LETTER

doi:10.1038/nature13838

Structure of the immature HIV–1 capsid in intact virus particles at 8.8 Å resolution

Florian K. M. Schur[1,2], Wim J. H. Hagen[1], Michaela Rumlová[3,4], Tomáš Ruml[5], Barbara Müller[2,6], Hans-Georg Kräusslich[2,6] & John A. G. Briggs[1,2]

Human immunodeficiency virus type 1 (HIV-1) assembly proceeds in two stages. First, the 55 kilodalton viral Gag polyprotein assembles into a hexameric protein lattice at the plasma membrane of the infected cell, inducing budding and release of an immature particle. Second, Gag is cleaved by the viral protease, leading to internal rearrangement of the virus into the mature, infectious form[1]. Immature and mature HIV-1 particles are heterogeneous in size and morphology, preventing high-resolution analysis of their protein arrangement *in situ* by conventional structural biology methods. Here we apply cryo-electron tomography and sub-tomogram averaging methods to resolve the structure of the capsid lattice within intact immature HIV-1 particles at subnanometre resolution, allowing unambiguous positioning of all α-helices. The resulting model reveals tertiary and quaternary structural interactions that mediate HIV-1 assembly. Strikingly, these interactions differ from those predicted by the current model based on *in vitro*-assembled arrays of Gag-derived proteins from Mason–Pfizer monkey virus[2]. To validate this difference, we solve the structure of the capsid lattice within intact immature Mason–Pfizer monkey virus particles. Comparison with the immature HIV-1 structure reveals that retroviral capsid proteins, while having conserved tertiary structures, adopt different quaternary arrangements during virus assembly. The approach demonstrated here should be applicable to determine structures of other proteins at subnanometre resolution within heterogeneous environments.

HIV-1 Gag consists of separately folded domains connected by flexible linkers. Its amino (N)-terminal matrix (MA) domain directs Gag to the plasma membrane, and the subsequent two capsid (CA) domains mediate protein–protein interactions in the mature and immature particle; they are followed by the RNA-binding nucleocapsid (NC) domain. Spacer-peptides SP1 and SP2 are positioned between CA and NC, and between NC and the carboxy (C)-terminal p6 domain, respectively[3]. The immature protein shell constitutes a hexameric Gag array bound to the viral membrane, forming a truncated sphere with irregular defects[4]. Proteolytic cleavage at five positions in Gag during maturation leads to a dramatic structural rearrangement, with cleaved CA forming the characteristic conical HIV-1 core encapsulating the now condensed NC–RNA complex.

High-resolution structures of all isolated Gag domains are available[5], and a structural model of the mature HIV-1 capsid core has recently been generated[6]. However, the flexible nature of the uncleaved polyprotein has made structural studies of the immature Gag conformation more challenging. We recently described the structure of an *in vitro*-assembled immature-like tubular array of a Mason–Pfizer monkey virus (M-PMV) Gag-derived protein at ~8 Å resolution[2], which currently represents the most detailed model for an immature Gag lattice. For immature HIV-1, a partial model has been generated based on *in vitro*-assembled tubular arrays of Gag-derived protein in which a mutation stabilizes the C-terminal CA domain (CA-CTD) in an immature-like form, while the N-terminal CA domain (CA-NTD) adopts a presumably non-physiological form[7].

Subtomogram averaging has been used to solve low-resolution structures of biological molecules within pleiotropic native environments[8], including structures of the immature HIV-1 Gag lattice within virus particles at ~20 Å resolution[4,9]. Higher-resolution structures have not yet been obtained for any component of such heterogeneous viruses *in situ*.

Recently, we showed that optimized subtomogram averaging methods can recover structural data at below 10 Å resolution[10]. Their application to immature HIV-1 might enable determination of a subnanometre-resolution structure of Gag within intact virus. Therefore, we purified immature HIV-1 particles treated with the protease inhibitor amprenavir (APV) (Extended Data Fig. 1a) and prepared them for cryo-electron tomography. We collected and reconstructed 70 cryo-electron tomograms as described[10] and segregated them into two data sets for independent processing. Reconstructed tomograms contained 224 immature HIV-1 particles with variable diameters and the characteristic striated density corresponding to the Gag lattice[11] (Fig. 1a, Extended Data Fig. 2a and Extended Data Table 1). Subtomograms extracted along the capsid layers were iteratively aligned and averaged in three dimensions. The aligned positions of subtomograms within individual virus particles indicated a continuous but incomplete ~8 nm spaced hexameric lattice as previously described[4] (Extended Data Fig. 2a). The CA layer within the final average of all data was resolved at 8.8 Å (Extended Data Fig. 2b), showing well-resolved rod-shaped densities for α-helices (Fig. 1b–d and Extended Data Fig. 2d–f). The general domain arrangement seen in the APV-inhibited HIV-1 was confirmed in an 11 Å resolution structure of HIV-1 with an inactivating protease mutation (Extended Data Figs 1b and 2d–f).

The final structure shows hexamerically arranged densities in the CA and SP1 regions, similar to those previously described at low resolution[4,9] (Fig. 1b–d). No well-ordered densities were observed for MA or NC (Extended Data Fig. 2c), indicating that they do not adopt a defined spatial arrangement relative to CA, consistent with the flexibility of the inter-domain linkers. In the region presumably corresponding to SP1, six rod-like densities were observed, arranged in a tube around a hollow centre at the sixfold symmetry axis (Fig. 1d and Supplementary Video 1), consistent with previous models suggesting this region assembles a six-helix bundle[7,9,12].

We performed rigid-body fitting of high-resolution structures for CA-NTD (Protein Data Bank (PDB) 1L6N[13]) and CA-CTD (PDB 3DS2 (ref. 14)) into the outer and inner layers of the CA density, respectively, and refined the fit using molecular dynamics flexible fitting[15] (Extended Data Fig. 3a, b). This provided a model showing the tertiary and quaternary interactions in the CA layer of the immature virus (Figs 1b–d, 2 and Supplementary Video 1). For both domains a single unambiguous fit was identified accommodating all 11 α-helices of CA within appropriate

[1]Structural and Computational Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. [2]Molecular Medicine Partnership Unit, European Molecular Biology Laboratory/Universitätsklinikum Heidelberg, Heidelberg, Germany. [3]Institute of Organic Chemistry and Biochemistry (IOCB), Academy of Sciences of the Czech Republic, v.v.i., IOCB & Gilead Research Center, Flemingovo nám. 2, 166 10 Prague, Czech Republic. [4]Department of Biotechnology, Institute of Chemical Technology, Prague, Technická 5, 166 28, Prague, Czech Republic. [5]Department of Biochemistry and Microbiology, Institute of Chemical Technology, Prague, Technická 5, 166 28, Prague, Czech Republic. [6]Department of Infectious Diseases, Virology, Universitätsklinikum Heidelberg, Im Neuenheimer Feld 324, 69120 Heidelberg, Germany.

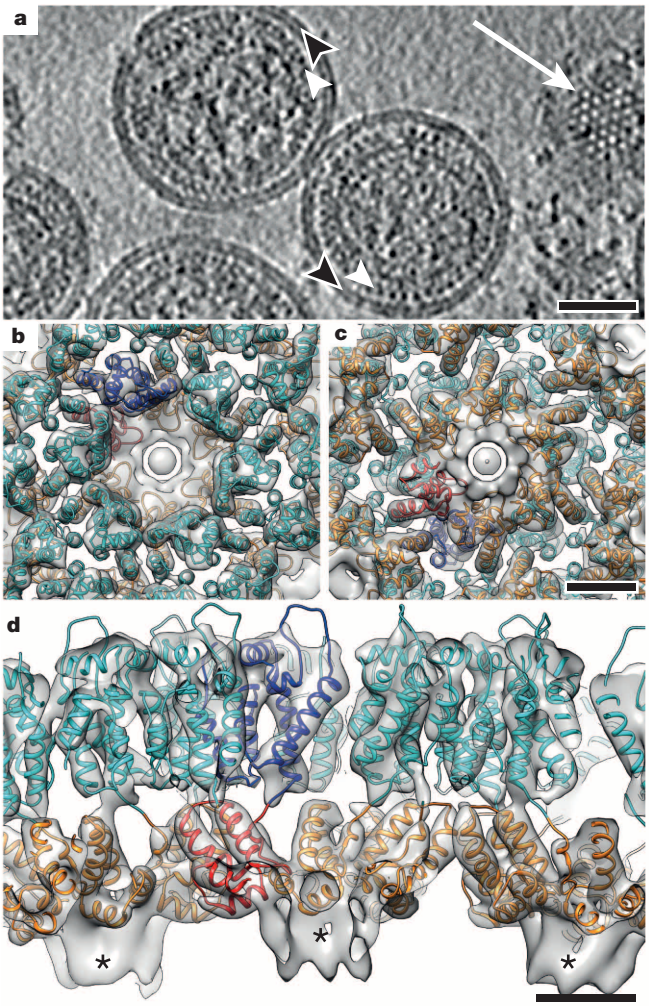


Figure 1 | Structure of the CA-SP1 lattice within immature HIV-1. **a**, Computational slice through a Gaussian-filtered tomogram containing immature HIV-1 particles treated with APV. White arrowheads indicate the immature CA layer; black arrowheads indicate the membrane; white arrow marks a grazing slice through the CA layer illustrating the hexagonal lattice. Scale bar, 50 nm. **b**–**d**, Isosurface representation of the final structure showing the immature CA-SP1 lattice viewed from outside the virus (**b**), from inside (**c**) and in an orthogonal view (**d**). Isosurface threshold value is 2σ away from the mean. High-resolution structures for CA-NTD (cyan) and CA-CTD (orange) have been fitted into the density. An individual capsid monomer is coloured blue/red. Unfilled densities marked with asterisks correspond to the SP1 region. Scale bar, 25 Å. See also Supplementary Video 1.

densities in the electron microscopy map. At lower isosurface thresholds, densities corresponding to the peptide chain linking helix 7 in CA-NTD with helix 8 in CA-CTD, as well as the cyclophilin-A binding loop, were observed (Extended Data Fig. 3b–d and Supplementary Video 2).

The arrangement of the CA-CTD is similar to those in partly immature-like *in vitro*-assembled tubular arrays of HIV-1 (ref. 7) (Extended Data Fig. 4) and tubular arrays of M-PMV Gag[2]. Between hexamers, two CA-CTD domains form a homo-dimeric interface mediated by helix 9 (Fig. 2c). The positions of assembly-critical hydrophobic residues W316 and M317 in helix 9 (refs 16, 17) are consistent with a contribution to the interaction interface (Fig. 2f). Around the hexamer, contacts between adjacent CA-CTD monomers include part of the major homology region (I285–L304)[18], consistent with a role for this region in stabilizing the hexamer. Other residues, where mutations are known to disrupt viral assembly[16,17], are close to inter-protein interfaces (Fig. 2d–f). These observations point to the central role of CA-CTD in assembling and stabilizing the immature lattice.

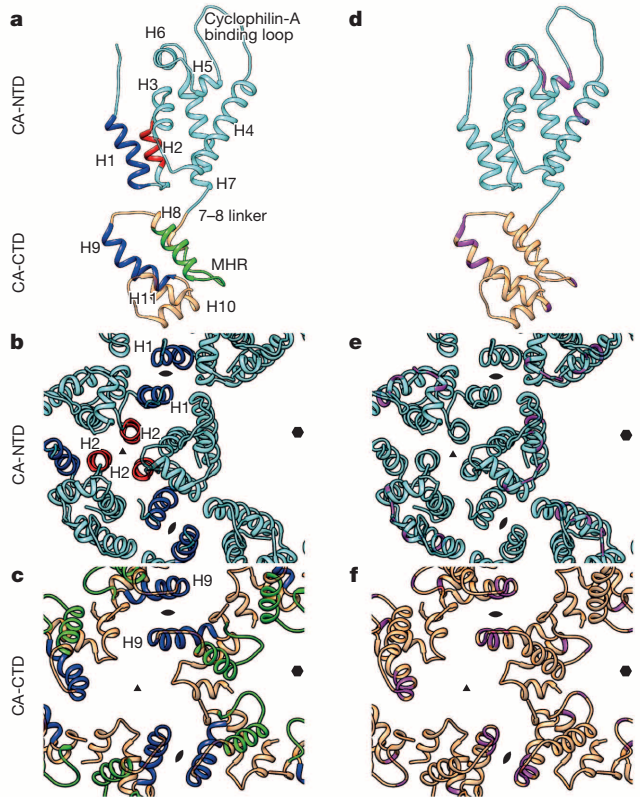


Figure 2 | Interactions stabilizing the immature HIV-1 capsid lattice. **a**, An annotated CA monomer, indicating helices contributing to homo-dimeric and trimeric CA-NTD and CA-CTD interactions in blue and red, respectively. Interactions around the hexamer involve residues in the major homology region (green). **b**, The CA-NTD can form intra-hexameric interactions, dimeric inter-hexameric contacts between helices 1 and trimeric inter-hexameric interactions between helices 2. **c**, In the CA-CTD a dimeric interaction between helices 9 is observed. **d**–**f**, Views as in **a**–**c**: residues in which mutations disrupt particle assembly are coloured magenta. Sixfold, threefold and twofold symmetry axes are indicated by black hexagons, triangles and ellipses, respectively.

The CA-NTD and CA-CTD abut each other at two positions in the immature lattice (Supplementary Video 2): close to residues S310–Q311 in helix 9 and E161–K162 in helix 1, and close to residues M276–S278 in helix 7 and K302–R305 in helix 8 (Extended Data Fig. 3c, d). Unlike in the mature capsid, there is no large CA-CTD:CA-NTD interface[19,20]. This is consistent with the ability of CA-CTD to assemble an immature-like arrangement independent of the CA-NTD arrangement *in vitro*[7], and to induce release of virus-like particles from cells in Gag protein mutants lacking CA-NTD[21].

Virus-like particles lacking CA-NTD have aberrant size distributions[21], indicating a role for CA-NTD in determining particle morphology. Within the immature virus, each CA-NTD domain can interact with five neighbouring CA-NTD domains (Fig. 2b). Around the hexamer, the top of helix 4 in each CA-NTD interacts with a patch of residues between helices 5 and 6 in the neighbouring CA-NTD (Supplementary Video 2). A homo-dimeric interface is formed between helix 1 from two adjacent hexamers. A homo-trimeric, three-helix bundle interface is formed by helix 2 from neighbouring hexamers. The weaker density of helix 2 compared with other helices within our electron microscopy density suggests it may have some flexibility within the lattice. Together these interfaces comprise an extensive network of interactions. The integrity of the lattice could be maintained by any combination of two of these three interfaces, and mutation in a single interface may therefore not interfere with assembly. This may explain why mutation of CA-NTD residues close to inter-hexamer protein–protein interfaces has a mild or no effect on virus assembly[16,22].

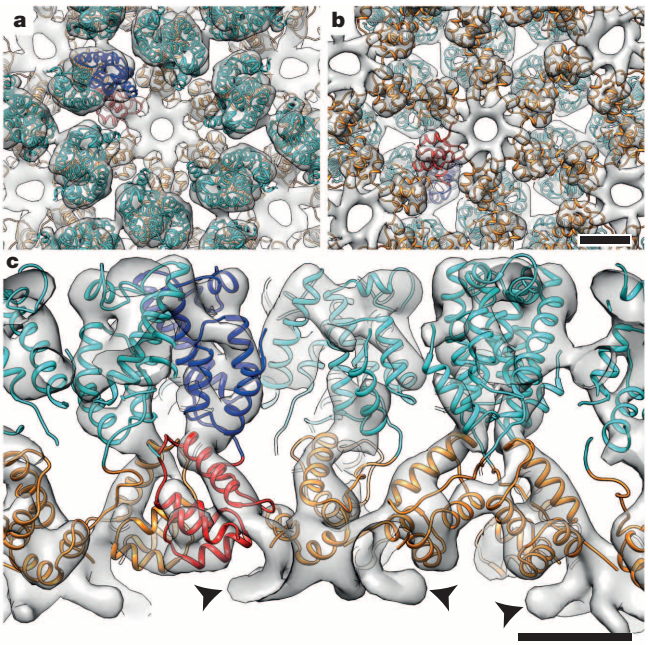


Figure 3 | Structure of the capsid lattice within immature M-PMV (D26N) particles. **a**–**c**, Isosurface representation of the immature M-PMV CA lattice viewed from outside the virus (**a**), from inside (**b**) and in an orthogonal view (**c**). Partly resolved α-helices are visible, allowing positioning of high-resolution structures for CA-NTD and CA-CTD into the density. Colouring as in Fig. 1. The splayed densities annotated with black arrowheads correspond to the 'spacer-peptide-like' region in M-PMV. Isosurface treshold is 2.5σ away from the mean. Scale bar, 25 Å.

Mutation of certain residues in helices 4–6 has been shown to interfere with virus assembly[16]. While some of these residues may contribute to the intra-hexamer interfaces, others face the central hole of the hexamer (Fig. 2e). They may modulate the structural integrity of CA-NTD, represent interfaces required transiently during assembly, or be a binding site for unknown host factors regulating assembly. The arrangement of CA-NTD domains leaves reported binding sites of host cofactors such as cyclophilin-A accessible[23].

Surprisingly, the arrangement of CA-NTD domains is completely different from the current model for immature retroviruses based on the structure of *in vitro*-assembled tubular arrays of M-PMV Gag[2]. This could be due to a true difference between the immature lattices in HIV-1 and M-PMV, or because the tubular array structure represents an artefact of *in vitro* assembly. We purified intact immature M-PMV containing an inactivating protease mutation (D26N) (Extended Data Figs 1c and 5a). Following the same protocol as for HIV, we obtained a 9.7 Å resolution reconstruction of the CA layer (Extended Data Fig. 5b–f) from which we generated a structural model (Fig. 3). The arrangement of CA domains in intact immature M-PMV matches the arrangement within *in vitro*-assembled M-PMV tubular arrays[2] (Extended Data Fig. 6), and is indeed different from that observed in immature HIV-1 (Figs 1b–d and 4). Thus, although retroviral CA proteins have conserved tertiary structures (Extended Data Fig. 7), they adopt different quaternary arrangements within the immature virus (Extended Data Fig. 8). While HIV-1 CA-NTD forms an extensive network of interactions linking within and between hexamers, the primary interaction of M-PMV CA-NTD is a larger dimeric interface linking two neighbouring hexamers. How these different arrangements have evolved can only be speculated. We suspect their functional relevance relates to the site of assembly: the lentivirus HIV-1 makes use of the plasma membrane as a scaffold, while the betaretrovirus M-PMV assembles within the cytoplasm[24] and may require stronger interactions to link Gag hexamers.

The arrangement of CA in immature HIV-1 is very different from that suggested for the mature HIV-1 capsid core[6,19,25] (Extended Data Fig. 8). In both immature and mature states, CA-CTD dimerization is mediated by helix 9, but with a different relative orientation of the two monomers (Fig. 4 and Extended Data Fig. 4d). This and the dramatic rearrangement of the CA-NTD from immature to mature particles would require breaking essentially all interactions of the immature lattice and forming new contacts during maturation (Supplementary Video 3). This could be achieved by two routes: (1) by disassembly of the immature lattice into monomers or dimers linked by the helix 9 interface; (2) by disassembly of the immature lattice into hexamers or higher-order oligomers. Route 2 would imply a two-step 'molecular dance'. First, the hexameric interactions in the CA-CTD-SP1 region (which can form even when CA-NTD is not properly arranged[7]) could stabilize the hexamer while cleavage between MA and CA induces the CA-NTD to move around the flexible 7–8 linker into its mature hexameric arrangement.

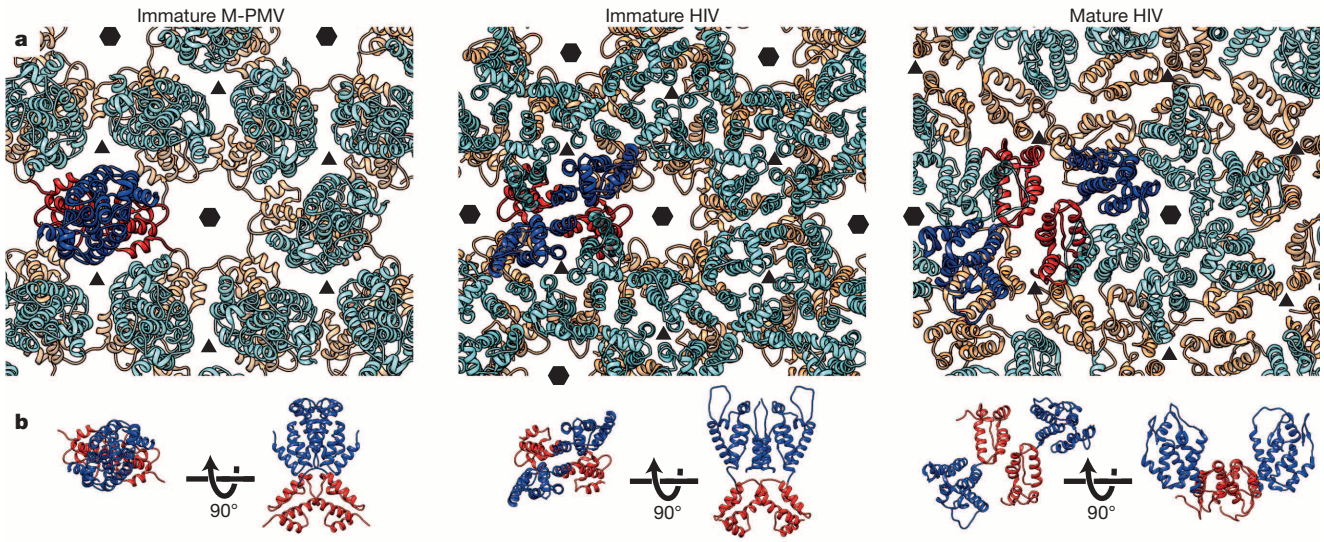


Figure 4 | Comparison of the CA arrangement in immature and mature retroviruses. **a**, The arrangement of the CA lattices in immature M-PMV and immature and mature HIV particles. CA-NTD and CA-CTD domains are coloured in cyan/blue and orange/red, respectively. See also Supplementary Video 3. Symmetry axes are marked as in Fig. 2. **b**, The structures of CA dimers from the lattices in **a** are shown in top and orthogonal views. CA-CTD interactions are always mediated by helices 9, although the dimer rotates around this interaction in the transition from the immature to the mature form (see Extended Data Fig. 4). CA-NTD interactions are in each case entirely different. See also Extended Data Fig. 8.

Second, these now mature CA-NTD interactions would stabilize the hexamer while cleavage between CA and SP1 induces maturation of CA-CTD.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 May; accepted 3 September 2014.
Published online 2 November 2014.

1. Sundquist, W. I. & Krausslich, H. G. HIV-1 assembly, budding, and maturation. *Cold Spring Harb. Perspec. Med.* **2,** a006924 (2012).
2. Bharat, T. A. *et al.* Structure of the immature retroviral capsid at 8 Å resolution by cryo-electron microscopy. *Nature* **487,** 385–389 (2012).
3. Bell, N. M. & Lever, A. M. HIV Gag polyprotein: processing and early viral particle assembly. *Trends Microbiol.* **21,** 136–144 (2013).
4. Briggs, J. A. *et al.* Structure and assembly of immature HIV. *Proc. Natl Acad. Sci. USA* **106,** 11090–11095 (2009).
5. Briggs, J. A. & Krausslich, H. G. The molecular architecture of HIV. *J. Mol. Biol.* **410,** 491–500 (2011).
6. Zhao, G. *et al.* Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **497,** 643–646 (2013).
7. Bharat, T. A. *et al.* Cryo-electron microscopy of tubular arrays of HIV-1 Gag resolves structures essential for immature virus assembly. *Proc. Natl Acad. Sci. USA* **111,** 8233–8238 (2014).
8. Briggs, J. A. Structural biology *in situ* — the potential of subtomogram averaging. *Curr. Opin. Struct. Biol.* **23,** 261–267 (2013).
9. Wright, E. R. *et al.* Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells. *EMBO J.* **26,** 2218–2226 (2007).
10. Schur, F. K., Hagen, W. J., de Marco, A. & Briggs, J. A. Determination of protein structure at 8.5 Å resolution using cryo-electron tomography and sub-tomogram averaging. *J. Struct. Biol.* **184,** 394–400 (2013).
11. Fuller, S. D., Wilk, T., Gowen, B. E., Kräusslich, H.-G. & Vogt, V. M. Cryo-electron microscopy reveals ordered domains in the immature HIV-1 particle. *Curr. Biol.* **7,** 729–738 (1997).
12. Accola, M. A., Höglund, S. & Göttlinger, H. G. A. Putative α-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).
13. Tang, C., Ndassa, Y. & Summers, M. F. Structure of the N-terminal 283-residue fragment of the immature HIV-1 Gag polyprotein. *Nature Struct. Biol.* **9,** 537–543 (2002).
14. Bartonova, V. *et al.* Residues in the HIV-1 capsid assembly inhibitor binding site are essential for maintaining the assembly-competent quaternary structure of the capsid protein. *J. Biol. Chem.* **283,** 32024–32033 (2008).
15. Trabuco, L. G., Villa, E., Schreiner, E., Harrison, C. B. & Schulten, K. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods* **49,** 174–180 (2009).
16. von Schwedler, U. K., Stray, K. M., Garrus, J. E. & Sundquist, W. I. Functional surfaces of the human immunodeficiency virus type 1 capsid protein. *J. Virol.* **77,** 5439–5450 (2003).
17. Chu, H.-H., Chang, Y.-F. & Wang, C.-T. Mutations in the α-helix directly C-terminal to the major homology region of human immunodeficiency virus type 1 capsid protein disrupt gag multimerization and markedly impair virus particle production. *J. Biomed. Sci.* **13,** 645–656 (2006).
18. Mammano, F., Ohagen, A., Höglund, S. & Göttlinger, H. G. Role of the major homology region of human immunodeficiency virus type 1 in virion morphogenesis. *J. Virol.* **68,** 4927–4936 (1994).
19. Pornillos, O. *et al.* X-ray structures of the hexameric building block of the HIV capsid. *Cell* **137,** 1282–1292 (2009).
20. Monroe, E. B., Kang, S., Kyere, S. K., Li, R. & Prevelige, P. E., Jr. Hydrogen/deuterium exchange analysis of HIV-1 capsid assembly and maturation. *Structure* **18,** 1483–1491 (2010).
21. Borsetti, A., Öhagen, Å. & Göttlinger, H. G. The C-terminal half of the human immunodeficiency virus type 1 gag precursor is sufficient for efficient particle assembly. *J. Virol.* **72,** 9313–9317 (1998).
22. von Schwedler, U. K. *et al.* Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly. *EMBO J.* **17,** 1555–1568 (1998).
23. Franke, E. K., Yuan, H. E. H. & Luban, J. Specific incorporation of cyclophilin A into HIV-1 virions. *Nature* **372,** 359–362 (1994).
24. Chopra, H. C. & Mason, M. M. A new virus in a spontaneous mammary tumor of a rhesus monkey. *Cancer Res.* **30,** 2081–2086 (1970).
25. Pornillos, O., Ganser-Pornillos, B. K. & Yeager, M. Atomic-level modelling of the HIV capsid. *Nature* **469,** 424–427 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements This study was supported by Deutsche Forschungsgemeinschaft grants BR 3635/2-1 to J.A.G.B., KR 906/7-1 to H.-G.K. and by Grant Agency of the Czech Republic 14-15326S to M.R. The Briggs laboratory acknowledges financial support from the European Molecular Biology Laboratory and from the Chica und Heinz Schaller Stiftung. We thank B. Glass, M. Anders and S. Mattei for preparation of samples, and R. Hadravova, K. H. Bui, F. Thommen, M. Schorb, S. Dodonova, S. Glatt, P. Ulbrich and T. Bharat for technical support and/or discussion. This study was technically supported by the European Molecular Biology Laboratory IT services unit.

Author Contributions F.K.M.S., M.R., T.R., B.M., H.-G.K. and J.A.G.B. designed and interpreted experiments. F.K.M.S. and W.J.H.H. collected data, F.K.M.S. performed image processing, and F.K.M.S. and J.A.G.B. analysed data. F.K.M.S. and J.A.G.B. wrote the manuscript with support from all authors.

Author Information Cryo-electron microscopy structures and a representative tomogram have been deposited in the Electron Microscopy Data Bank under accession numbers EMD-2706, EMD-2707 and EMD-2708, and the fitted HIV atomic model in the PDB under accession number 4USN. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.A.G.B. (john.briggs@embl.de).

METHODS

Virus particle production and purification. HEK293T cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum (FCS; Biochrom), penicillin (100 IU ml^{-1}), streptomycin (100 μg ml^{-1}) and 4 mM glutamine. For the protease-inhibitor-treated sample, cells were transfected with proviral plasmid pNLC4-3 (ref. 26) using polyethyleneimine following standard procedures. A final concentration of 5 μM amprenavir[27] (obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, National Institute of Allergy and Infectious Diseases, National Institutes of Health) was added to the tissue culture medium at the time of transfection. For the protease mutant HIV and M-PMV sample, cells were transfected with plasmids pNL4-3(PR-) (in case of HIV D25A) or pSHRM15 (for M-PMV D26N). At 44 hours post-transfection, tissue culture supernatant was harvested and filtered through 0.45 μm nitrocellulose filters. Virus was enriched by ultracentrifugation through a 20% (w/w) sucrose cushion and further purified by centrifugation through an iodixanol density gradient[28]. Concentrated virus samples were re-suspended in 25 mM MES pH 6.0, 150 mM NaCl or phosphate buffered saline (PBS), and treated with 1.2% paraformaldehyde for 1 h on ice and stored in aliquots at −80 °C. Purities of samples and Gag maturation state were verified by SDS–polyacrylamide gel electrophoresis followed by silver staining and immunoblotting (Extended Data Fig. 1).

Cryo-electron microscopy and image processing. Degassed C-Flat 2/2-2C or 2/2-3C grids were glow discharged for 30 s at 20 mA. Virus solution was diluted with PBS containing 10 nm colloidal gold and 2 μl of this mixture was then applied to grids and plunge frozen in liquid ethane using an FEI Vitrobot Mark 2. The grids were stored under liquid nitrogen conditions until imaging.

Data acquisition and image processing for immature HIV treated with APV and the immature M-PMV protease mutant (D26N) was performed as previously described[10] (Extended Data Table 1). In brief, tilt series were imaged on a FEI Titan Krios electron microscope operated at 200 keV, with a GIF2002 post-column energy filter (using a slit width of 20 eV) and a 2k × 2k Gatan Multiscan 795 CCD (charge-coupled device) camera. For navigation and search purposes, low-magnification montages were acquired using SerialEM[29], and tilt series were acquired at appropriate positions using FEI tomography software version 4 in automated batch mode. The nominal magnification was 42,000, giving a calibrated pixel size of 2.025 Å. The tilt range was from −45° to +60° in 3° steps, collecting first from 0° to −45° and then from 3° to 60°. Tilt series were collected at a range of nominal defoci between −1.2 and −4.5 μm. The total dose applied to each tilt series was approximately 40 e $Å^{-2}$. Tomograms were reconstructed using the IMOD software suite[30].

Image acquisition for protease defective HIV (D25A) was performed on an FEI Titan Krios electron microscope operated at 300 keV equipped with a Falcon II direct electron detector. Acquisition parameters were the same as described above, but using defoci ranging from −2.0 to −5.5 μm and having a calibrated pixel size of 2.281 Å per pixel. The total dose was approximately 60 e $Å^{-2}$. The tilt range was either from −45° to +60° in 3° steps or from −42° to +42° in 3° steps.

Image processing was performed identically for all samples. Before image processing, tomograms were split into two half-data sets with each set containing roughly the same number of virus particles and the same distribution of defoci, to obtain two independent reconstructions. Subtomogram averaging calculations were performed using Matlab scripts derived from the AV3 (ref. 31) and TOM[32] packages. Generation of masks and alignment of the two final references before Fourier shell correlation (FSC) measurement was done using the Dynamo software package[33]. Visualization of tomograms was performed in either IMOD[30] or Amira (FEI Visualization Sciences Group) using the electron microscopy toolbox[34].

Initial processing was performed on 4× binned, non-contrast transfer function (CTF) corrected data. For each half-data set, one tomogram was chosen to obtain an initial structure. Cubic subtomograms with a size of $(340 Å)^3$ were extracted from the surface of a sphere with its centre being the centre of the virus and the radius corresponding to the mean radius at the CA-level. The subtomograms were assigned initial angles based on the geometry of the sphere and then averaged to obtain a smooth reference. Subtomograms were aligned and averaged in an iterative manner against this reference in six dimensions to obtain an initial structure[31]. Subsequently all subtomograms within each half-data set were aligned and averaged against their respective independent starting reference for six iterations. After the first two iterations a cross-correlation-based cleaning was performed to remove subtomograms that contained no density corresponding to the Gag layer.

The defocus of each tomogram was measured by fitting theoretical CTF-curves to averaged power spectra from 512 pixel × 512 pixel tiles generated from all images in a tilt series using Matlab. CTF correction used the program 'ctf phase flip' implemented in IMOD[35]. Subvolumes with a size of $(284 Å)^3$ were extracted from unbinned, CTF-corrected tomograms at the positions determined in the 4× binned alignments, and subjected to four further alignment iterations. The sixfold symmetry inherent in the structure was applied at each step. The two final references were aligned, multiplied with a Gaussian filtered mask and the FSC was calculated. For the APV-treated HIV sample, comparison of the averages of all subtomograms after alignment (174,360 and 168,144 asymmetric units in each of the half-data sets, respectively) by FSC indicated a resolution of 8.9 Å. After removing subtomograms with lower cross-correlation values, final averages were generated from 94,986 and 99,744 asymmetric units in each of the half-data sets, giving a slightly improved resolution of 8.8 Å, but not perceptibly changing the structure. Subsequently, we averaged the two half-data set maps and sharpened them with an empirically determined negative B-factor of $-1{,}200 Å^2$, while filtering to the resolution determined at the 0.143 FSC threshold[36].

For the M-PMV protease mutant (D26N) sample, 39 tomograms were reconstructed from which 84 viruses were then extracted. The processing was performed as for the protease-inhibitor treated HIV sample. A total of 77,520 asymmetric units from both half-data sets were averaged to generate the final volume and the resolution was determined by FSC to be 9.7 Å at the 0.143 criterion. Finally the structure was sharpened applying a negative B-factor of $-1{,}000 Å^2$, while filtering to the resolution determined at the 0.143 FSC threshold.

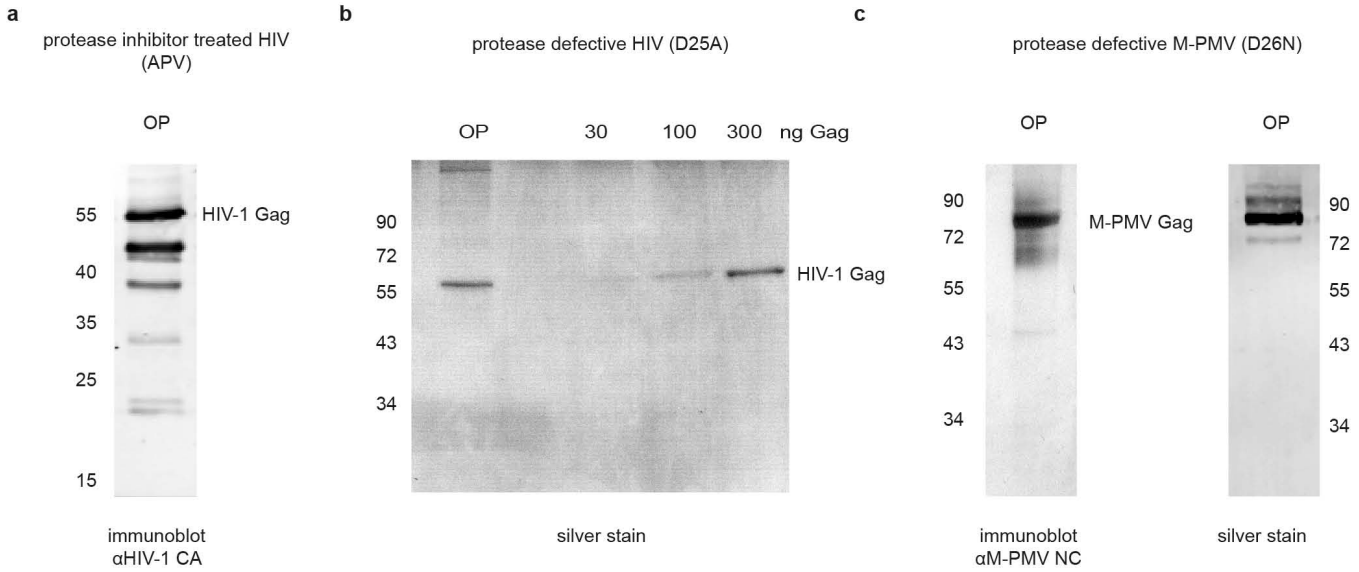
For a summary of image-processing statistics and information on the protease defective HIV (D25A) sample, see Extended Data Table 1.

Structure fitting and analysis. Visualization and rigid body docking was performed in UCSF Chimera[37]. For flexible fitting into the protease-inhibitor treated HIV map, the nuclear magnetic resonance structure for CA-NTD (PDB 1L6N, chain 1) and the crystal structure for CA-CTD (PDB 3DS2, one monomer) were used. Initially, both models were rigid-body docked into the electron microscopy density to establish their correct relative position using the 'Fit in map' option and were then joined using the 'Adjust bond' option in Chimera. We truncated the joined model from residues 142 to 351 and added PDB 1U57 (ref. 38) (chain 3) (residues GVGGPG HKARVLAEAMSQVT) to generate a preliminary model of the SP1 region. Finally, 30 copies of the joined monomers were rigid-body docked into the electron microscopy map. The outer 12 capsid monomers and all residues corresponding to SP1 were included in the simulations to prevent excessive movement of residues into empty densities but were removed after the simulation and not interpreted. All flexible fitting simulations were performed using molecular dynamics flexible fitting[15] and NAMD version 2.9 (ref. 39). The simulations were performed until the backbone atoms stopped moving and the fit stabilized. The flexible fitting resulted in only minor changes (Extended Data Fig. 3a, b), highlighting the strong initial correspondence between our map and published models of the individual domains.

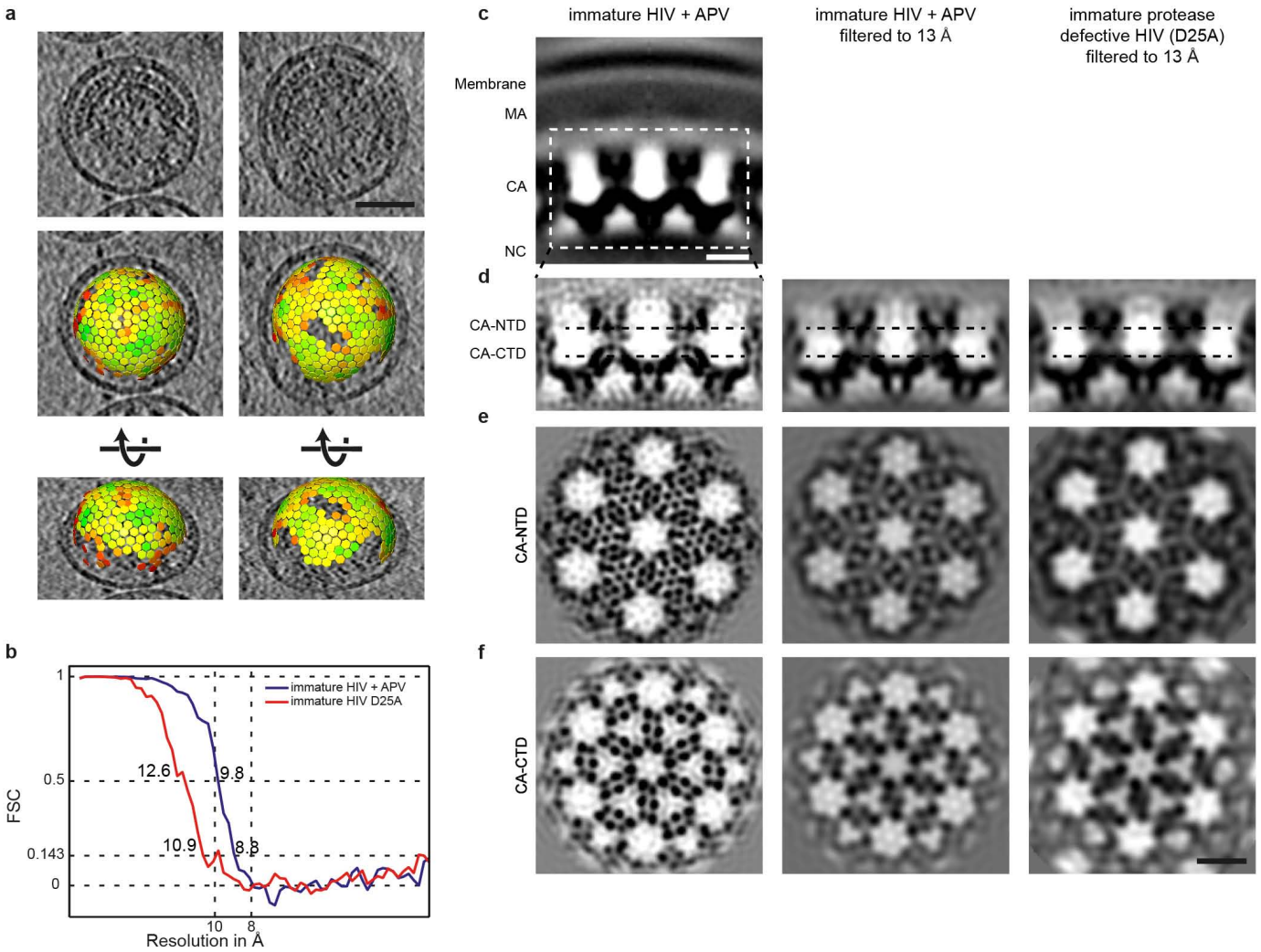
To generate the structural model for M-PMV, the PDB for CA-NTD 2KGF[40] (chain 1, residues 317–439) was combined with a homology model of M-PMV CA-CTD (residues 440–516) based on PDB 3H4E[19] (chain L) and fitted into the volume as a rigid body. The mature HIV-1 structure presented for comparison is PDB 3J34 (ref. 6).

26. Bohne, J. & Kräusslich, H.-G. Mutation of the major 5′ splice site renders a CMV-driven HIV-1 proviral clone Tat-dependent: connections between transcription and splicing. *FEBS Lett.* **563,** 113–118 (2004).
27. Fung, H. B., Kirschenbaum, H. L. & Hameed, R. Amprenavir: a new human immunodeficiency virus type 1 protease inhibitor. *Clin. Ther.* **22,** 549–572 (2000).
28. Dettenhofer, M. & Yu, X.-F. Highly purified human immunodeficiency virus type 1 reveals a virtual absence of Vif in virions. *J. Virol.* **73,** 1460–1467 (1999).
29. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152,** 36–51 (2005).
30. 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).
31. Forster, F., Medalia, O., Zauberman, N., Baumeister, W. & Fass, D. Retrovirus envelope protein complex structure in situ studied by cryo-electron tomography. *Proc. Natl Acad. Sci. USA* **102,** 4729–4734 (2005).
32. Nickell, S. *et al.* TOM software toolbox: acquisition and analysis for electron tomography. *J. Struct. Biol.* **149,** 227–234 (2005).
33. 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).
34. Pruggnaller, S., Mayr, M. & Frangakis, A. S. A visualization and segmentation toolbox for electron microscopy. *J. Struct. Biol.* **164,** 161–165 (2008).
35. Xiong, Q., Morphew, M. K., Schwartz, C. L., Hoenger, A. H. & Mastronarde, D. N. CTF determination and correction for low dose tomographic tilt series. *J. Struct. Biol.* **168,** 378–387 (2009).
36. 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).
37. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
38. Morellet, N., Druillennec, 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).
39. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26,** 1781–1802 (2005).
40. Macek, P. *et al.* NMR structure of the N-terminal domain of capsid protein from the Mason–Pfizer monkey virus. *J. Mol. Biol.* **392,** 100–114 (2009).

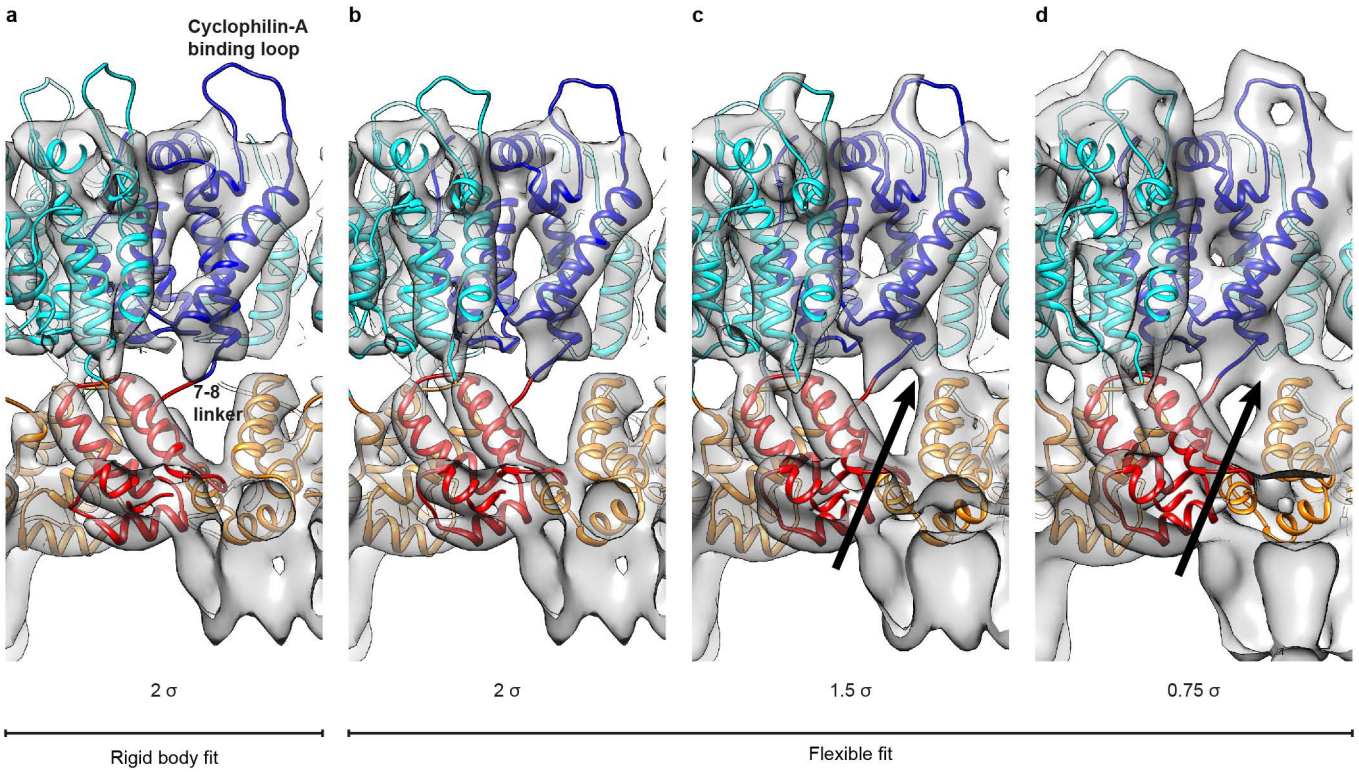
41. de Marco, A. *et al.* Role of the SP2 domain and its proteolytic cleavage in HIV-1 structural maturation and infectivity. *J. Virol.* **86,** 13708–13716 (2012).
42. Kingston, R. L. *et al.* Structure and self-association of the Rous sarcoma virus capsid protein. *Structure* **8,** 617–628 (2000).
43. Campos-Olivas, R., Newman, J. L. & Summers, M. F. Solution structure and dynamics of the Rous sarcoma virus capsid protein and comparison with capsid proteins of other retroviruses. *J. Mol. Biol.* **296,** 633–649 (2000).
44. Mortuza, G. B. *et al.* High-resolution structure of a retroviral capsid hexameric amino-terminal domain. *Nature* **431,** 481–485 (2004).
45. Cornilescu, C. C., Bouamr, F., Yao, X., Carter, C. & Tjandra, N. Structural analysis of the N-terminal domain of the human T-cell leukemia virus capsid protein. *J. Mol. Biol.* **306,** 783–797 (2001).
46. Khorasanizadeh, S., Campos-Olivas, R., Clark, C. & Summers, M. Sequence-specific 1H, 13C and 15N chemical shift assignment and secondary structure of the HTLV-I capsid protein. *J. Biomol. NMR* **14,** 199–200 (1999).



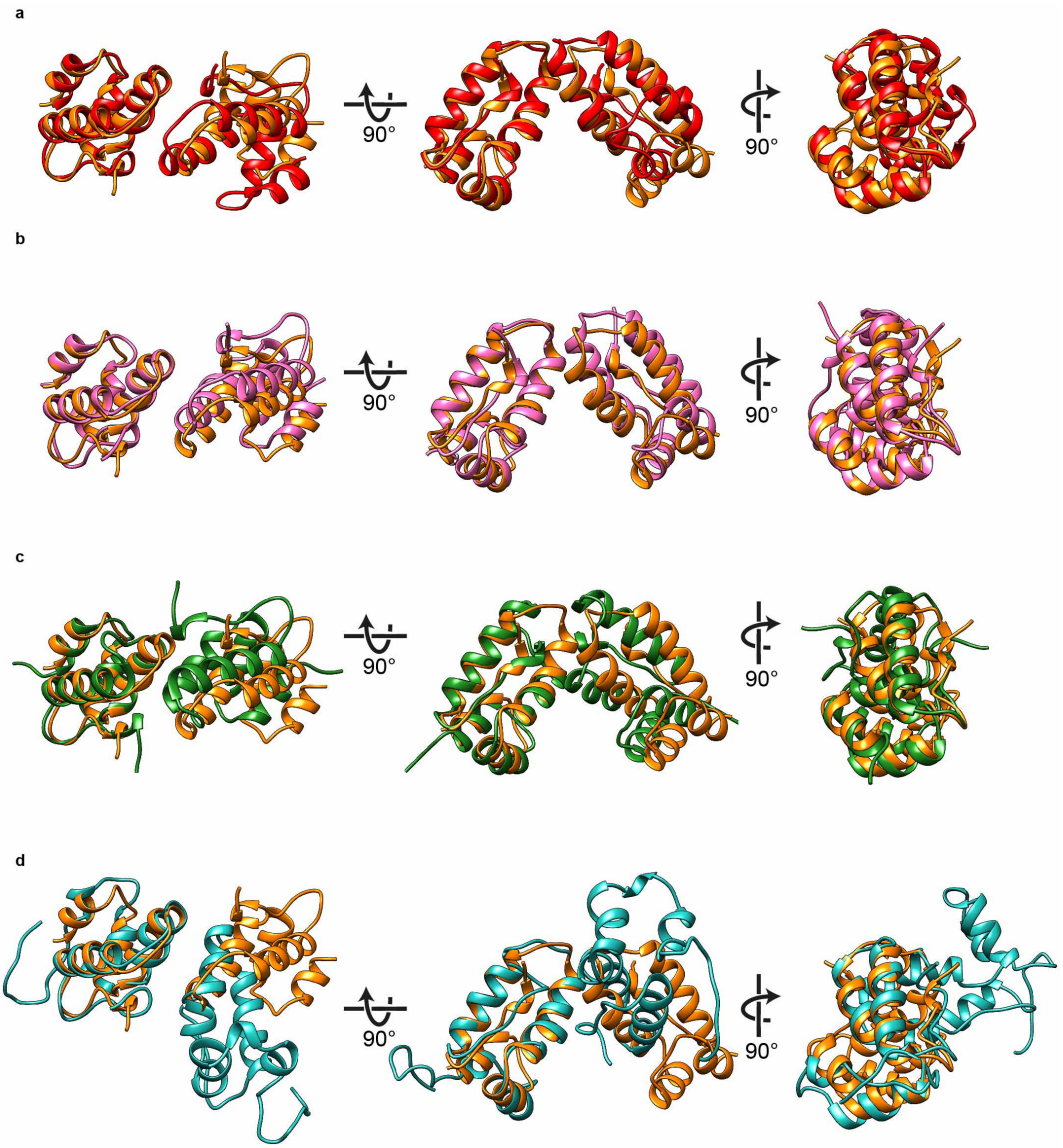
Extended Data Figure 1 | Characterization of virus preparations used in this study. Samples of Optiprep gradient (OP)-purified particles were separated by SDS–polyacrylamide gel electrophoresis. Proteins were visualized by silver staining or immunoblot, respectively, as indicated. For immunoblot analysis, proteins were transferred to nitrocellulose membranes by semi-dry blotting. Membranes were probed with polyclonal antiserum raised against recombinant HIV-1 CA (**a**) or M-PMV NC (**c**), respectively. Bound antibodies were visualized by quantitative immunodetection on a LiCor Odyssey imager, using secondary antibodies and protocols according to the manufacturer's instructions. **a**, HIV-1 particles prepared in the presence of 5 μM APV. Note that residual processing of Gag has occurred due to incomplete protease inhibition. The main additional Gag product (~50 kDa) corresponds to Gag lacking the C-terminal p6-region; we have previously shown that cleavages downstream of SP1 do not disrupt the immature Gag lattice[41]. **b**, Immature HIV-1 particles prepared from cells transfected with pNL4-3 (PR-, D25A) compared with a standard of purified Gag. Note that Gag is completely uncleaved in this case. **c**, Immature protease defective M-PMV particles purified from cells transfected with plasmid pSHRM15 (D26N). Positions of molecular mass standards (in kilodaltons) are indicated.



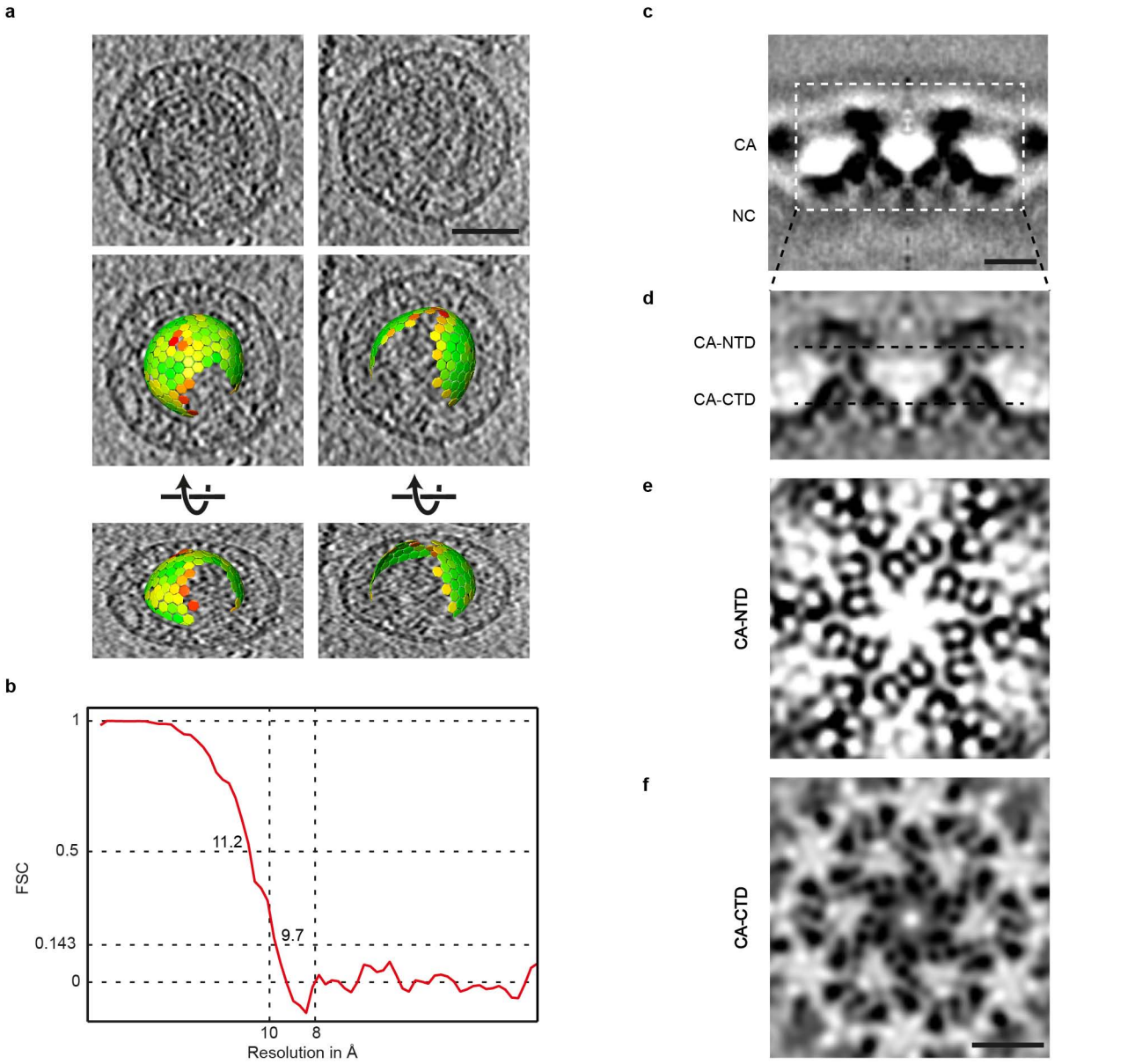
Extended Data Figure 2 | Cryo-electron tomography and subtomogram averaging reconstruction of the immature HIV-1 lattice. **a**, Within pleiotropic HIV particles the polyprotein Gag forms incomplete hexameric lattices of heterogeneous shapes. Tomographic slices of two representative viruses are shown. Underneath the membrane, ordered protein density corresponding to capsid can be seen. In the middle and lower panels, virus particles are displayed with their respective output lattice maps as derived from subtomogram averaging. The colour of the hexagons denotes the cross-correlation coefficient (CCC) of the alignment ranging from red (low CCC (0.04)) to green (high CCC (0.19)). In the lower panel, the view has been tilted to reveal regions with no lattice. Areas with no protein density underneath the membrane are devoid of a hexameric lattice. Scale bar, 50 nm. **b**, FSC between two half-data sets (each of which was independently aligned and averaged starting from independent references) of immature HIV treated with APV (blue line) showing a resolution of 9.8/8.8 Å at the 0.5/0.143 criterion, respectively. The equivalent FSC for a lower-resolution structure of the protease defective HIV (D25A) sample (red line) shows a resolution of 12.6/10.9 Å at the 0.5/0.143 criterion. **c**, Central orthoslice through the final average of the protease-inhibitor-treated HIV reconstruction (no inverse B-factor was applied). Ordered density is only observed in the CA-SP1 region. The dashed rectangle indicates the region shown in **d**. **d**, Orthoslice through the CA-SP1 region in the final average of HIV-1 + APV (left, corrected with an inverse B-factor of -1200 Å^2), the same structure filtered to 13 Å (middle) and of the protease defective HIV-1 (D25A) preparation filtered to 13 Å (right), indicating that protease-inhibited and protease mutant structures have the same domain arrangement. The dashed lines indicate the positions of horizontal orthoslices in **e** and **f**. Scale bar, 50 Å.



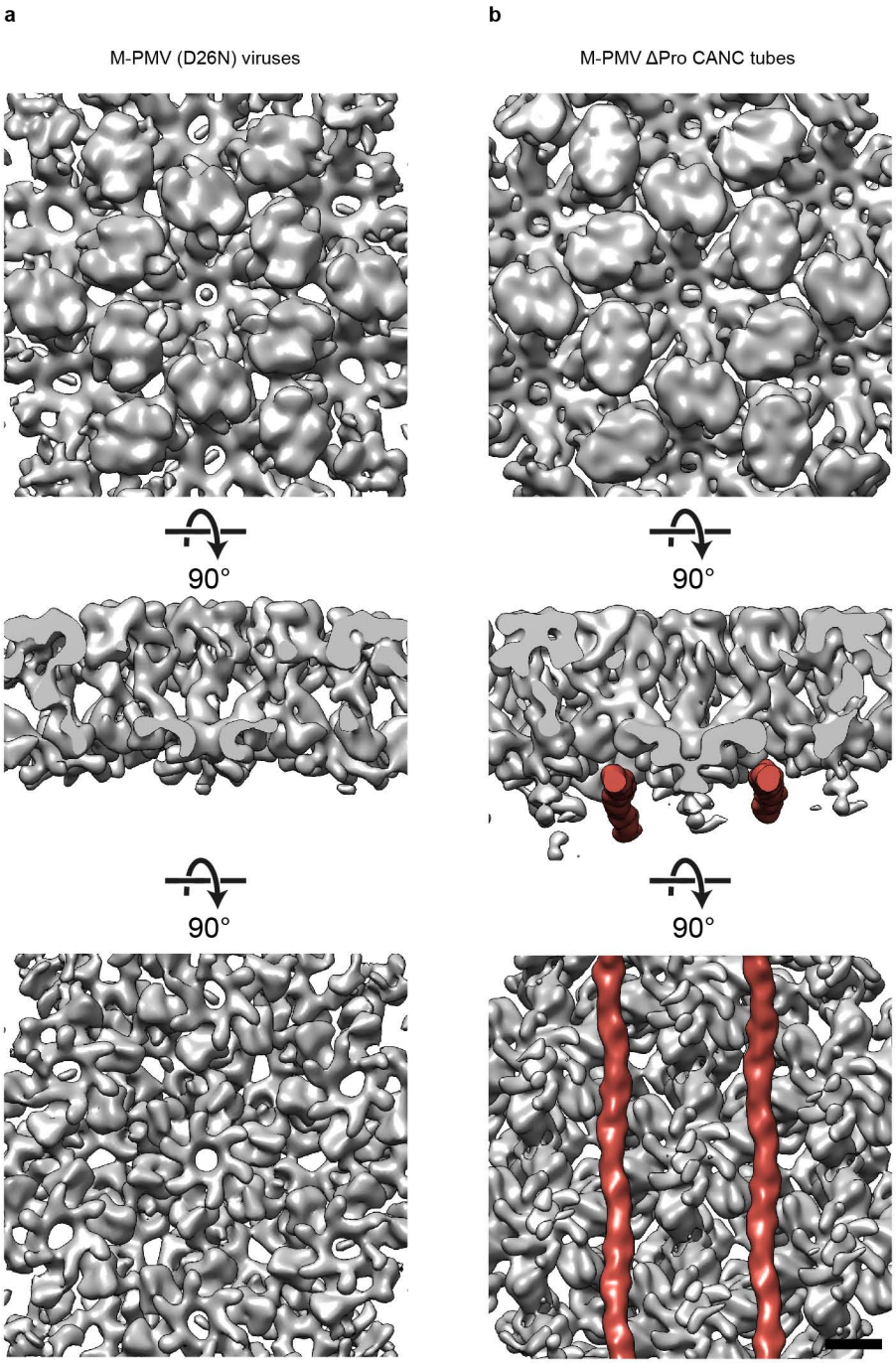
Extended Data Figure 3 | Molecular dynamics flexible fitting of high-resolution HIV-1 CA structures into the electron microscopy density. a, The rigid body fit used as a starting model for flexible fitting superimposed onto the electron microscopy map. **b**–**d**, The final flexible fit superimposed onto the electron microscopy map viewed at three different isosurface thresholds. The flexible fitting resulted in only minor movements relative to the initial rigid-body fit, implying that individual CA domains do not undergo large changes in structure between the isolated and assembled protein domains. At lower isosurface thresholds, densities corresponding to the cyclophilin-A binding loop and the 7–8 linker can be seen (positions annotated in **a**). One of the positions at which CA-NTD and CA-CTD contact one another is marked by an arrow. Isosurface values are indicated in the figure (threshold value is σ away from the mean).



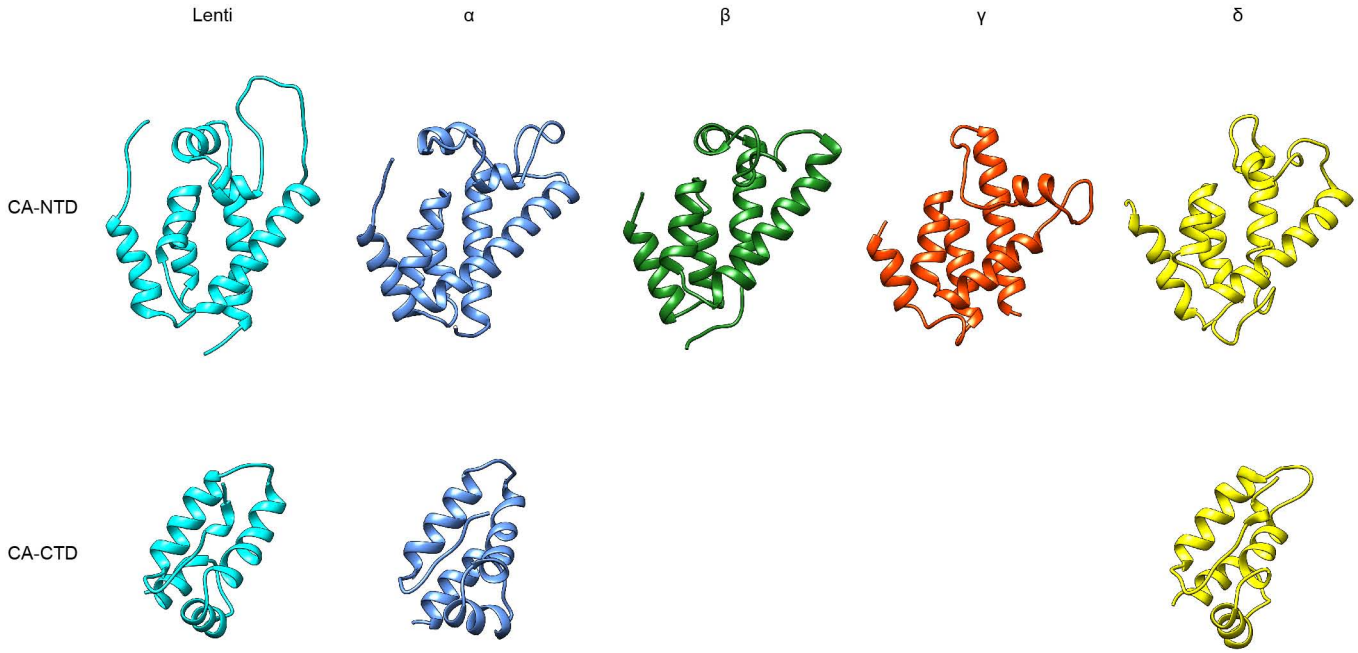
Extended Data Figure 4 | Comparison of different retroviral CA-CTD dimer structures. **a–d**, Comparison of different CA-CTD dimers aligned against the immature flexibly fitted HIV-1 CA-CTD dimer obtained in this study (orange). **a**, PDB 3DS2 (red, the crystal structure that most closely resembles the immature CA-CTD); **b**, the structure recently obtained by fitting a crystal dimer into *in vitro*-assembled HIV-1 tubes[7] (pink); **c**, the immature M-PMV dimer based on a homology model fitted into the immature M-PMV electron microscopy density generated in this study (green); **d**, a mature HIV-1 CA-CTD dimer from PDB 3J34 (ref. 6) (light blue, chain A and f). The backbone root mean squared deviations between the superimposed structures are 6.8 Å (**a**), 3.2 Å (**b**), 6.4 Å (**c**) and 17.9 Å (**d**).



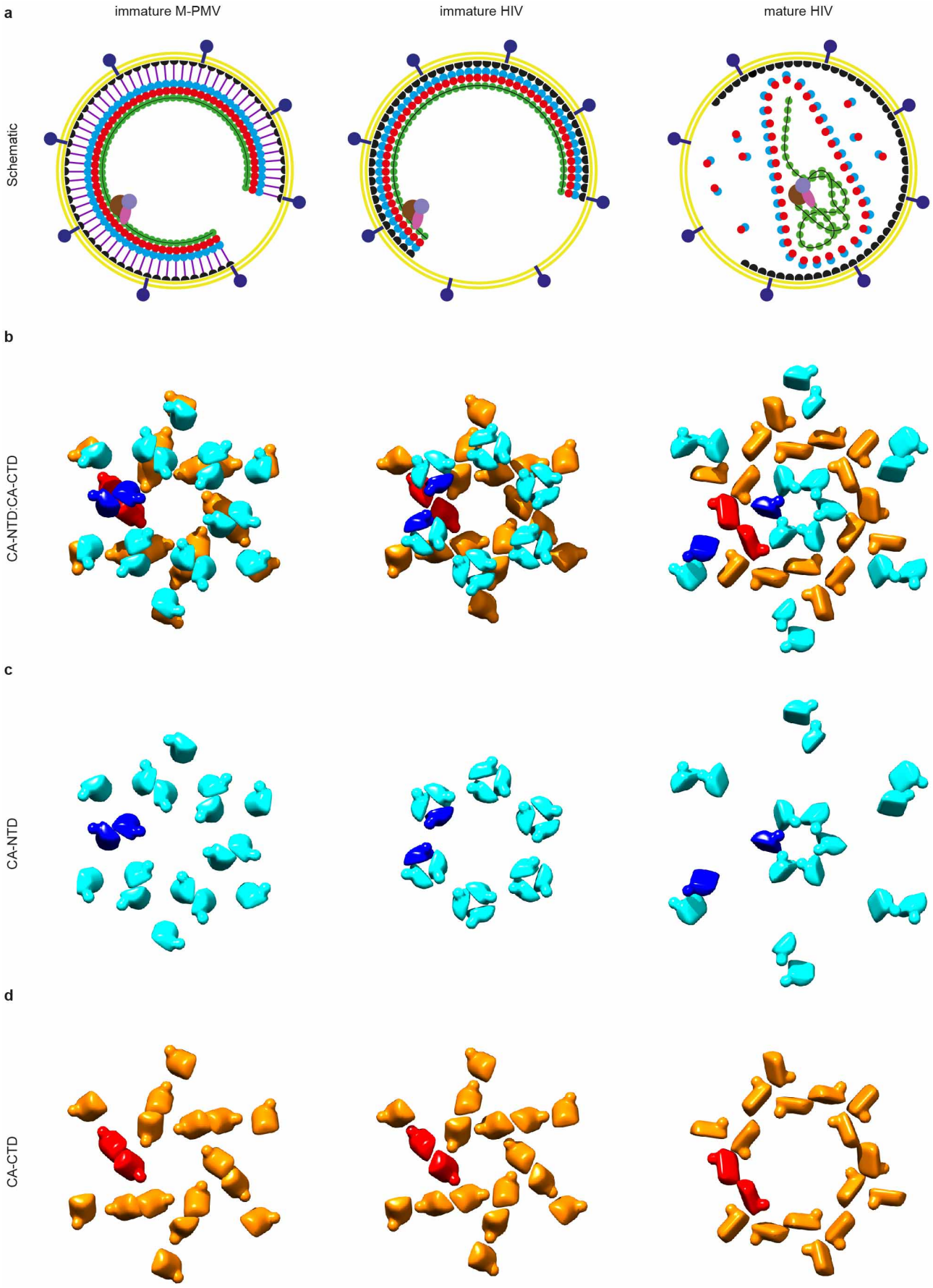
Extended Data Figure 5 | Cryo-electron tomography and subtomogram averaging reconstruction of immature M-PMV (D26N) viruses. **a**, Slices through two tomograms containing immature M-PMV (D26N) particles. Using subtomogram averaging, the position and the arrangement of the hexameric capsid lattice can be resolved (middle and lower panels). See Extended Data Fig. 2 for explanations. CCC values range from 0.05 to 0.14. Scale bar, 50 nm. **b**, FSC of the two independently aligned and averaged half-data sets of the M-PMV (D26N) sample. The resolution corresponds to 11.2 and 9.7 Å at the 0.5/0.143 criterion, respectively. **c**, Side view orthoslice through the final average of the protease defective M-PMV reconstruction (no inverse B-factor was applied). No ordered densities are observed except for the CA region. The dashed rectangle indicates the region shown in **d**. **d**, Orthoslice through the CA region in the final average corrected with a B-factor of $-1{,}000$ Å^2. The dashed lines indicate the positions of horizontal orthoslices in **e** and **f**. Scale bar, 50 Å.



Extended Data Figure 6 | Comparison of immature protease defective M-PMV (D26N) particles and M-PMV ΔPro CANC tubes. **a**, An isosurface representation of the structure generated from M-PMV (D26N) virus particles, as presented in this study. The structure is shown from the outside of the virus (top), in a horizontal slice (middle) and from the inside of the virus (bottom). **b**, For comparison, equivalent views of the structure from *in vitro*-assembled M-PMV ΔPro CANC tubular arrays (EMD2089)[2] are shown. Both structures have been filtered to 10 Å and the threshold was set to 2σ away from the mean. The additional density observed for the M-PMV ΔPro CANC tube structure (coloured red in the middle and bottom panel) is thought to be a nucleic acid structure that is ordered in the tubular arrays[2]. Scale bar, 25 Å.



Extended Data Figure 7 | Structural conservation of retroviral CA domains. Known atomic structures of CA-NTD and CA-CTD domains from the lentivirus[13,14], α[42,43], β[40], γ[44] and δ[45,46] retrovirus families. The tertiary structure of CA domains is highly conserved between different retroviruses.



Extended Data Figure 8 | Comparison of CA domain arrangements in immature M-PMV, immature HIV and mature HIV. **a**, Two-dimensional schematic representations of immature M-PMV, immature HIV and mature HIV particles. Retroviruses bud from infected cells in an immature form. The Gag polyprotein is radially arranged underneath the host-derived plasma membrane (yellow) and includes the membrane associated matrix domain (MA, black), the bipartite capsid domain (CA, blue and red) and the nucleic-acid-bound nucleocapsid (NC, green). M-PMV possesses two additional domains positioned between MA and CA, termed pp24 and p12 (purple), leading to a bigger spatial separation of CA and MA compared with HIV. The viral protease cleaves the Gag polyprotein at defined positions, triggering maturation. This process leads to a rearrangement of the domains, giving rise to the mature, infectious virus. **b**–**d**, Schematic diagram representing the arrangement of the CA-NTD and CA-CTD domains within immature M-PMV and HIV and within mature HIV. CA-NTD and CA-CTD molecules are represented by cyan/blue- and orange/red-coloured solids, respectively. The solids are positioned at the exact positions and orientations at which the high-resolution structures fitted into the electron microscopy densities. For CA-NTD the N terminus and for CA-CTD the C terminus are represented as spherical extensions. The shown schematics were generated in UCSF Chimera by defining the translational and rotational matrix of each fitted CA domain, and applying it to the solid representations. **b**, Both CA-NTD (cyan/blue) and CA-CTD (orange/red) domains are shown; **c**, CA-NTD only; **d**, CA-CTD only.

Extended Data Table 1 | Data processing statistics

	Immature protease inhibitor treated (APV) HIV viruses	Immature protease defective HIV (D25A) viruses	Immature protease defective M-PMV (D26N) viruses
Microscope	FEI Titan Krios	FEI Titan Krios	FEI Titan Krios
Voltage (keV)	200	300	200
Detector	Gatan multiscan 795 CCD camera	FEI Falcon II Direct detector	Gatan multiscan 795 CCD camera
Energy-filter	GATAN GIF2002	None	GATAN GIF2002
Å/pixel	2.025	2.281	2.025
Defocus range (microns)	-1.2 to -4.0	-2.0 to -5.5	-1.5 to -4.5
Total Dose (electrons/Å^2)	40	60	40
Tomograms reconstructed	70	33	39
Viruses extracted	224	116	84
Asymmetric units Set A	94,986	33,762	34,818
Asymmetric units Set B	99,744	29,352	42,702
Final resolution (0.5/0.143 FSC) in Å	9.8/8.8	12.6/10.9	11.2/9.7

Summary of image acquisition and processing statistics for the data sets discussed in this study. Note that each reconstruction was performed on two independent half-data sets.