, without significantly altering the CA_{CTD} dimer and trimer interfaces^{4,5}.

Our results, obtained via a classification and alignment approach accounting for asymmetry of CA units in polyhedron assemblies, now extend these findings and provide a more detailed understanding of the dynamic CA properties that underlie curved retroviral CA lattices.

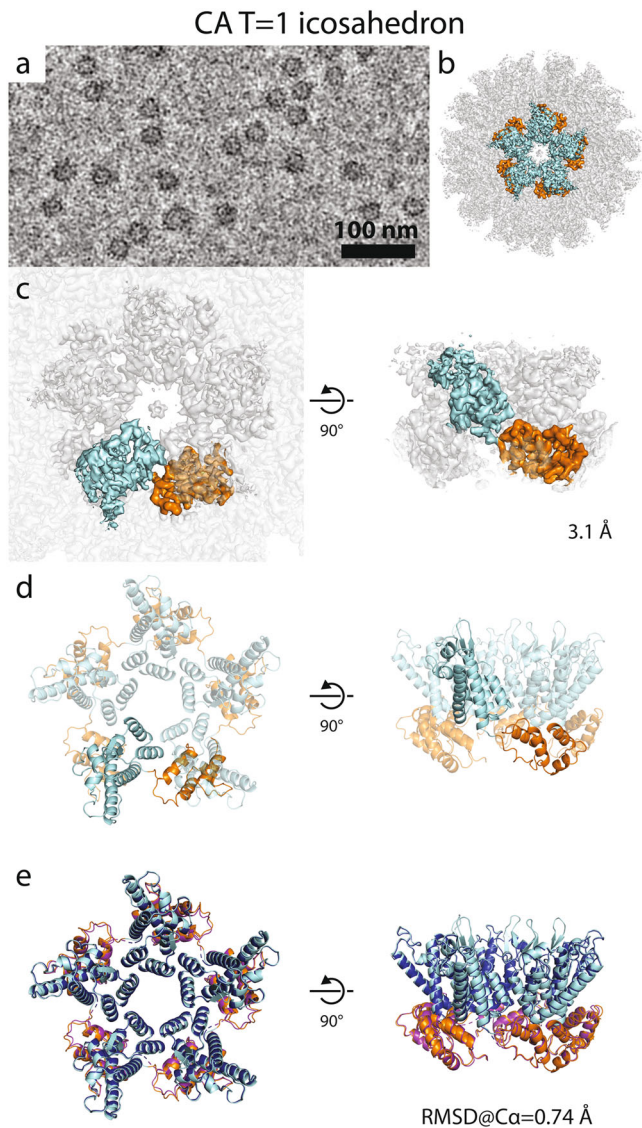


Fig. 5 Structure of the RSV $T = 1$ icosahedron pentamer. **a** Filtered cryo-EM micrograph of $T = 1$ icosahedrons. The image is representative of the 2394 micrograph dataset. **b** Isosurface representation of the CA $T = 1$ icosahedron determined by single-particle cryo-EM. One pentamer is highlighted in colors: cyan—CA_{NTD}; orange—CA_{CTD}. **c** Isosurface representations of one $T = 1$ pentamer from the structure shown in **b**. The density in the center of the pentamer, coordinated by K17 and R21, may correspond to phosphate ions, which are the main negatively charged component of the assembly buffer. **d** Model of the RSV CA pentamer refined from the structure shown in **b**. **e** Comparison of the real-space refined model of the CA pentamer from $T = 1$ particles and rigid-body fitted model of CA pentamer from polyhedrons. The low RMSD value of their backbone atoms shows that the two models are almost identical.

In line with the observations made in HIV-1 (ref. 4), mature RSV CA forms pseudo-symmetric, intrinsically curved hexamers. However, due to the highly pleomorphic shape of our CLPs, the structural plasticity of the RSV hexamer is significantly more pronounced, resulting in a large structural variety of hexamers adopting twofold and threefold-symmetric assemblies, as well as asymmetric structures, which are distorted by local curvature and geometrical context. Hence, while hexamers are building blocks of any mature CA lattice, their intrinsic structure and flexibility needs to vary to adapt different requirements in mature core formation. The observed increased flexibility of the intra-

hexameric contacts corresponds with increased rigidity of inter-unit interfaces. However, the degree of curvature variability in the mature lattice still requires different connections between the units. Hence, instead of having a continuum of different inter-hexameric and pentamer interactions gradually adjusting the orientation between two units, two populations of the CTD dimer interface are observed. In contrast to the pliability of the RSV CA hexamer, the CA pentamer represents a rigid building block that acts as the local organization center that defines hexamer shape and, depending on the position on the lattice, overall core architecture. A similar lattice organization featuring plastic hexamers, invariant pentamer geometry, and two distinct CA_{CTD} dimer populations has been previously reported for in vitro assembled icosahedral CLPs derived from the human endogenous retrovirus HML2 (ref. 27), suggesting conserved mature CA assembly mechanisms for building fullerene-like structures of a wide range of shapes (from icosahedra to polyhedrons) and sizes.

The increased deformability of hexamers in RSV is different to what has been previously found in HIV-1. For example, when compared to RSV, the ~ 9 Å resolution structure of HIV-1 pentamer from authentic viral cores⁹ does not reveal any deviation of its surrounding hexamers from true sixfold symmetry (Supplementary Fig. 9a), suggesting an overall lower flexibility of the HIV-1 hexameric CA unit in the mature assembly. Cone-shaped cores, as predominantly found in HIV-1, are determined by their distribution of pentamers, with seven pentamers on the wide and five pentamers on the narrow end. An increased rigidity of HIV-1 hexamers could result in accumulated strain and tension around pentamers, resulting in their nonrandom distribution on mature cores, where pentamers locate to positions of defined curvature at the cone ends. This geometrical limitation could result in the more homogenous core size and shape distribution in HIV-1. In contrast, the increased flexibility of the RSV hexamer allows a more random distribution of pentamers in mature RSV cores. This enables RSV CA to even form inter-pentameric contacts, which can be regularly observed on polyhedrons and in particular $T = 1$ icosahedral particles. No such pentamer-pentamer contacts have been yet observed for HIV-1 indicating that the described geometrical limitations in HIV-1 might render such interfaces unfavorable.

Further work on authentic CA cores or pleomorphic mature CLPs for other retroviruses will be required to determine the full spectrum of intra- and inter-hexameric CA interactions, and if the rigidity of the pentamer as organizing center is a conserved feature of mature retroviral CA assemblies. Taken together, these findings shed light on the role for IP6 in the regulation of mature RSV assembly, and also on the structural principles underlying retroviral mature core formation and shaping.

Methods

Protein purification and in vitro assembly. GagΔMBDΔPR and CASPNC proteins were purified using the SUMO-tag system²⁸, as previously described in ref. 14. The resulting SUMO purified proteins have an ectopic serine residue, which we previously reported does not influence assembly^{13,15}. RSV CA protein was expressed in *E. coli* and purified using standard affinity and size-exclusion chromatography (SEC) described briefly here. Bacterial pellets were resuspended in buffer (50 mM Tris-HCl pH 8 and 2 mM TCEP), lysed by sonication, and the lysate cleared by ultracentrifugation. Nucleic acid was precipitated by the addition of 0.03% (v/v) polyethyleneimine followed by centrifugation. Ammonium sulfate was added to $\sim 20\%$ saturation of the supernatant, and the precipitate was pelleted by centrifugation. The pellet was resuspended in buffer (50 mM Tris-HCl and 2 mM TCEP), and flowed through tandem QFF and SP columns (GE). CA protein was present in the flow through, and was $\sim 95\%$ pure. Flow through was subjected to SEC through a Superdex 75 column (GE). Peak fractions corresponding to the CA protein were concentrated to 15–20 mg/mL, flash frozen in liquid nitrogen, and stored at -80 °C.

In vitro assembly of GagΔMBDΔPR was performed by dilution assembly, as previously described²⁹. CASPNC was assembled via dialysis described briefly here. A total of 30 μ L containing 50 μ M protein in storage buffer (20 mM Tris-HCl pH 8,

500 mM NaCl, and 2 mM TCEP) supplemented with $\sim 5 \mu\text{M}$ GT25 oligo, IP6 (if present) was dialyzed against assembly buffer (20 mM MES pH 6.2, 100 mM NaCl, 2 mM TCEP, IP6 (if present)) at 4 °C for 4 h. Gag Δ MBD Δ PR and CASPNC were spotted onto EM grids, negative stained with uranyl acetate, and imaged via TEM. In vitro assembly of CA with IP6 and NaPO₄ was done by dilution assembly. Protein was diluted into assembly buffer (20 mM Tris-HCl pH 8 and 2 mM TCEP) supplemented with NaPO₄ or IP6, to a final concentration of 290 μM . Samples were incubated at 22 °C for 30 min and stored at 4 °C until spotted onto TEM grids.

Cell culture, virus release, and infectivity. HEK293FTs and IPPK-KO cells were maintained, and plated for transfections and transductions, as previously described¹³. RSV particles for cellular assays were produced and cells transduced in a similar manner as for HIV, as previously described¹³. The flow cytometry gating strategy is shown in Supplementary Fig. 10. Western blots and their analysis were performed as previously described, and briefly here¹³. A rabbit anti-RSV-capsid antibody (prepared in-house) was used at a 1:500 dilution in 5% nonfat dry milk in PBS-Tween20 for the 1 h primary application at room temperature. A goat anti-rabbit peroxidase-conjugated antibody (Sigma, A0545) was used at a 1:10,000 dilution in 5% nonfat dry milk in PBS-Tween20 for the 1 h secondary application at room temperature. Per the manufacturer's protocols, membranes were stripped with Restore PLUS Western Blot Stripping Buffer (Thermo, 46430), re-blocked, incubated with a mouse anti-GAPDH antibody (Santa Cruz Biotechnology, SC-47724) at a 1:500 dilution, washed, incubated with goat anti-mouse-HRP (Sigma, A5278) at a 1:10,000 dilution, and imaged. Blot images were converted to 8-bit format, Fiji's (ImageJ) gel analysis tools were used to calculate blot densities, and values were exported to a CSV file for subsequent analysis in RStudio.

Single-particle cryo-EM. For cryo-EM 500 μM CA was assembled in 1 M NaPO₄ at pH 8. The assembly reaction was diluted 1:4 to 250 μM NaPO₄, and 3 μL were spotted onto glow discharged (45 s, 20 mA) 2/2-3C C-flat grids. Samples were vitrified in liquid ethane using a Vitrobot Mark 4 (blot time of 2.5 s and a blot force of 0). Imaging was done at 200 kV on a Thermo Fisher Talos Arctica TEM equipped with a BioQuantum post column energy filter and Gatan K3 direct detector, using the SerialEM software package³⁰. The data were collected with a nominal magnification of 63,000 \times in super-resolution mode with a physical pixel size of 1.25 Å/pixel. A total of 50 frames were captured as movies with a total dose of $50 e^{-}/\text{Å}^2$. Image processing was done in RELION 3.1 (ref. 31), and software was maintained by SBGrid³². Specifically, motion correction was done using MOTIONCOR2 (ref. 33) in the RELION wrapper. CTF estimation and correction was performed using GCTF³⁴ in the RELION wrapper. For $T = 1$ structure determination, 2394 micrographs were taken, from which 374 particles were manually picked and 2D classified to generate templates for auto-picking. Two rounds of auto-picking and 2D classification resulted in 150,599 extracted particles. These particles were refined against EMD-5772 (ref. 24; low-pass filtered to 60 Å) to a resolution of 4.6 Å. Bayesian particle polishing, further iterative refinement, 2D and 3D classification, and CTF refinement resulted in a final set of 21,498 particles that produced a 3D reconstruction with a masked 0.143 (Fourier shell correlation) FSC resolution of 3.1 Å. $T = 3$ CLPs were selected from the same initial 2D classification job to generate templates for auto-picking. Three rounds of auto-picking and 2D classification resulted in 2309 particles. Subsequent rounds of 2D and 3D classification resulted in a set of 599 particles refined using EMD-5773 (ref. 24) low-pass filtered to 60 Å as a starting model to 8.6 Å. Bayesian particle polishing, further iterative refinement, 2d and 3d classification, and CTF refinement resulted in a set of 406 particles with a masked 0.143 estimated resolution of 7.6 Å.

Cryo-electron tomography. Cryo-ET sample preparation and data collection were done essentially as described previously^{7,14}. 10 nm colloidal gold was added to CASPNC CLPs and 2.5 μL of this solution was then applied to degassed and glow discharged 2/2-2C C-flat grids. The samples were vitrified in liquid ethane using a Vitrobot Mark 4 (blot time of 2 s, wait time of 5 s, blot force 0) and stored in liquid nitrogen until imaging.

Tilt series were acquired on an FEI Titan Krios, operated at 300 keV, equipped with a Gatan Quantum 967 LS energy filter and a Gatan K2xp direct electron detector, using the SerialEM software package³⁰. The slit width of the filter was set to 20 eV. Areas of interest for high-resolution data collection were identified in low magnification montages. Prior to tomogram acquisition, gain references were acquired and the filter was fully tuned. Microscope tuning was performed using the FEI AutoCTF software³⁵.

The nominal magnification was 105,000 \times , resulting in a pixel size of 1.328 Å/pixel. Tilt series were acquired using a dose-symmetric tilt-scheme³⁶, with a tilt range from 0° to -66° and +63° in 3° steps. Tilt images were acquired as 8k \times 8k super-resolution movies, consisting of ten frames. The dose rate was set at $\sim 2.5 e^{-}/\text{Å}^2/\text{s}$ and aiming for a target dose of $\sim 156 e^{-}/\text{Å}^2/\text{tilt series}$. Tilt series were collected at nominal defocus between -1.5 and -4 μm . Data acquisition information is also provided in Supplementary Table 1.

Cryo-ET image processing. K2 super-resolution movies were aligned on-the-fly during data acquisition using the SerialEMCCD frame alignment plugin and tilt series were automatically saved as 2 \times binned mrc stacks. Prior to further

processing, tilt series were sorted and bad tilts (e.g., images with large shifts or objects blocking the beam) were removed, using MATLAB scripts.

CTF estimation was performed using CTFFIND4 (version 4.1.10)³⁷ on each tilt individually. Images were low-pass filtered according to their cumulative electron dose³⁸. Tilt series alignment of the exposure-filtered micrographs was performed in the IMOD software package³⁹ and 3D-CTF-corrected tomograms were reconstructed by weighted back-projection using novaCTF⁴⁰. The CTF correction method was multiplication with astigmatism correction and a Z-slab thickness of 15 nm. The resulting tomograms were consecutively binned two, four, and eight times via Fourier cropping.

In total 49 tomograms were used for further processing. Subtomogram averaging was performed in Dynamo⁴¹. Additional intermediate processing steps and steps for context-based classification (as described below) were done using MATLAB scripts, employing functions from Dynamo, AV3 (ref. 42), and TOM⁴³ software packages. Complete statistics for the sizes of the different datasets are given in Supplementary Table 2.

Subtomogram averaging of regular CASPNC tubes. In order to determine the structure of the CA hexamer from regular tube CLPs, subtomograms extracted along the tubular surface were subjected to iterative averaging and alignment, as previously described¹⁴. Briefly, initial subtomogram positions and Euler angles were set based on tube radius and splines fitted to the respective tubular axes, using MATLAB and 3dmod software³⁹.

Initially, a single tube, derived from a tomogram with a nominal defocus of -2.5 μm was chosen to derive a de novo starting reference at bin8. This reference was then used for the alignment of the entire tube CLP dataset consisting of 236 tubes.

Bin8 starting extraction positions on tube surfaces were oversampled ~ 4 times. After the first round of bin8 alignment, all subvolumes that had converged onto the same position of the tubular lattice were removed using a subvolume-to-subvolume distance cutoff. Subvolumes that contained no protein density or did not align against the reference were removed based on a cross-correlation (CC) threshold.

In total, ten rounds of alignment were performed, progressively reducing the angular search range, adapting low-pass filters and reducing binning. At every unbinning step (e.g., going from bin8 to bin4) subtomograms were re-extracted from the respective unbinned tomograms at positions determined in the previous binned alignments. New averages were then generated using Euler angles determined in the previous binned alignment round. C2 symmetry was applied throughout the entire processing of this tubular dataset.

At bin2, data were split into even/odd half-sets, which from hereon were treated completely independently. Up to bin2 alignment low-pass filter settings were restricted to 32 Å, ensuring that even/odd datasets were not aligned on identical features beyond this resolution.

Final bin1 averages displayed varying anisotropy due to a preferential orientation of the tubes in the tomograms with respect to the tilt axis. To compensate for the anisotropy in the averages, weighted averaging of the 1 \times binned data were performed using modified wedge masks^{14,44}. Mask-corrected FSC, using gaussian-filtered masks, was employed to determine the resolution at the 0.143 criterion. The half maps were then combined, sharpened, and filtered to the measured resolution⁴⁵.

Lattice annotation of polyhedrons. In order to initially annotate the CA lattices from all CLPs (in particular irregular polyhedrons, but also again tubes), an in-house written MATLAB script for template matching was used, in part employing functions from the Dynamo⁴¹, AV3 (ref. 42), and TOM⁴³ software packages. Based on previous observations of mature retroviral CA assemblies³⁹, the CA lattice of RSV CASPNC CLPs was assumed to consist of C2-symmetric CA hexamers (as seen in our regular tube CLPs), C6-symmetric CA hexamers, and C5-symmetric CA pentamers (in polyhedral CLPs). Therefore, three different templates following these symmetries were generated in a similar fashion as described above for the C2-symmetric hexamer, but with the following differences. Subvolumes were first aligned in bin8. Subsequently, three additional bin4 alignment rounds allowed sorting of subtomograms into pentamers (C5) and hexamers (C2 and C6), to generate three new averages for these CA assembly types, which were then used as templates for template matching on bin8 tomograms.

For template matching, a cylindrical mask encompassing the central CA pentamer or hexamer and one ring of adjacent hexamers was applied to the template. Both the reference and tomogram were band-pass filtered (425 Å-30 Å) before CC calculation. Phi and non-Phi angles were scanned over a range of 360° in 10° and 15° steps, respectively. To select the highest scoring CC peaks from the resulting CC array, a lattice connectivity analysis was performed. The following criteria were used to assess pairs of local CC maxima determined for hexamers/pentamers to define a maximum-connectivity network (which is forming a distinct CLP): minimum spacing between hexamers/pentamers = 60 Å, maximum spacing between hexamers/pentamers = 110 Å, minimum curvature (i.e., difference between the tangential plane and a vector connecting the respective pair of peaks) = -15°, maximum curvature = 40°, maximum difference of normal vectors corresponding to the respective CC peaks = 45°. Only networks containing >20 CC peaks for either of the three used templates were considered as CLPs and used for subsequent steps.

Context-based classification of CA subtomograms. The irregularity of the CA lattice in polyhedron CASPNC assemblies limited the achievable resolution using conventional subtomogram averaging (as described for the regular tube CLPs). To overcome this limitation, subvolumes were classified according to their local geometrical context. Specifically, pairs of adjacent pentamers/hexamers (so called unit pairs, where a unit refers to a pentamer or hexamer, resulting in possible unit pairs of hexamer/hexamer, hexamer/pentamer, or pentamer/pentamer) were sorted into classes based on their local lattice arrangement.

To this end, two classification criteria were used, based on the following observations of our data:

- i) pentamers within polyhedrons and hexamers from tubes are regular objects with defined C5 and C2 symmetry, respectively (Fig. 2a-c);
- ii) the hexamers in the vicinity of a pentamer have varying geometry (Supplementary Fig. 6);
- iii) hexamers, which are further distant from a pentamer, adopt a more tube-like geometry with C2 symmetry.

The first criterion for dimer classification is the position of pentamer(s) in a given unit pair. Each hexamer adjacent to a pentamer has all of its dimeric connections numbered relative to the position of the pentamer (Supplementary Fig. 4a), resulting in numbers 1'-6' for the first unit (i.e., unit A), and 1''-6'' for the second (i.e., unit B). Number 1' or 1'' indicates that one unit from the dimer is pentamer. Numbers 2' (which is the same as 6'') or 6' (which is the same as 2'') indicate that a pentamer neighbor is shared between unit A and B. Numbers 3-5 indicate a pentamer that is directly in the vicinity of only one unit of the pair.

The second criterion for dimer classification is the orientation of the unit pair with respect to the tube geometry (Supplementary Fig. 4b). In a regular tube each hexamer is surrounded by six C2 hexamers (except for the tube ends). There are therefore three distinct orientations of two neighboring hexamers with respect to the tube geometry. This criterion was used for classification of hexamer-hexamer dimers, in which at least one did not neighbor with any pentamer.

Using these criteria, unit pairs were sorted into six groups, which were then further divided into subclasses (Supplementary Fig. 4c). Group 1 comprises dimers, in which one or both of the respective units are pentamers (in positions 1' and 1''). For the initial consensus alignment, any further pentamers in positions 2-6, were not considered during the classification. For the final round of consensus alignment, however, additional surrounding pentamers in positions 2-6 (with respect to the hexamer) were used to subclassify the hexamer-pentamer connections. Group 2 comprises two neighboring hexamers, which share one or two pentamer neighbors in positions 2 and 6. Similarly, as in the previous group, further pentamers in positions 3'-5' are only considered in subclasses in the final round of consensus alignment. Group 3 encompasses dimers of hexamers, in which each hexamer has exactly one pentamer neighbor in positions 3, 4, or 5. Group 4 contains dimers, in which only one hexamer has a pentamer neighbor (positions 3, 4, or 5). Since the geometry of the second hexamer is not constrained by a pentamer in its vicinity, the dimer orientation with respect to the tubular geometry is used to further sort particles (Supplementary Fig. 4c). Group 5 contains dimers of hexamers without pentamers in position 2-6 (i.e., in regular CASPNC tubes not having pentamers). Again, classification according to the orientation with respect to the tube geometry was used to yield three subclasses. Group 6 comprises dimers, in which hexamer A contains pentamers in two out of three positions: 3', 4', and 5', and hexamer B does not have any pentamer in its vicinity. The angle between the dimer and the tube axis is in this case neglected, since further classification would yield very sparsely populated classes. This context-based classification laid foundation for the consensus alignment, as specified below.

Consensus alignment of tubular and polyhedral CASPNC CLPs. Instead of refining the positions and orientations of isolated hexamer and pentamer units, an alignment of unit pairs classified into one of the six groups was performed. To this end, for each unit pair a set of alignments, always masking the central and one of the adjacent units, was performed. Hence, the number of alignments N per unit corresponded to the number of its directly surrounding units in the lattice, yielding $N(x, y, z)$ coordinates and $N(\phi, \theta, \psi)$ Euler angle triplets. The N set of alignments was then used to determine a consensus alignment for a unit depending on its neighbors. The consensus position of the unit center was calculated as weighted mean of the (x, y, z) coordinate sets (Eq. (1)).

$$(X, Y, Z) = \frac{\sum_{i=1}^N c c_i^* (x_i, y_i, z_i)}{\sum_{i=1}^N c c_i} \quad (1)$$

Where x_i, y_i, z_i and X, Y, Z are the positions of a given particle in Cartesian coordinates for different measurements and their consensus, respectively, and $c c_i$ is CC.

The consensus orientation was calculated as the weighted sum of the normal vectors corresponding to the respective Euler angle triplets (ϕ, θ, ψ) (Eq. (2)).

$$\vec{n} = \frac{\sum_{i=1}^N c c_i^* (\vec{n}_i)}{\left| \sum_{i=1}^N c c_i^* (\vec{n}_i) \right|} \quad (2)$$

Where \vec{n}_i and \vec{n} are normal vectors obtained for different measurements and their

consensus, respectively. The CC against the respective classes was used as a weighting function.

An example of the consensus alignment for a hexamer in the lattice of a polyhedron is given in Supplementary Fig. 5b.

For the alignment, the coordinates determined using template matching in bin8 tomograms were first unbinned to bin4, and subjected to two rounds of (standard) subtomogram alignment. Four and two rounds of consensus alignment classified to groups 1-6 was performed with bin4 and bin2 data, respectively. One final round of consensus alignment with the group subclasses was then performed with bin2 data.

The context-based classification and consensus alignment also provided information on the CA lattice geometry in polyhedrons, specifically the tilt, twist, and distance geometry of unit pairs in the different classes (Supplementary Fig. 11). This information was utilized to further refine the unit positions in each class based on the mean geometry of the lattice in each class. To this end, the coordinates and Euler angles of each unit pair were optimized to minimize tilt, twist, and distance deviation of each unit pair from the mean value of the respective class.

This geometry optimization was performed before and after the last round of consensus alignment on 2× binned data, and resulted in improved averages and slightly increased resolution, compared to the consensus alignment alone.

As in the case of processing of the C2-hexamer from tubes, the data were split into half-sets at the bin2 stage. Specifically, in this case, the data were split into two half-sets of comparable size on the CLP level, in order to ensure complete separation of the subvolumes. Again, low-pass filters were set not beyond 32 Å prior to this step, to avoid aligning on higher resolution features within the unsplit dataset.

In order to generate the final class averages, two rounds of conventional subtomogram alignment were performed on unbinned data using the translational and orientational parameters from the previous binned consensus alignments. At this step, boxes were centered at the dimer interface and hence no more consensus refinement was performed.

Data visualization, model building, and analysis. Cryo-EM and cryo-ET data and structures were visualized in IMOD³⁹ and UCSF Chimera⁴⁶. Visualization of CLP lattice maps was done using the PlaceObject plugin in UCSF Chimera⁴².

The resolution of our electron microscopy maps for the C2-symmetric hexamer obtained from the regular CASPNC tubes and the CA pentamer from T1 particles allowed us to refine existing models of RSV CA (pdb 3TIR)²². At the obtained resolution for the C2-symmetric hexamer (4.3 Å), the helical pitch was visible and also several large side chains (e.g., tryptophans, phenylalanines, arginines, or lysines) could be identified. Small side chains and also negatively charged side chains were not clearly visible at this resolution. Therefore, rotamer refinements of residues were not possible.

The pseudo-atomic model of RSV CA (pdb 3TIR) was used as a starting model for refinement into the T1 particle map. The CA_{NTD} and CA_{CTD} of one CA monomer of pdb 3TIR were independently placed into the EM density, using the rigid-body fitting option in UCSF Chimera. Subsequently the linker connecting the two CA domains was joined in Coot. To account for the different monomer-monomer interactions in the RSV CA pentamer, the monomers were replicated according to the inherent fivefold symmetry of the map and an additional ring of CA_{CTD}'s was rigid-body fitted into the EM densities of the surrounding CA pentamers. A map segment (defined by a mask extending 3 Å around the rigid-body fitted model) was extracted, and real-space coordinate refinement against the EM density was performed using Phenix⁴⁷. This was iterated with manual model building in Coot⁴⁸, similar as described previously¹⁴. In brief, secondary structure restraints and non-crystallographic symmetry restraints were applied throughout all refinements. Each Phenix iteration consisted of five macro cycles, in which simulated annealing was performed in every macro cycle. Atomic displacement parameter refinement was performed at the end of each iteration.

For modeling the mature CA assembly in the C2-symmetric hexamer in regular CASPNC tubes, one CA monomer of pdb 3TIR was used. The monomer was rigid-body fitted into the EM density three times to accommodate the three symmetry-independent copies of CA in the C2-symmetric hexamer. The residues in the first strand of the N-terminal β-hairpin were removed, as no ordered density corresponding to these residues was present. As symmetry-independent monomers show differences in their respective orientations of their CA_{NTD} and CA_{CTD}, the fit was further manually refined in Coot. Subsequently, the symmetry-independent monomers were expanded according to the C2 symmetry of the cryo-EM density, completing the hexameric CA assembly. An additional ring of CA_{CTD}'s was rigid-body fitted into the EM densities of the surrounding CA hexamers, to shield the inner hexameric ring. Refinement was then performed similarly as described above, iterating between automatic refinement in Phenix and manual model building in Coot.

The model of the CA monomer derived from real-space refinement into the T1 particles was used to generate pseudo-atomic models of the hexameric and pentameric CA assembly in the CASPNC polyhedrons. CA_{NTD} and CA_{CTD} domains were separately fitted into the respective maps, using the rigid-body fit functionality in UCSF Chimera. For CASPNC polyhedron reconstructions, no further real-space refinement of the models was performed.

The quality of the pentamer CA model derived from T1 particles and the C2-symmetric hexamer CA model from regular tube CLPs was validated, using MOLPROBITY⁴⁹.

All comparisons and RMSD calculations were performed in UCSF Chimera or Matlab, between the C-alpha backbone atoms of selected residues.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The electron microscopy density maps and the corresponding models and a representative tomogram have been deposited in the Electron Microscopy Data Bank (accession codes: EMD-12485, EMD-12486, EMD-12487, EMD-12488, EMD-12489, EMD-12490, EMD-12491, EMD-12492, EMD-12493, EMD-12494, EMD-12495, EMD-12496, EMD-12497, EMD-12498, EMD-12499, EMD-12500, EMD-12501, EMD-12502, EMD-12503, EMD-12504, EMD-12505, EMD-12506, EMD-12507, EMD-12508, EMD-12509, EMD-12510, EMD-12511, and EMD-12774) and Protein Data Bank (accession codes: 7NO0, 7NO1, 7NO2, 7NO3, 7NO4, 7NO5, 7NO6, 7NO7, 7NO8, 7NO9, 7NOA, 7NOB, 7NOC, 7NOD, 7NOE, 7NOF, 7NOG, 7NOH, 7NOI, 7NOJ, 7NOK, 7NOL, 7NOM, 7NOM, 7NOO, 7NOP, and 7NOQ), respectively. The pdb we used as a starting model to refine RSV CA into our structure was PDB 3TIR. Source data are provided with this paper.

Code availability

MATLAB scripts for context-based classification and consensus alignments, as well as documentation on their usage are available for download at <https://schurlab.ist.ac.at/downloads/>.

Received: 13 November 2020; Accepted: 3 May 2021;

Published online: 28 May 2021

References

- Pornillos, O., Ganser-Pornillos, B. K., Kelly, B. N. & Yeager, M. X-ray structures of the hexameric building block of the HIV capsid. *Cell* **137**, 1282–1292 (2009).
- Pornillos, O., Ganser-Pornillos, B. K. & Yeager, M. Atomic-level modelling of the HIV capsid. *Nature* **469**, 424–427 (2011).
- Zhao, G., Perilla, J. R., Yufenyuy, E. L. & Zhang, P. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **497**, 643–646 (2013).
- Ni, T., Gerard, S., Zhao, G. & Zhang, P. Intrinsic curvature of the HIV-1 CA hexamer underlies capsid topology and interaction with cyclophilin A. *Nat. Struct. Mol. Biol.* **27**, 855–862 (2020).
- Lu, M., Russell, R. W., Bryer, A. J. & Polenova, T. Atomic-resolution structure of HIV-1 capsid tubes by magic-angle spinning NMR. *Nat. Struct. Mol. Biol.* **27**, 863–869 (2020).
- Obr, M. & Schur, F. K. M. Structural analysis of pleomorphic and asymmetric viruses using cryo-electron tomography and subtomogram averaging. *Adv. Virus Res.* **105**, 117–159 (2019).
- Schur, F. K. M., Obr, M., Hagen, W. J. H. & Briggs, J. A. G. An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science* **353**, 506–508 (2016).
- Schur, F. K. M., Hagen, W. J. H., Rumlová, M. & Briggs, J. A. G. Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution. *Nature* **517**, 505–508 (2015).
- Mattei, S., Glass, B., Hagen, W. J. H. & Briggs, J. A. G. The structure and flexibility of conical HIV-1 capsids determined within intact virions. *Science* **354**, 1434–1437 (2016).
- Mattei, S., Schur, F. K. M. & Briggs, J. A. G. Retrovirus maturation - an extraordinary structural transformation. *Curr. Opin. Virol.* **18**, 27–35 (2016).
- Campbell, E. M. & Hope, T. J. HIV-1 capsid: the multifaceted key player in HIV-1 infection. *Nat. Rev. Microbiol.* **13**, 471–483 (2015).
- Qu, K., Glass, B., Doležal, M. & Briggs, J. A. G. Structure and architecture of immature and mature murine leukemia virus capsids. *Proc. Natl Acad. Sci. USA* **115**, E11751–E11760 (2018).
- Ricana, C. L., Lyddon, T. D., Dick, R. A. & Johnson, M. C. Primate lentiviruses require Inositol hexakisphosphate (IP6) or inositol pentakisphosphate (IP5) for the production of viral particles. *PLoS Pathog.* **16**, e1008646 (2020).
- Dick, R. A., Xu, C., Morado, D. R. & Schur, F. K. M. Structures of immature EIAV Gag lattices reveal a conserved role for IP6 in lentivirus assembly. *PLoS Pathog.* **16**, e1008277 (2020).
- Dick, R. A., Zadrozny, K. K., Xu, C. & Vogt, V. M. Inositol phosphates are assembly co-factors for HIV-1. *Nature* **560**, 509–512 (2018).
- Bush, D. L. & Vogt, V. M. In vitro assembly of retroviruses. *Annu. Rev. Virol.* **1**, 561–580 (2014).
- Bharat, T. A. M., Davey, N. E., Ulbrich, P. & Briggs, J. A. G. Structure of the immature retroviral capsid at 8 Å resolution by cryo-electron microscopy. *Nature* **487**, 385–389 (2012).
- Joshi, S. M. & Vogt, V. M. Role of the Rous sarcoma virus p10 domain in shape determination of Gag virus-like particles assembled in vitro and within *Escherichia coli*. *J. Virol.* **74**, 10260–10268 (2000).
- Campbell, S. & Vogt, V. M. In vitro assembly of virus-like particles with Rous sarcoma virus Gag deletion mutants: identification of the p10 domain as a morphological determinant in the formation of spherical particles. *J. Virol.* **71**, 4425–4435 (1997).
- Campbell, S. & Vogt, V. M. Self-assembly in vitro of purified CA-NC proteins from Rous sarcoma virus and human immunodeficiency virus type 1. *J. Virol.* **69**, 6487–6497 (1995).
- Butan, C., Winkler, D. C., Heymann, J. B. & Steven, A. C. RSV capsid polymorphism correlates with polymerization efficiency and envelope glycoprotein content: implications that nucleation controls morphogenesis. *J. Mol. Biol.* **376**, 1168–1181 (2008).
- Bailey, G. D., Hyun, J. K., Mitra, A. K. & Kingston, R. L. A structural model for the generation of continuous curvature on the surface of a retroviral capsid. *J. Mol. Biol.* **417**, 212–223 (2012).
- Cardone, G., Purdy, J. G., Cheng, N. & Steven, A. C. Visualization of a missing link in retrovirus capsid assembly. *Nature* **457**, 694–698 (2009).
- Keller, P. W., Huang, R. K., England, M. R. & 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.* **87**, 13655–13664 (2013).
- Mallery, D. L., Márquez, C. L., McEwan, W. A. & James, L. C. IP6 is an HIV pocket factor that prevents capsid collapse and promotes DNA synthesis. *Elife* **7**, e35335 (2018).
- Ganser, B. K., Li, S., Klishko, V. Y. & Sundquist, W. I. Assembly and analysis of conical models for the HIV-1 core. *Science* **283**, 80–83 (1999).
- Acton, O., Grant, T., Nicastrò, G. & Rosenthal, P. B. Structural basis for fullerene geometry in a human endogenous retrovirus capsid. *Nat. Commun.* **10**, 1–13 (2019).
- Malakhov, M. P., Mattern, M. R., Malakhova, O. A. & Butt, T. R. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J. Struct. Funct. Genomics* **5**, 75–86 (2004).
- Schur, F. K. M., Dick, R. A., Hagen, W. J. H. & Briggs, J. A. G. The structure of immature virus-like Rous sarcoma virus gag particles reveals a structural role for the p10 domain in assembly. *J. Virol.* **89**, 10294–10302 (2015).
- Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
- Zivanov, J., Nakane, T., Forsberg, B. O. & Scheres, S. H. W. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, e42166 (2018).
- Morin, A., Eisenbraun, B., Key, J. & Sliz, P. Collaboration gets the most out of software. *Elife* **2**, e01456 (2013).
- Zheng, S. Q., Palovcak, E., Armache, J. P. & Agard, D. A. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
- Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
- Vulović, M., Franken, E., Ravelli, R. B. G. & Rieger, B. Precise and unbiased estimation of astigmatism and defocus in transmission electron microscopy. *Ultramicroscopy* **116**, 115–134 (2012).
- Hagen, W. J. H., Wan, W. & Briggs, J. A. G. Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J. Struct. Biol.* **197**, 191–198 (2017).
- Rohou, A. & Grigorieff, N. CTFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
- Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *Elife* **4**, e06980 (2015).
- Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).
- Turoňová, B., Schur, F. K. M., Wan, W. & Briggs, J. A. G. Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4 Å. *J. Struct. Biol.* **199**, 187–195 (2017).
- Castano-Diez, D., Kudryashev, M., Arheit, M. & Stahlberg, H. Dynamo: a flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. *J. Struct. Biol.* **178**, 139–151 (2012).
- Förster, F., Medalia, O., Zauberman, N. & Fass, D. Retrovirus envelope protein complex structure in situ studied by cryo-electron tomography. *Proc. Natl Acad. Sci. USA* **102**, 4729–4734 (2005).

43. Nickell, S., Förster, F., Linaroudis, A. & Plitzko, J. M. TOM software toolbox: acquisition and analysis for electron tomography. *J. Struct. Biol.* **149**, 227–234 (2005).
44. Wan, W., Kolesnikova, L., Clarke, M. & Briggs, J. A. G. Structure and assembly of the Ebola virus nucleocapsid. *Nature* **551**, 394–397 (2017).
45. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
46. Pettersen, E. F., Goddard, T. D., Huang, C. C. & Ferrin, T. E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
47. Adams, P. D., Afonine, P. V., Bunkoczi, G. & Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D* **66**, 213–221 (2010).
48. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D* **60**, 2126–2132 (2004).
49. Chen, V. B., Arendall, W. B. III, Headd, J. J. & Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D* **66**, 12–21 (2010).

Acknowledgements

This work was funded by the National Institute of Allergy and Infectious Diseases under awards R01AI147890 to R.A.D., R01AI150454 to V.M.V., R35GM136258 in support of J.-P.R.F., and the Austrian Science Fund (FWF) grant P31445 to F.K.M.S. Access to high-resolution cryo-ET data acquisition at EMBL Heidelberg was supported by iNEXT (grant no. 653706), funded by the Horizon 2020 program of the European Union (PID 4246). We thank Wim Hagen and Felix Weis at EMBL Heidelberg for support in cryo-ET data acquisition. This work made use of the Cornell Center for Materials Research Shared Facilities, which are supported through the NSF MRSEC program (DMR-179875). This research was also supported by the Scientific Service Units (SSUs) of IST Austria through resources provided by Scientific Computing (SciComp), the Life Science Facility (LSF), and the Electron Microscopy Facility (EMF).

Author contributions

Conceptualization, supervision, and funding: F.K.M.S. and R.A.D.; data collection M.O., C.L.R., N.N., J.-P.R.F., M.K., A.T., F.K.M.S., and R.A.D.; methods, validation, analysis:

M.O., J.-P.R.F., M.C.J., F.K.M.S., and R.A.D.; data curation and writing—original draft: M.O., F.K.M.S., and R.A.D.; writing—review and editing: M.O., C.L.R., J.-P.R.F., V.M.V., F.K.M.S., and R.A.D.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-23506-0>.

Correspondence and requests for materials should be addressed to F.K.M.S. or R.A.D.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contributions to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021