1 CryoET structures of immature HIV Gag reveal a complete six-helix bundle and
2 stabilizing small molecules distinct from IP6

3

4 Luiza Mendonça¹, Dapeng Sun^{1,2}, Jiying Ning², Jiwei Liu¹, Abhay Kotecha³, Mateusz
5 Olek^{4,5}, Thomas Frosio⁴, Xiaofeng Fu², Benjamin A Himes², Alex B. Kleinpeter⁶, Eric O.
6 Freed⁶, Jing Zhou⁷, Christopher Aiken⁷, Peijun Zhang^{1,2,4*}

7

8 ¹Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of
9 Oxford, Oxford, OX3 7BN, UK;

10 ²Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh,
11 PA 15260, USA;

12 ³Thermo Fisher Scientific, Achtseweg Noorg 5, 5651 GG, Eindhoven, The Netherlands;

13 ⁴Electron Bio-Imaging Centre, Diamond Light Source, Harwell Science and Innovation
14 Campus, Didcot OX11 0DE, UK;

15 ⁵Department of Chemistry, University of York, York, YO10 5DD, UK;

16 ⁶Virus-Cell Interaction Section, HIV Dynamics and Replication Program, Center for Cancer
17 Research, National Cancer Institute, Frederick, MD 21702-1201, USA;

18 ⁷Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical
19 Center, Nashville TN, 37232, USA

20

21 *Correspondence to Peijun Zhang, peijun@strubi.ox.ac.uk.

22 **Abstract**

23

24 Gag is the major HIV-1 structural polyprotein precursor. The Gag SP1 domain with the last
 25 residues of CA have been hypothesized to form a six-helix bundle necessary for particle
 26 assembly, but this bundle has not been fully resolved. Here, we determined the structures of
 27 complete CA-SP1 six-helix bundle connecting to the NC domain, from both *in vitro* Gag
 28 assemblies and viral-like particles (VLPs) carrying a T8I mutation in SP1, to near-atomic
 29 resolutions using cryoET and subtomogram averaging. The structures revealed novel
 30 densities, however distinct from IP6, inside the six-helix bundle of Gag assemblies,
 31 stabilizing the immature lattice. Interestingly, the T8I mutation impaired proteolytic cleavage
 32 of Gag at both SP1 boundaries. Our findings signify the involvement of small molecules in
 33 immature Gag assembly and provide the structural basis for development of small molecule
 34 inhibitors that stabilize SP1 helix, thus interfere with PR-mediated virus maturation.

35 **Introduction**

36

37 Gag is the structural precursor protein in HIV-1. It is a polyprotein synthesized from the
 38 unspliced viral mRNA, and is cleaved by the viral protease (PR) during the maturation stage
 39 of the replicative cycle into 6 proteins and peptides: matrix (MA), capsid (CA), spacer
 40 peptide 1 (SP1), nucleocapsid (NC), spacer peptide 2 (SP2) and p6¹. Gag plays a major role
 41 during viral particle assembly, with each of its protein domains performing distinct functions.
 42 MA is responsible for membrane targeting. NC is responsible for RNA encapsidation. The
 43 CA and SP1 region multimerizes forming the immature viral lattice, comprised of
 44 interconnected hexamers. The Gag hexamer is, therefore, the assembly unit of the immature
 45 lattice. Proteolytic cleavage of Gag results in the transition of the immature Gag lattice to a
 46 mature capsid formed by the CA protein²⁻⁴.

47

48 The SP1 segment of Gag is 14 amino acids long and is delimited by the very first (SP1|NC)
 49 and the very last (CA|SP1) cleavage sites processed during HIV-1 maturation. SP1 plays a
 50 major role in the higher-order multimerization of Gag during viral assembly and is essential
 51 for viral infectivity⁵⁻⁷. Molecular dynamics and circular dichroism studies indicated that at
 52 high local concentrations, the stretch of residues comprising the last 8 amino acids of CA and
 53 the SP1 form a six-helix bundle^{5,8}. It has been hypothesized that the amphipathic nature of
 54 this region would lead to association of six CA-SP1 molecules, thus burying the hydrophobic
 55 residues in the internal face of the bundle, thereby stabilizing the Gag hexamer⁸. An NMR
 56 study of this region indicated that it adopts an alpha-helical conformation in 30% TFE⁹, and
 57 at high protein concentrations in aqueous solutions⁸. Magic Angle Spinning (MAS) NMR of
 58 CA-SP1 tubular assemblies demonstrated that the SP1 peptide is highly dynamic¹⁰. The
 59 structures of immature CA lattice and the first 8 residues of SP1 were determined previously

60 by cryoEM and X-ray crystallography^{11,12}. Both structures show a six-helix bundle formed by
61 a stretch of 16 residues, 8 from the CA C-terminal domain (CA_{CTD}) and 8 from SP1.
62 However, the rest of SP1 and its connection to NC could not be resolved in these structures.
63
64 Other factors are considered to be important for immature lattice formation and stability, in
65 particular the small molecule inositol hexakisphosphate (IP6)¹³⁻¹⁶. The CA-SP1 six-helix
66 bundle contains two rings of positive charge formed by 12 lysines, K158 and K227, in the
67 CA_{CTD}. The cryoEM and x-ray crystal structures indicate that all the lysine side chains point
68 toward the center of the Gag hexamer, creating a strong positive potential^{11,12}. IP6 was found
69 to neutralize the positive charges at the center of the Gag hexamer, thus acting as a co-factor
70 stabilizing the Gag hexamer. Other polyanionic molecules may also fulfil this role¹⁷, but they
71 remain to be identified and characterized.
72
73 The SP1 region is thought to be the target of multiple maturation inhibitors (MIs), such as
74 Bevirimat (BVM) and PF-46396 (PF-96)¹⁸⁻²³. MIs block CA|SP1 cleavage, presumably by
75 stabilizing the CA-SP1 region, and/or obstructing protease binding, thus impairing viral
76 infectivity^{11,12,24}. Natural polymorphisms as well as escape mutations in the CA_{CTD} and SP1
77 can render MIs ineffective^{25,26}. Compensatory mutations selected during the passage of MI-
78 dependent viral mutants were shown to phenocopy the effect of MIs. One such compensatory
79 mutation is the single amino acid change, SP1 T8I, which is particularly interesting. It was
80 selected independently during the propagation of PF-96 and BVM-dependent viruses in the
81 absence of these MIs^{22,25}. The T8I mutation was found to promote the assembly of the Gag
82 lattice and stabilize the alpha-helical conformation of CA-SP1^{19,24}.
83

84 Here we report the structures of two types of immature Gag assemblies bearing the T8I
 85 mutation: one from purified Gag assemblies made in bacterial cells where the Gag
 86 polyprotein was expressed; the other from purified VLPs produced in mammalian cells. We
 87 solved the structures of full-length CA-SP1 from both assemblies at 4.2 Å and 4.5 Å
 88 resolutions using cryo-electron tomography (cryoET) and subtomogram averaging (STA)²⁷.
 89 Our maps show a fully extended six-helix bundle that connects the CA_{CTD} to the NC-RNA
 90 density layers and a possible organization of the NC-RNA. Intriguingly, novel densities,
 91 other than IP6, were identified to coordinate lysine residues in a manner similar to IP6,
 92 stabilizing CA-SP1 six-helix bundle. We further investigated the effect of T8I mutation on
 93 Gag processing, demonstrating that PR-mediated cleavage at both SP1 boundaries was
 94 impaired.

95

96 **Results**

97

98 **CryoET and subtomogram averaging of GagT8I assemblies and VLPs**

99 To resolve the structure of the full CA-SP1 six helix bundle, we utilized HIV-1 Gag proteins
 100 containing the T8I mutation at SP1 for stabilizing bundle assembly. SP1 T8I is a
 101 compensatory mutation that emerged in MI-dependent viruses subject to the selective
 102 pressure of replicating under sub-optimal concentrations of MIs. This mutation was shown to
 103 stabilize immature assembly^{22, 24, 25}. We used two different Gag assemblies for structure
 104 analysis by cryoET: Gag_{ΔMA}T8I protein assemblies expressed in *E. coli* (Fig. 1a and b) and
 105 GagT8I VLPs produced in HEK293T cells (Fig. 1a and d). The Gag_{ΔMA}T8I construct lacks
 106 most of the globular domain of MA (residues 15 to 100) and p6 (Fig. 1a). Wild-type (WT)
 107 Gag_{ΔMA} is a widely used construct for HIV-1 Gag lattice investigation^{11,13,28-30}. Upon
 108 induction in *E. coli*, the WT Gag_{ΔMA} protein was expressed, purified from bacterial lysates,
 109 and assembled *in vitro* into spherical particles bearing Gag immature lattices in the presence
 110 of oligonucleotides. The Gag_{ΔMA}T8I protein, however, was not soluble when released from
 111 bacteria owing to its self-assembly into spherical particles inside the bacterial cells
 112 (Supplementary Fig. 1). We therefore purified Gag_{ΔMA}T8I assemblies directly from lysed
 113 bacterial cells. The GagT8I VLPs were produced by transfection of a codon-optimized Gag
 114 expression vector into HEK293T cells and purified from the cell supernatant, similar to the
 115 WT Gag VLPs³¹.

116

117 A total of 90, 13 and 46 tilt series were collected for Gag_{ΔMA}T8I assemblies in the absence
 118 and presence of IP6 and GagT8I VLPs, respectively. Details of data acquisition and image
 119 processing are summarized in Table 1, and data processing workflow is illustrated in
 120 Supplementary Figure 2. We created a cryoET Toolbox for the on-the-fly processing of

cryoET data as they were acquired (Supplementary Fig. 2). emClarity, a software package developed in our group³², was used for subtomogram alignment and averaging. A low-pass filtered (25 Å) Gag hexamer density map (EMD-8403)³³ was used for template-matching in emClarity, to automatically extract the position and initial orientation of the Gag hexamer in tomograms, as shown in Figure 1c and e. Upon iterative refinement (Supplementary Fig. 2), the final density maps of the Gag hexamer from Gag_{ΔMA}T8I assemblies in the absence and presence of IP6 and GagT8I VLPs were obtained at 4.2 Å, 4.2 Å and 4.5 Å resolution, respectively (Supplementary Fig. 3c). The resulting maps confirm an immature Gag lattice from Gag_{ΔMA}T8I assemblies and GagT8I VLPs, which is composed of interconnected hexamers formed by the CA N-terminal domain (CA_{NTD}), CA_{CTD} and an extended six-helix bundle formed by CA_{CTD} and SP1 peptide (Fig. 1f).

GagT8I assembles into a full six-helix bundle further stabilized by small molecules

The subtomogram-averaged maps at 4.2 Å and 4.5 Å resolutions enabled refinement of atomic models of CA and SP1 (Fig. 2a-b). The refinement statistics are summarized in Table 1. Overall, the structures of Gag_{ΔMA}T8I and GagT8I VLPs are very similar, with an RMSD of 0.577 Å (Fig. 2a-b). The complete CA-SP1 helix and the NC-RNA density layer which SP1 connects are clearly resolved in both maps (Fig. 2a-b), which were incomplete or missing from the previous CA and SP1 structures^{11,12}. Nonetheless, interesting differences could be observed between the two. The GagT8I VLP subtomogram average shows a density that extends beyond the N-terminus of CA, where 5 residues in the junction region between MA and CA are resolved and modelled (Fig. 2c, bottom). This is because GagT8I VLPs have a complete MA domain that interacts with the lipid envelope; by contrast, in the Gag_{ΔMA}T8I assemblies the globular domain of MA is missing and Gag is not constrained by a lipid envelope (Supplementary Fig. 4). The refined atomic model shows that the T8I mutation

stabilizes the Gag hexamer by enhancing the hydrophobicity in one face of the amphipathic CA-SP1 helix (Supplementary Fig. 5). This confirms the hypothesis that the Gag hexamer is formed by the association and burial of the hydrophobic CA-SP1 face in the internal channel of the six-helix bundle^{7,8}.

Intriguingly, we observed a strong density inside the space formed by the six-helix bundle. This density is coordinated by two lysine rings formed by residues K158 and K227 near the bottom of the CA_{CTD} and above the bundle (Fig. 2d). A close inspection revealed that the profile and position of the internal density are different in the Gag_{ΔMA}T8I and GagT8I VLPs maps (Fig. 2d, red arrows). For GagT8I VLPs produced in mammalian cells, this density is consistent with IP6, as suggested in previous studies^{13,14} (Fig. 2d, middle). IP6 is known to associate with the HIV-1 immature lattice and is present at cytoplasmic concentrations estimated at 20-50 nM in mammalian cells³⁴. *E. coli* cells, however, lack the enzyme responsible for biosynthesis of the inositol ring and are thus thought to lack inositides³⁵. Considering that Gag_{ΔMA}T8I assemblies were produced in *E. coli* cells, it seems unlikely that the density in Gag_{ΔMA}T8I map is IP6 (Fig. 2d, left). In fact, when the same Gag_{ΔMA}T8I assembly is incubated with a buffer containing 10 μM IP6, this internal density reverts to a lower position closely resembling that seen in GagT8I VLPs (Fig. 2d, right). It is plausible that a small anionic molecule other than IP6 assists in the charge neutralization of the lysine rings and promoted assembly in *E. coli*-expressed Gag_{ΔMA}T8I. Consistent with this, a similar density was also present at the same location in a previously published subtomogram average of bacterially expressed WT Gag_{ΔMA} assemblies (Fig. 3c, blue)¹¹. Further studies are needed to identify such small molecules. Additional small molecules may also be involved in stabilization of the six-helix bundle and HIV-1 immature lattice, as another set of densities were observed inside the six-helix bundle below the lysine rings. These densities differ

between the Gag_{ΔMA}T8I and GagT8I VLP maps (Fig. 2d). In Gag_{ΔMA}T8I, this density sits in the highly hydrophobic environment made by the L231 residue (L363 in Gag), whereas in GagT8I VLPs, this density is closer to the M4 residues (M367 in Gag). An exciting prospect from this study is that small molecules can be located and differentiated in subtomogram averages of *in situ* structures, though a better resolution would help in their identification.

The T8I mutation reduces SP1 dynamics and impairs proteolytic cleavage

In the previous WT Gag_{ΔMA} map (EMD-3782) and its corresponding atomic model (PDB 5L93), the SP1 region after T8 was not resolved¹¹. The SP1 helix in the Gag_{ΔMA}T8I map is complete and extends to the SP1 and NC interface (Fig. 3a-c), therefore allowing model building of the C-terminal portion of SP1, from T8I to M14 (Fig. 3d). The T8I mutation was shown to quench the dynamics of SP1 by MAS NMR²⁴, and potentially stabilize a continuous helical conformation. It is worth noting that the WT Gag_{ΔMA} map (EMD-3782) was determined in the presence of BVM. The density corresponding to BVM is visible in the WT Gag_{ΔMA} map (Fig. 3c, pink arrow), but not in Gag_{ΔMA} T8I map (Fig. 2d, left panel). Yet, the observed unknown densities coordinated by K158 and K227 overlapped well in both WT Gag_{ΔMA} and Gag_{ΔMA}T8I maps (Fig. 3c, red arrow), even though the two proteins were expressed and assembled independently in different manners in the absence of IP6.

The SP1 peptide is flanked by the first (SP1|NC) and the last (CA|SP1) cleavage sites in the HIV-1 maturation cascade. If the T8I mutation stabilizes the extended six-helix bundle, this may affect PR processing of these cleavage sites. We thus tested the effect of the T8I mutation on PR-mediated cleavage of Gag assemblies using *in vitro* cleavage assays (Fig. 3e-f)³⁶. We first analysed the concentration-dependent processing of WT Gag_{ΔMA} and Gag_{ΔMA}T8I by PR. WT Gag_{ΔMA} processing showed the established ordering of cleavage as

the PR concentration increased, with the first major cleavage products MA-CA-SP1 and NC-SP2 (cleaving at SP1|NC), followed by CA-SP1 (cleaving at MA|CA), and completed with CA and NC (Fig. 3e). In contrast, not only was a much higher PR concentration required to initiate processing of Gag_{ΔMA}T8I, but also the order of cleavage was altered (Fig. 3e). The first cleavage occurred at MA|CA site, suggesting an impaired or delayed cleavage at the SP1|NC site. Furthermore, the CA|SP1 site was also protected from cleavage, such that very little CA was produced. The same trend was observed when we investigated the kinetics of PR processing *in vitro* (Fig. 3f). At the end point of PR processing (120 minutes), CA-SP1 was the major final product for Gag_{ΔMA}T8I, instead of CA. Therefore, cleavage at both proteolytic sites flanking SP1 are impaired by the T8I mutation, and/or the small molecules, which supports the formation of a stable and extended SP1 helix.

The NC-RNA layer is defined but variable in the SP1-NC domain organization

We next sought to investigate how the SP1 six-helix bundle connects to the NC-RNA region, which has been intractable previously. As shown in Fig. 4a, three layers of densities were observed in tomographic slices of the Gag_{ΔMA}T8I particles. Radial density profiles from three different particles of similar size displayed three distinct peaks corresponding to the CA_{NTD}, CA_{CTD} and NC-RNA density layers (Fig. 4a)³⁷. While CA_{NTD}, CA_{CTD} and SP1 are rigidly connected as seen in the Gag_{ΔMA}T8I map, NC-RNA is flexibly linked to SP1, resulting in a smeared density in Gag_{ΔMA}T8I subtomogram averages (Fig. 3a).

To gain structural information on the SP1-NC-RNA region, we carried out localized 3D alignment and classification, focusing on the SP1-NC-RNA region, using PCA and k-means subtomogram classification tools implemented in emClarity without imposing any symmetry (Supplementary Fig. 6a). Among 32 resulting 3D classes, we identified multiple classes that

seemed to share common organization of the NC-RNA layer (Supplementary Fig. 6b, red boxes) and averaged 6747 particles from these classes to generate a density map for the SP1-NC-RNA region (Fig. 4b-c). A composite molecule of CA_{CTD}-SP1-NC was created based on the refined atomic model CA-SP1 and the published NC-RNA model (PDB 1F6U)³⁸, which was then used separately for rigid body fitting of the NC and the RNA domains in the density map. We performed this for each CA monomer in the model, resulting in two plausible models of SP1-NC-RNA organization (Fig. 4b-c). To bridge the space between the six-helix bundle density and the NC-RNA density layer, it was necessary to model the N-terminus of NC as a coil, up to residue Asn8. In both models, the NC-RNA complex is at the density layer below and surrounding the six-helix bundle (Fig. 4c). As HIV-1 NC is very flexible and the viral RNA can adopt many conformations, it is plausible that there exist multiple modes of SP1 to NC connection, including the two modes described here. Taken together, a stretched coil conformation at the SP1 and NC junction and/or dynamic unfolding of this region could facilitate PR processing. As such, the helix stabilization imparted by the T8I mutation may explain the observed delays in SP1|NC cleavage.

Discussion

HIV-1 Gag is a multidomain and multifunctional protein responsible for membrane targeting, RNA encapsidation, particle assembly and budding. The CA and SP1 domains of Gag form the viral immature hexagonal lattice. Residues that make up the CA-SP1 helix are remarkably conserved and mutations in this region lead to loss of infectivity and/or assembly defects^{8,19,39,40}.

The SP1-T8I mutation is a stabilizing compensatory mutation selected for during the passage of MI-dependent virus. Our refined atomic model shows that this mutation increases the hydrophobicity at the inner face of the six-helix bundle, thus stabilizing it (Supplementary Fig. 5). It has been shown that IP6 is also involved in Gag hexamer stabilization by neutralizing positive charges of lysine residues at the base of the CA_{CTD}¹³⁻¹⁶. Here, we show that small molecules other than IP6 can neutralize charged lysine rings to stabilize the immature lattice, demonstrating the potential of subtomogram averaging to localize and differentiate small molecules in *in situ* structures, and possibility of developing small molecules as MIs..

HIV-1 maturation is a highly regulated process. The sequence of cleavage events is thought to be determined by the affinity constants and processing kinetics between the different cleavage sites in Gag and HIV-1 PR⁴¹⁻⁴³. Events that alter the kinetics or sequence of maturation invariably lead to loss of infectivity^{19,24,44,45}. T8I mutation was acquired as a second-site mutation that restores assembly and maturation to the BVM- and PF-96-dependent mutants, but the T8I single mutant exhibited a maturation defect with inefficient cleavage at CA-SP1. The inference is that the original resistant mutants exhibited unstable Gag lattices in the absence of the compound, which stabilizes the lattice, and T8I rescues the dependence on the compound by stabilizing the lattice, which becomes hyper-stable in the absence of a destabilizing mutation. We show that the T8I mutation alters the cleavage sequence in HIV-1 maturation and impairs processing of both SP1 boundaries, suggesting that a finely tuned SP1 stability is required to balance between lattice assembly, proper maturation and infectivity.

SP1 is critically important as one of the major targets for therapeutic drugs against HIV-1 and AIDS. The structure of full-length CA-SP1 in immature Gag assemblies derived from this study provides a blueprint for future development of small molecule inhibitors that can lock SP1 in a stable helical conformation and thus interfere with PR-mediated processing and virus maturation, and block HIV-1 infection.

Materials and Methods

Particle production and purification

The Gag Δ MA T8I mutant was constructed from pET2 PRR400 by site-direct mutagenesis. Protein was expressed in *E. coli*, Rosetta 2 (DE3), cultured in Luria-Bertani medium, and induced with 0.5 mM isopropylthiogalactoside at 23 °C for 16 hrs. Gag Δ MA T8I assemblies were formed inside of bacterial cells and purified directly from cell lysate. The cell pellets were collected and resuspended in lysis buffer (25 mM Tris, pH 7.5, 0.5 M NaCl) and broken with a microfluidizer. Subsequently, the lysate was centrifuged at 5000g for 10min to remove cell debris. The supernatant was collected and subjected to sucrose gradient centrifugation (30-70%) with a 15% sucrose cushion on the top. Gradient was spun at 210,000g for 18 hrs. Particles were collected from the pellet and resuspended in lysis buffer with and without 10 μ M IP6 and used for CryoEM.

The GagT8I mutant VLPs were produced from HEK293T cells transfected with a codon-optimized full-length Gag expression plasmid (pCMV-Gag-opt)⁴⁶. The culture supernatants were harvested approximately 40 hours after transfection, filtered through a 0.45-micron pore-sized filter, and pelleted through a cushion of 20% sucrose (wt/vol) in STE buffer (20

mM Tris-HCl pH 7.4, 100 mM NaCl, 1 mM EDTA). The VLPs were gently resuspended in STE buffer and frozen at -80°C.

CryoEM sample preparation

A homemade manual plunger was used for cryoEM grid preparation. Quantifoil R2/2 copper grids were freshly glow discharged at 30 mA for 30 seconds. For Gag_{ΔMA}T8I assemblies, 2 μl sample was applied to the carbon-side of the grid, followed by 1 μl 10 nm fiducial gold bead solution, and blotted from the backside since samples are very fragile and sensitive to blotting paper. 1 μl sample buffer was then added on the grid backside to facilitate the filter paper blotting from the backside. For GagT8I VLPs, 1 μl 10 nm fiducial gold beads were applied to the grid first, followed by 1~2 μl sample, and blotted from the frontside. The grid after blotting was then quickly immersed into liquid ethane cooled by liquid nitrogen. Frozen grids were stored in liquid nitrogen until data collection.

CryoET data acquisition

Data for Gag_{ΔMA}T8I assemblies and GagT8I VLPs were collected at the Electron Bio-Imaging Centre at Diamond Light Source (eBIC-DLS) in the United Kingdom. Tilt-series were acquired on a Gatan K2-Summit detector in super-resolution mode using a Thermo Fisher Scientific KriosG3i microscope operating at 300 kV equipped with a Gatan energy filter (slit width 20 eV; GIF Quantum LS, Gatan). Tilt-series were recorded at a nominal magnification of 105,000×, corresponding to a calibrated pixel size of 1.35 Å. A dose-symmetric scheme⁴⁷ was used to collect tilt-series from -60° to 60° or -54° to 54° at a step size of 3° using SerialEM software⁴⁸. At each tilt, a movie stack consisting of 5-10 frames was recorded with a set dose rate of 3-6 e⁻/Å²/sec. Tilt-series were collected at a range of nominal defocus between -1.0 and -4.0 μm and a target total dose of 110-120 e⁻/Å² was

applied over the entire series. For Gag_{ΔMA}T8I with 10 uM IP6 sample, data were collected at the Thermo Fisher Scientific RnD division in Eindhoven, The Netherlands. Tilt-series were acquired on a Falcon4 detector in EER (electron event representation)^{49,50} format using a Thermo Fisher Scientific KriosG4 operating at 300 kV equipped with a CFEG and SelectrisX energy filter (slit width 10 eV; Thermo Fisher Scientific). Tilt-series were recorded at a nominal magnification of 105,000×, corresponding to a calibrated pixel size of 1.196 Å. A dose-symmetric scheme was used to collect tilt-series from -60° to 60° at a step size of 3° using TEM Tomography software. The CFEG was automatically flashed every 8hrs. At each tilt, a movie stack consisting of 217 EER frames was recorded with a dose rate of 4.6 e⁻/px/sec and a total dose of 2.98 e⁻/Å² per tilt. Tilt-series were collected at a nominal defocus range between -1.0 and -4.0 μm and a target total dose of 122 e⁻/Å² was applied over the entire tilt-series. Further details are given in Table 1.

CryoET image processing

CryoET Toolbox (https://github.com/ffyr2w/cet_toolbox) was used for on-the-fly pre-processing of K2 datasets (freely available). In summary, movie frames were Fourier-cropped to a final pixel size of 1.35 Å and motion corrected by averaging frames for each tilt using program “alignframes” implemented in the IMOD package⁵¹ or MotionCor2⁵². For the Falcon4 data, 30 EER frames were grouped to create 7 dose fractions from 210 frames (last 7 frames were discarded) and motion corrected using Relion3.1⁵³. Tilt series were aligned using the default parameters in IMOD version 4.10.22 with the eTomo interface⁵¹, using gold-fiducial markers. The alignment parameters including the projection transformations, local alignments, and fitted tilt angles were then passed to emClarity³² framework. Subtomogram alignment and averaging were carried out in emClarity following the published protocol³². Briefly, the workflow was as follows: Particles were picked from 6× binned non-

CTF-corrected tomograms by emClarity template matching function using EMD-8403³³ low-pass filtered to 25 Å as template. The template matching results were cleaned automatically on basis of geometrical restraints using ‘removeNeighbours’ function implemented in emClarity. Only particles that had at least 3 neighbours within 100 Å and oriented within 20° were retained. Subtomograms at the air-water interface were manually discarded using IMOD⁵¹. Following template matching, the data set was randomly split into two groups based on tomogram of origin, rather than randomly by sub-tomograms, which were processed independently for all subsequent steps. The initial positions and orientations of the first cycle averaging come from template matching results. C6 symmetry was applied throughout all sub-tomogram averaging procedures. The 3D alignment procedures were carried out gradually with binning of 5, 4, 3, 2 and 1. At each binning, duplicate particles were removed and the tilt-series geometry was refined using the positions of subtomograms as fiducial markers (TomoCPR). The Fourier Shell Correlation was calculated by the gold-standard method from even and odd data sets. Density maps were visualized in Chimera⁵⁴ or PyMol (Schrödinger, Inc.). A diagram for the data processing is presented in Supplementary Fig. 2.

Model building and refinement

An atomic model was built and refined against the Gag_{ΔMA}T8I density map. The crystal structure (PDB 4XFX)⁵⁵ and cryoEM structure (PDB 5L93)¹¹ of HIV-1 CA were used as an initial reference. At 4.2 Å resolution, positive and bulky side chains are clearly visible enabling reasonable positioning of the residues in the atomic models. Rotamers were generally not refined, unless there was good evidence for a different rotamer in our density map. For the SP1 region, a composite model was made by joining the SP1 residues from PDB 1U57⁹ to the 5L93 model (lacking the last 6 SP1 residues) to serve as initial reference for the model refinement.

371

372 The model refinement procedure involved iterative manual refinement in Coot⁵⁶ followed by
373 3 rounds of Phenix cryoEM real-space refinement⁵⁷. After each round of real-space
374 refinement, Ramachandran and rotamer outliers were manually refined in Coot, and another
375 round of real-space refinement was performed until good model statistics were achieved. In
376 order to ensure reasonable modelling in the SP1 region, the density map of Gag_{ΔMA}T8I was
377 sharpened with a B factor of 120 in combination with a 1% FDR confidence map⁵⁸ and a
378 locally sharpened B-factor map⁵⁹ for the manual refinement in Coot.

379

380 **Subtomogram classification and modelling of the NC-RNA layer**

381 For focused alignment and classification at the SP1-NC junction, a $49 \times 49 \times 47$ Å mask was
382 applied to the region that comprised the CA-CTD, the six-helix bundle and the NC-RNA
383 layer. PCA and k-means classification implemented in emClarity were used to identify
384 common organization features in NC-RNA layer in the subtomograms. No symmetry was
385 imposed and tomograms were binned by a factor of 3 (equivalent pixel size 4.05 Å).
386 Subtomograms were classified into 32 classes and among those 3 major classes were
387 combined, comprising 6747 subvolumes.

388

389 A composite model was built in Coot⁵⁶ using the refined Gag_{ΔMA}T8I model and the published
390 NC-ΨRNA_{SL2} model (PDB 1F6U)³⁸. UCSF Chimera⁵⁴ was used for rigid body fitting of the
391 NC and RNA region of the composite model into the NC-RNA density map. This was
392 repeated for each CA monomer. Bond lengths were regularized in Coot by using the
393 Regularize Zone from residue M14 in SP1 to residue Q9 in NC.

394

395 ***In vitro* cleavage of Gag assemblies**

PR digestion experiments with recombinant WT and T8I assemblies were performed by the addition of various concentrations of PR to the Gag WT assembly mixture and T8I cell lysate which was pre-treated and diluted to 2mg/ml, and incubated at 37°C for 2hrs. For kinetic analysis of Gag WT and T8I cleavage, 3.3μM of HIV-1 PR was added to the Gag assembly mixture and incubated at 37°C, at different time points, 4ul of the digestion reaction mixture was taken out and mixed with NuPAGE® LDS Sample Buffer (Invitrogen) to stop the reaction, and then subjected to NuPAGE Novex 4–12% Bis-Tris gel (Invitrogen) for cleavage products analysis and visualized by Coomassie blue staining.

Data availability

The cryoEM density maps CA-SP1 hexamer from Gag_{ΔMA}T8I assemblies in the absence and presence of IP6 and GagT8I VLPs were deposited in the EMDB under accession code EMD-11894, EMD-11899, and EMD-11897. The refined models were deposited in PDB under accession codes 7ASH and 7ASI, respectively.

Acknowledgements

We thank Dr. Teresa Brosenitsch for critical reading of the manuscript. We are grateful to Drs Alistair Siebert, Kyle Dent, Andy Howe and Lingbo Yu for technical support. We thank Reint Boer Iwema for support with TEM Tomography software. We acknowledge Diamond for access and support of the CryoEM facilities at the UK national electron bio-imaging centre (eBIC, proposal NR18477, NR21005 and NT21004), funded by the Wellcome Trust, MRC and BBSRC. We also thank Thermo Fisher Scientific for microscope access. The computational aspects of this research were supported by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z and the NIHR Oxford BRC. This work was supported by the National Institutes of Health P50 grant AI150481 (P.Z.), the UK Wellcome Trust Investigator

Award 206422/Z/17/Z (P.Z.), and the UK Biotechnology and Biological Sciences Research Council grant BB/S003339/1 (P.Z.).

Author contributions

P.Z. conceived the research and designed the experiments. J.N. purified Gag Δ MA T8I assemblies and performed *in vitro* protease cleavage experiments. J.Z. and C.A. purified GagT8I VLPs. X.F. prepared cryoEM samples. J.L. and A.K. collected cryoET data. T.F. developed CryoET Toolbox used for pre-processing. L.M., D.S. and B.H. performed subtomogram averaging using emClarity. L.M., and M.O. refined the structural models. A.B.K. performed functional assays. L.M., E.O.F., D.S. and P.Z. analyzed data and wrote the paper with support from all the authors.

Competing interests

A.K. is an employee of Thermo Fisher Scientific.

References

1. Aiken, C. & Zhang, P. HIV-1 Maturation. in *Advances in HIV-1 Assembly and Release* (ed. Freed, E.O.) 153-166 (Springer New York, New York, NY, 2013).
2. Byeon, I.J. et al. Structural convergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. *Cell* **139**, 780-790 (2009).
3. Ni, T. et al. Intrinsic curvature of the HIV-1 CA hexamer underlies capsid topology and interaction with cyclophilin A. *Nat Struct Mol Biol* **27**, 855-862 (2020).
4. Zhao, G. et al. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **497**, 643-646 (2013).
5. Accola, M.A., Hoglund, S. & Gottlinger, H.G. A putative alpha-helical structure which overlaps the capsid-p2 boundary in the human immunodeficiency virus type 1 Gag precursor is crucial for viral particle assembly. *J Virol* **72**, 2072-2078 (1998).
6. Krausslich, H.G., Facke, M., Heuser, A.M., Konvalinka, J. & Zentgraf, H. The spacer peptide between human immunodeficiency virus capsid and nucleocapsid proteins is essential for ordered assembly and viral infectivity. *J Virol* **69**, 3407-3419 (1995).
7. Liang, C. et al. Characterization of a putative alpha-helix across the capsid-SP1 boundary that is critical for the multimerization of human immunodeficiency virus type 1 gag. *J Virol* **76**, 11729-11737 (2002).
8. Datta, S.A. et al. On the role of the SP1 domain in HIV-1 particle assembly: a molecular switch? *J Virol* **85**, 4111-4121 (2011).
9. Morellet, N., Druillenec, S., Lenoir, C., Bouaziz, S. & Roques, B.P. Helical structure determined by NMR of the HIV-1 (345-392)Gag sequence, surrounding p2: implications for particle assembly and RNA packaging. *Protein Sci* **14**, 375-386 (2005).
10. Han, Y. et al. Magic angle spinning NMR reveals sequence-dependent structural plasticity, dynamics, and the spacer peptide 1 conformation in HIV-1 capsid protein assemblies. *J Am Chem Soc* **135**, 17793-17803 (2013).
11. Schur, F.K. et al. An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science* **353**, 506-508 (2016).
12. Wagner, J.M. et al. Crystal structure of an HIV assembly and maturation switch. *Elife* **5**(2016).
13. Dick, R.A. et al. Structures of immature EIAV Gag lattices reveal a conserved role for IP6 in lentivirus assembly. *PLoS Pathog* **16**, e1008277 (2020).

- 479 14. Dick, R.A. et al. Inositol phosphates are assembly co-factors for
480 HIV-1. *Nature* **560**, 509–512 (2018).
- 481 15. Mallery, D.L. et al. Cellular IP6 Levels Limit HIV Production while
482 Viruses that Cannot Efficiently Package IP6 Are Attenuated for
483 Infection and Replication. *Cell Rep* **29**, 3983–3996 e3984 (2019).
- 484 16. Mallery, D.L. et al. IP6 is an HIV pocket factor that prevents capsid
485 collapse and promotes DNA synthesis. *Elife* **7**(2018).
- 486 17. Campbell, S. et al. Modulation of HIV-like particle assembly in vitro
487 by inositol phosphates. *Proc Natl Acad Sci U S A* **98**, 10875–10879
488 (2001).
- 489 18. Blair, W.S. et al. New small-molecule inhibitor class targeting human
490 immunodeficiency virus type 1 virion maturation. *Antimicrob Agents*
491 *Chemother* **53**, 5080–5087 (2009).
- 492 19. Fontana, J. et al. Identification of an HIV-1 Mutation in Spacer
493 Peptide 1 That Stabilizes the Immature CA-SP1 Lattice. *J Virol* **90**,
494 972–978 (2016).
- 495 20. Li, F. et al. PA-457: a potent HIV inhibitor that disrupts core
496 condensation by targeting a late step in Gag processing. *Proc Natl*
497 *Acad Sci U S A* **100**, 13555–13560 (2003).
- 498 21. Nguyen, A.T. et al. The prototype HIV-1 maturation inhibitor,
499 bevirimat, binds to the CA-SP1 cleavage site in immature Gag
500 particles. *Retrovirology* **8**, 101 (2011).
- 501 22. Waki, K. et al. Structural and functional insights into the HIV-1
502 maturation inhibitor binding pocket. *PLoS Pathog* **8**, e1002997 (2012).
- 503 23. Zhou, J. et al. Small-molecule inhibition of human immunodeficiency
504 virus type 1 replication by specific targeting of the final step of
505 virion maturation. *J Virol* **78**, 922–929 (2004).
- 506 24. Wang, M. et al. Quenching protein dynamics interferes with HIV capsid
507 maturation. *Nat Commun* **8**, 1779 (2017).
- 508 25. Adamson, C.S. et al. In vitro resistance to the human
509 immunodeficiency virus type 1 maturation inhibitor PA-457
510 (Bevirimat). *J Virol* **80**, 10957–10971 (2006).
- 511 26. Adamson, C.S., Sakalian, M., Salzwedel, K. & Freed, E.O.
512 Polymorphisms in Gag spacer peptide 1 confer varying levels of
513 resistance to the HIV- 1 maturation inhibitor bevirimat.
514 *Retrovirology* **7**, 36 (2010).
- 515 27. Zhang, P. Advances in cryo-electron tomography and subtomogram
516 averaging and classification. *Curr Opin Struct Biol* **58**, 249–258
517 (2019).
- 518 28. Bartonova, V. et al. Residues in the HIV-1 capsid assembly inhibitor
519 binding site are essential for maintaining the assembly-competent
520 quaternary structure of the capsid protein. *J Biol Chem* **283**, 32024–
521 32033 (2008).

29. Gross, I., Hohenberg, H., Huckhagel, C. & Krausslich, H.G. N-Terminal extension of human immunodeficiency virus capsid protein converts the in vitro assembly phenotype from tubular to spherical particles. *J Virol* **72**, 4798–4810 (1998).
30. Gross, I. et al. A conformational switch controlling HIV-1 morphogenesis. *EMBO J* **19**, 103–113 (2000).
31. Martin, J.L., Cao, S., Maldonado, J.O., Zhang, W. & Mansky, L.M. Distinct Particle Morphologies Revealed through Comparative Parallel Analyses of Retrovirus-Like Particles. *J Virol* **90**, 8074–8084 (2016).
32. Himes, B.A. & Zhang, P. emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nat Methods* **15**, 955–961 (2018).
33. Ning, J. et al. In vitro protease cleavage and computer simulations reveal the HIV-1 capsid maturation pathway. *Nat Commun* **7**, 13689 (2016).
34. Barker, C.J., Wright, J., Hughes, P.J., Kirk, C.J. & Michell, R.H. Complex changes in cellular inositol phosphate complement accompany transit through the cell cycle. *Biochem J* **380**, 465–473 (2004).
35. Das-Chatterjee, A. et al. Introgression of a novel salt-tolerant L-myo-inositol 1-phosphate synthase from *Porteresia coarctata* (Roxb.) Tateoka (PcIN01) confers salt tolerance to evolutionary diverse organisms. *FEBS Lett* **580**, 3980–3988 (2006).
36. Meng, X. et al. Protease cleavage leads to formation of mature trimer interface in HIV-1 capsid. *PLoS Pathog* **8**, e1002886 (2012).
37. Zhou, L. et al. Low-dose phase retrieval of biological specimens using cryo-electron ptychography. *Nat Commun* **11**, 2773 (2020).
38. Amarasinghe, G.K. et al. NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. *J Mol Biol* **301**, 491–511 (2000).
39. Melamed, D. et al. The conserved carboxy terminus of the capsid domain of human immunodeficiency virus type 1 gag protein is important for virion assembly and release. *J Virol* **78**, 9675–9688 (2004).
40. Pettit, S.C. et al. The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *J Virol* **68**, 8017–8027 (1994).
41. Louis, J.M., Weber, I.T., Tozser, J., Clore, G.M. & Gronenborn, A.M. HIV-1 protease: maturation, enzyme specificity, and drug resistance. *Adv Pharmacol* **49**, 111–146 (2000).
42. Pettit, S.C., Clemente, J.C., Jeung, J.A., Dunn, B.M. & Kaplan, A.H. Ordered processing of the human immunodeficiency virus type 1 GagPol precursor is influenced by the context of the embedded viral protease. *J Virol* **79**, 10601–10607 (2005).

43. Pettit, S.C., Lindquist, J.N., Kaplan, A.H. & Swanstrom, R. Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-Pol precursor are cleaved by the viral protease at different rates. *Retrovirology* **2**, 66 (2005).
44. de Marco, A. et al. Role of the SP2 domain and its proteolytic cleavage in HIV-1 structural maturation and infectivity. *J Virol* **86**, 13708–13716 (2012).
45. Ludwig, C., Leiherer, A. & Wagner, R. Importance of protease cleavage sites within and flanking human immunodeficiency virus type 1 transframe protein p6* for spatiotemporal regulation of protease activation. *J Virol* **82**, 4573–4584 (2008).
46. Huang, Y., Kong, W.P. & Nabel, G.J. Human immunodeficiency virus type 1-specific immunity after genetic immunization is enhanced by modification of Gag and Pol expression. *J Virol* **75**, 4947–4951 (2001).
47. 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).
48. Mastronarde, D.N. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36–51 (2005).
49. Guo, H. et al. Electron-event representation data enable efficient cryoEM file storage with full preservation of spatial and temporal resolution. *IUCrJ* **7**, 860–869 (2020).
50. Nakane, T. et al. Single-particle cryo-EM at atomic resolution. *Nature* (2020).
51. 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).
52. Zheng, S.Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331–332 (2017).
53. Scheres, S.H.W. Amyloid structure determination in RELION-3.1. *Acta Crystallogr D Struct Biol* **76**, 94–101 (2020).
54. Pettersen, E.F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612 (2004).
55. Gres, A.T. et al. STRUCTURAL VIROLOGY. X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability. *Science* **349**, 99–103 (2015).
56. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132 (2004).
57. Afonine, P.V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531–544 (2018).

- 609 58. Beckers, M., Jakobi, A.J. & Sachse, C. Thresholding of cryo-EM
610 density maps by false discovery rate control. *IUCrJ* **6**, 18–33 (2019).
611 59. Jakobi, A.J., Wilmanns, M. & Sachse, C. Model-based local density
612 sharpening of cryo-EM maps. *Elife* **6**(2017).
613

Figure legends

Figure 1 | CryoET and subtomogram averaging of Gag T8I assemblies and VLPs. (a)

Organization of the two constructs used for sample production: Gag Δ MA T8I, used for Gag assemblies made in *E. coli* cells; and GagT8I, used for VLP production in HEK293T cells. Both constructs bear the T8I mutation in the SP1 region (T371I in Gag sequence). (b-c) Purified Gag Δ MA T8I assemblies from *E. coli* cells are shown in a representative tomographic slice (b) and emClarity template matching result (c). (d-e) Purified GagT8I VLPs produced in HEK293T cells shown in a representative tomographic slice (d) and emClarity template matching result (e). (f) Subtomogram average of Gag Δ MA T8I assemblies in an extended lattice, shown in top and side views. The central Gag Δ MA T8I hexamer is highlighted in yellow and within which a monomer is highlighted in blue (CA_{NTD}) and orange (CA_{CTD}-SP1). Scale bars are 50 nm. Each point in (c) and (e) represents the position of a Gag hexamer.

Figure 2 | Structures of immature GagT8I CA-SP1 by subtomogram averaging. (a-b)

The density maps of immature Gag CA-SP1 from Gag Δ MA T8I assemblies (a) and GagT8I VLPs (b), overlaid with the refined respective atomic models shown in blue (CA_{NTD}) and orange (CA_{CTD}-SP1). Domain regions are labeled. (c) Enlarged views of CypA-binding loop (CypA-BL) and N-terminal MA-CA linker in Gag Δ MA T8I assemblies (top) and GagT8I VLPs (bottom). (d) Detailed views of the internal densities within the six-helix bundle of Gag Δ MA T8I assemblies (left), GagT8I VLPs (middle) and Gag Δ MA T8I assemblies with 10 μ M IP6 (right). Black arrows point to the lysine residues from CA_{CTD}, K158 and K227, forming two-rings to coordinate a central density (red arrows). Blue arrows point to residues CA L231 in Gag Δ MA T8I assemblies and SP1 M4 in GagT8I VLPs. Dashed lines mark the same height in both density maps.

Figure 3 | The T8I mutation stabilizes the six-helix bundle and impairs proteolytic

processing of cleavage sites flanking SP1. (a-b) Comparison of the CA-SP1 hexamer in

Gag Δ MA T8I (a) and Gag Δ MA WT (EMD-3782) (b), shown in a side-view. The density map is overlaid with the respective atomic model. The dashed arrows point to the position of T8I in both maps. (c) An enlarged view of density overlay from the dashed box region in a and b.

The Gag Δ MA T8I density is in grey and the Gag Δ MA WT density is in blue. The red arrow points to the same central density shown in Figure 2d. The pink arrow points to BVM density in

EMD-3782. (d) Overlay of the refined atomic models from Gag_{ΔMA}T8I and Gag_{ΔMA}WT (PDB 5L93). The arrow points to the location of the T8I mutation and CA|SP1 and SP1|NC cleavage sites. (e-f) *In-vitro* PR cleavage assays. Gag_{ΔMA}WT and Gag_{ΔMA}T8I assemblies were incubated with recombinant HIV-1 PR at the indicated concentration (e) and for the indicated period of time (f), separated by SDS-PAGE and stained with Coomassie Blue. The protein bands are labelled.

Figure 4 | Characterization of the NC-RNA density layer. (a) Radial density profile of the Gag lattice. Top panel shows a tomographic slice of Gag_{ΔMA}T8I lattice domain organization where CA_{NTD}, CA_{CTD} and NC-RNA density layers are coloured in blue, orange and pink, respectively. Graph shows the radial density profiles of Gag_{ΔMA}T8I particles of similar size (n=3). Black line indicates the average intensity while blue lines depict the individual particle profiles. Density profile peaks corresponding to CA_{NTD}, CA_{CTD} and NC-RNA are shaded with the same colour scheme as in the top panel. (b) Top view from density map of six-helix bundle and the NC-RNA layer, overlaid with two plausible NC-RNA (ψ SL2 loop) models, red and pink, fitted into the density, leaving the remaining NC-RNA density unmodeled. (c-d) Enlarged side view from density map in (b).

666 **Table 1: Cryo-EM data collection, refinement and validation statistics**

	Gag _Δ MA T8I assemblies	GagT8I VLPs	Gag _Δ MA T8I assemblies + IP-6
Data acquisition and processing			
EM equipment	Titan Krios	Titan Krios	Titan Krios
Voltage (kV)	300	300	300
Detector	Gatan K2	Gatan K2	Falcon 4
Energy filter	Gatan GIF Quantum, 20 eV slit	Gatan GIF Quantum, 20 eV slit	Selectris X, 10 eV slit
Super-resolution mode	Yes	Yes	Yes
Pixel size (Å)	1.35	1.35	1.196
Total electron dose (e ⁻ /Å ²)	120	110	122
Dose rate (e ⁻ /Å ² /s)	3	3	3
Frame number	10	5	10
Acquisition scheme	-60/60°,3°	-54/54°,3°	-60/60°,3°
Defocus range (μm)	-1.0 ~ -4.0	-1.5 ~ -3.5	-1.0 ~ -3.5
Number of tilt-series	90	46	13
Software	SerialEM	SerialEM	TEM Tomography
Number of tomograms	90	46	13
Number of subtomograms	131,942	50,343	29,938
Symmetry imposed	C6	C6	C6
Resolution at 0.143 FSC	4.2 Å	4.5 Å	4.2 Å
Software	IMOD/emClarity	IMOD/emClarity	IMOD/emClarity
EMDB number	emd-11894	emd-11897	emd-11899
Refinement			
Number of residues	4230 (18 chains of 235 residues)	4230 (18 chains of 235 residues)	
All-atom clash score	22.36	27.09	
Favored rotamers	100%	99.48%	
Ramachandran outliers	0	0	
Ramachandran allowed	2.16%	3%	
Ramachandran favored	97.84%	97%	
C-beta deviations	0.93 %	0	
RMSD (angles, degree)	1.370 (162)	1.502 (90)	
RMSD (bonds, Å)	0.008 (0)	0.010 (0)	
PDB accession code	7ASH	7ASI	

667

668

Figure 1 | CryoET and subtomogram averaging of Gag T8I assemblies and VLPs. (a)

Organization of the two constructs used for sample production: Gag_{ΔMA}T8I, used for Gag assemblies made in *E. coli* cells; and GagT8I, used for VLP production in Hek-293T cells. Both constructs bear the T8I mutation in SP1 region (T371I in Gag sequence). (b-c) Purified Gag_{ΔMA}T8I assemblies from *E. coli* cells are shown in a representative tomographic slice (b) and emClarity template matching result (c). (d-e) Purified GagT8I VLPs produced in Hek-293T cells shown in a representative tomographic slice (d) and emClarity template matching result (e). (f) Subtomogram average of Gag_{ΔMA}T8I assemblies in an extended lattice, shown in top and side views. The central Gag_{ΔMA}T8I hexamer is highlighted in yellow and within which a monomer is highlighted in blue (CA_{NTD}) and orange (CA_{CTD}-SP1). Scale bars are 50 nm. Each point in (c) and (e) represents the position of a Gag hexamer.

Figure 2 | Structures of immature GagT8I CA-SP1 by subtomogram averaging. (a-b) The density maps of immature Gag CA-SP1 from Gag Δ MA T8I assemblies (a) and GagT8I VLPs (b), overlaid with the refined respective atomic models shown in blue (CA_{NTD}) and orange (CA_{CTD}-SP1). Domain regions are labeled. (c) Enlarged views of CypA-binding loop (CypA-BL) and N-terminal MA-CA linker in Gag Δ MA T8I assemblies (top) and GagT8I VLPs (bottom). (d) Detailed views of the internal densities within the six-helix bundle of Gag Δ MA T8I assemblies (left), GagT8I VLPs (middle) and Gag Δ MA T8I assemblies with 10 μ M IP-6 (right). Black arrows point to the lysine residues from CA_{CTD}, K158 and K227, forming two-rings to coordinate a central density (red arrows). Blue arrows point to residues CA L231 in Gag Δ MA T8I assemblies and SP1 M4 in GagT8I VLP. Dashed lines mark the same height in both density maps.

Figure 3 | T8I mutation stabilizes the six-helix bundle and impairs proteolytic processing of cleavage sites flanking SP1. (a-b) Comparison of the CA-SP1 hexamer in $Gag_{\Delta MA} T8I$ (a) and $Gag_{\Delta MA} WT$ (EMD-3782) (b), shown in a side-view. The density map is overlaid with the respective atomic model. The dashed arrows points to the position of T8I in both maps. (c) An enlarged view of density overlay from the dashed box region in a and b. The $Gag_{\Delta MA} T8I$ density is in grey and the $Gag_{\Delta MA} WT$ density is in blue. Red arrow points to the same central density shown in Figure 2d. Pink arrow points to BVM density in EMD-3782. (d) Overlay of the refined atomic models from $Gag_{\Delta MA} T8I$ and $Gag_{\Delta MA} WT$ (PDB 5L93). Arrow points to location of T8I mutation and CA|SP1 and SP1|NC cleavage sites. (e-f) *In-vitro* protease cleavage assays. $Gag_{\Delta MA} WT$ and $Gag_{\Delta MA} T8I$ assemblies were incubated with recombinant HIV-1 PR at the indicated concentration (e) and for the indicated period of time (f), separated in an SDS-PAGE and stained with Coomassie Blue. The protein bands are labelled.

Figure 4 | Characterization of NC-RNA density layer. (a) Radial density profile of Gag lattice. Top panel shows a tomographic slice of Gag_{ΔMA}T8I lattice domain organization where CA_{NTD}, CA_{CTD} and NC-RNA density layers are coloured in blue, orange and pink, respectively. Graph shows the radial density profiles of Gag_{ΔMA}T8I particles of similar size (n=3). Black line indicates the average intensity while blue lines depict the individual particle profiles. Density profile peaks corresponding to CA_{NTD}, CA_{CTD} and NC-RNA are shaded with the same colour scheme as in the top panel. (b) Top view from density map of six-helix bundle and the NC-RNA layer, overlaid with two plausible NC-RNA (ψ SL2 loop) models, red and pink, fitted into the density, leaving the remaining NC-RNA density unmodeled. (c-d) Enlarged side view from density map in (b).