

The stoichiometry of Gag protein in HIV-1

John A G Briggs¹, Martha N Simon², Ingolf Gross³, Hans-Georg Kräusslich⁴, Stephen D Fuller¹, Volker M Vogt³ & Marc C Johnson³

The major structural components of HIV-1 are encoded as a single polyprotein, Gag, which is sufficient for virus particle assembly. Initially, Gag forms an approximately spherical shell underlying the membrane of the immature particle. After proteolytic maturation of Gag, the capsid (CA) domain of Gag reforms into a conical shell enclosing the RNA genome. This mature shell contains 1,000–1,500 CA proteins assembled into a hexameric lattice with a spacing of 10 nm. By contrast, little is known about the structure of the immature virus. We used cryo-EM and scanning transmission EM to determine that an average (145 nm diameter) complete immature HIV particle contains ~5,000 structural (Gag) proteins, more than twice the number from previous estimates. In the immature virus, Gag forms a hexameric lattice with a spacing of 8.0 nm. Thus, less than half of the CA proteins form the mature core.

The immature HIV-1 Gag shell assembles with the N-terminal matrix domains interacting directly with the inner leaflet of the bilayer, and with the C-terminal ends pointing toward the particle center in a radial arrangement¹ (Fig. 1). When visualized by cryo-EM, the shell shows characteristic striations (Fig. 1b). Once active, the viral protease cleaves Gag into its constitutive components: matrix protein (MA), capsid protein (CA), nucleocapsid protein (NC) and p6 protein. After cleavage, CA reassembles to form the cone-shaped viral core, which contains the genomic RNA, NC, reverse transcriptase and other viral components (Fig. 1c). This maturation process is necessary for production of infectious virions. Interference with this process is the basis of some of the commonly used HIV therapies.

HIV-1 assembly has been widely studied both *in vitro* and *in vivo*. The CA and CA-NC portions of Gag can assemble as purified proteins *in vitro* into helical tubes. These are formed from hexameric rings ~9–10 nm apart^{2,3} that are assembled into a lattice with local p6 symmetry. These arrays are thought to mimic the arrangement of CA in the core of the mature virus, which is a cone- or tube-shaped structure consisting of ~1,000–1,500 Gag proteins⁴. Negatively stained two-dimensional crystals of recombinant His-tagged CA constructs adopt hexagonal p6 lattices with either 9-nm or 6.4-nm spacings, depending on the length of the His-tag⁵. It has been suggested that the 9-nm lattice corresponds to that seen in the helical tubes and that the 6.4-nm spacing may correspond to that in the immature virion⁵. Negatively stained virus-like particles produced by Gag expression in insect cells also show local areas composed of a 6.4-nm p6 hexagonal lattice⁶. In both two-dimensional crystals and virus-like particles, it is likely that substantial shrinkage of the lattice will result from a combination of negative staining and a reduction

in lattice curvature. There have been no direct observations of the Gag lattice spacing in the immature virion.

HIV virions have been typically reported to contain between 1,200 and 2,500 copies of the Gag protein^{7–10}. These estimates of stoichiometry are derived from four primary sources. The first is the empirical determination of the ratio of viral genomic RNA to CA, which corresponds to ~10⁴ virions pg⁻¹ CA antigen¹¹. This ratio yields an order-of-magnitude estimate of 2,500 CA proteins per virion. The second source is the comparison of the CA concentration in a sample with the number of particles counted by thin-section EM together with a known standard. This approach yields an estimate of 1,200 ± 700 CA proteins per virion¹², the accuracy of this number being limited by the inaccuracy in counting the virus particles. The third source is a comparison of the concentrations of CA and Env proteins in purified virus preparations coupled with a count of Env trimers per particle by negative-stain EM. The resulting estimate of 1,400 CA proteins per virion depends on the accurate quantification of the Env protein, which is difficult, as well as on the assumption that every Env protein formed a trimer that could be visualized by EM¹³. The fourth source is the recently reported stoichiometry of Rous sarcoma virus (RSV) Gag, together with the assumption that HIV stoichiometry is similar¹⁴.

Knowing the number of Gag proteins per virion is important for two reasons. First, the density and spacing of Gag in the immature particle underpin an understanding of the structure of the immature particle and of the changes that occur during maturation. Second, the Gag protein is the major component of the virion, making up approximately half of the total mass, and acts as the standard relative to which the numbers of other viral components are expressed. Knowing the stoichiometries of a wide range of viral

¹Division of Structural Biology, The Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Headington, Oxford OX3 7BN, UK. ²Department of Biology, Brookhaven National Laboratory, Upton, New York 11973, USA. ³Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA. ⁴Department of Virology, Hygiene-Institut, Im Neuenheimer Feld 324, D-69120 Heidelberg, Germany. Correspondence should be addressed to M.C.J. (mcj7@cornell.edu).

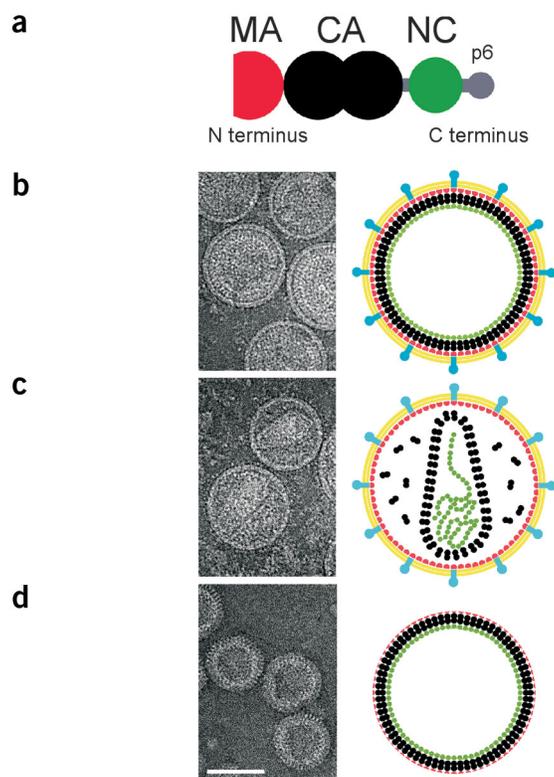


Figure 1 Cryo-electron micrographs and schematic representations of HIV-1 particles and HIV-1 Gag protein. (a) Schematic representation of Gag protein. Red, MA domain of Gag protein; black, CA domains of Gag; green, NC domain of Gag (bound to RNA); gray, linker peptides and p6 protein. (b) Immature HIV-1 virions. Blue, viral envelope proteins; yellow, lipid bilayer. (c) Mature HIV-1 virions. (d) *In vitro*-assembled HIV-1 Gag particles. Scale bar, 100 nm.

HIV-1 Gag protein-derived lattices show a p6 lattice containing six proteins per unit cell. This is the case *in vitro* for two-dimensional⁵ and tubular arrays² and *in vivo* for the mature HIV-1 core⁴. Furthermore, dimerization is known to be important in Gag assembly^{16–18}, and the C-terminal domain of CA dimerizes with a K_d of 10–20 μM (refs. 19,20). Dimeric interactions are not accommodated in p3 lattices containing only three proteins in the unit cell. It is therefore most probable that the unit cell contains six Gag proteins.

HIV-1 Gag-derived proteins can form spherical structures *in vitro* in the presence of nucleic acid^{21–23}. Apart from the absence of a membrane, these particles closely resemble immature HIV-1 and have a radial domain organization that is indistinguishable from that of the immature virion^{1,22}, suggesting that the Gag protein is assembled in the same manner in both particles. The *in vitro*-assembled particles are more regular in size than immature virions and, owing to the absence of a membrane, the Gag lattice is more clearly visible. *In vitro*-assembled HIV Gag particles were prepared as described^{21,22}, and imaged (Fig. 1d). The power spectra of the centers of images of 173 *in vitro*-assembled particles taken from three independent preparations were rotationally aligned and averaged. The averaged power spectrum again revealed a six-fold symmetrical pattern (Fig. 2a, right) with the (1,0) reflection at $6.9 \pm 0.2 \text{ nm}^{-1}$, equivalent to a unit cell center-to-center distance of $8.0 \pm 0.2 \text{ nm}$. The Gag protein therefore adopts the same packing arrangement in the *in vitro*-assembled particles as in the immature virion. The unit cell is substantially smaller than the 9.6-nm unit cell seen in the CA lattice of the mature viral core⁴.

The number of unit cells making up a particle can be calculated if the surface area of the lattice is known. We therefore carried out

components is therefore directly dependent on knowing the stoichiometry of Gag.

We have taken two approaches to measuring the number of Gag proteins in an HIV virion. We have observed the Gag protein lattice using cryo-EM and have measured the masses of individual particles using scanning transmission EM (STEM).

RESULTS

Cryo-EM

Immature HIV-1 virions were prepared from infected MT-4 cells treated with an inhibitor of the viral protease and imaged by cryo-EM as described^{1,15} (Fig. 1b). The power spectra of images of the centers of 145 immature virions, taken from two independent preparations, were iteratively rotationally aligned and averaged. The averaged power spectrum revealed a six-fold symmetrical pattern of peaks (Fig. 2a, left) consistent with Gag protein packing to form a p3 or p6 lattice. The peaks corresponding to the (1,0) reflection are at a spacing of $6.9 \pm 0.2 \text{ nm}^{-1}$, equivalent to a unit cell spacing of $8.0 \pm 0.2 \text{ nm}$.

Both p3 and p6 unit cells must contain a number of Gag proteins that is a multiple of three. As discussed above, previous observations of

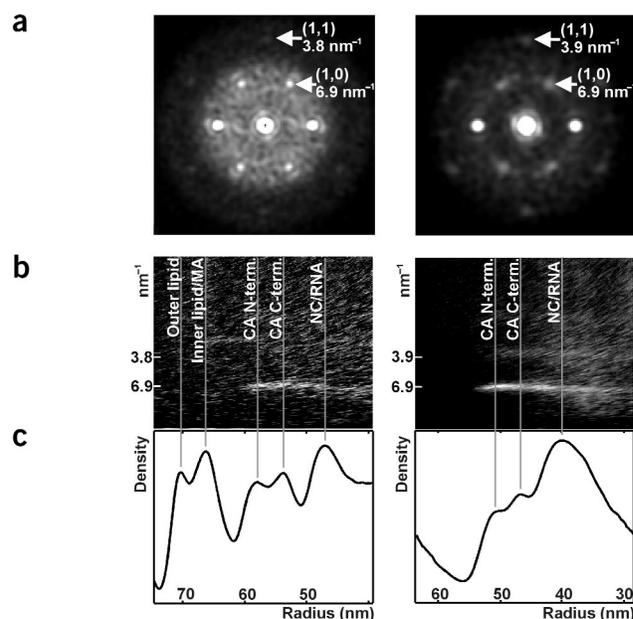


Figure 2 Image processing of cryo-electron micrographs of viral particles. Immature HIV-1 (left panels) and *in vitro*-assembled HIV-1 Gag particles (right panels) were analyzed. (a) Averaged aligned power spectra of the central regions of particles show a hexagonal pattern of peaks. Example (1,0) and (1,1) reflections are labeled. (b) Rings running parallel to the particle edge were extracted at each radius. 1D Fourier transforms along each ring (vertical columns) plotted against radius reveal from where the lattice reflections arise. Peaks are smeared horizontally to the right because the particle is viewed in projection. (c) The average mean density of each ring plotted against radius. Peaks represent structural features as indicated¹.

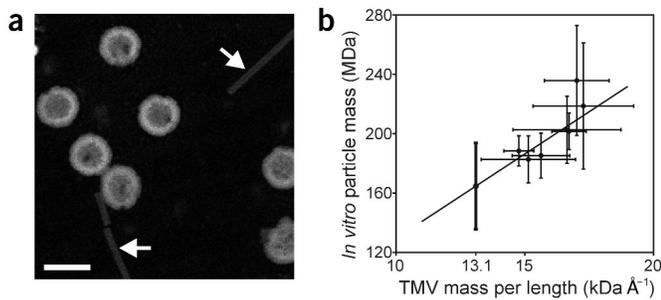


Figure 3 STEM of *in vitro*-assembled HIV-1 Gag protein. (a) Micrograph of *in vitro*-assembled particles mixed with tobacco mosaic virus rods (arrows). Scale bar, 100 nm. (b) Mass per unit length of TMV rods plotted against mass of *in vitro*-assembled particles for each grid. The best first-order fit is indicated, as well as the corrected mass of the *in vitro* particles identified from the value at the known TMV mass per unit length of 13.1 kDa Å⁻¹ (bold error bar).

further image processing to determine the radius within the particle at which the diffracting lattice is found. The characteristic striations visible at the edges of particles result from a side-on projection view of the Gag lattice. Image processing using tangential Fourier transforms revealed the relationship between the reflections from the Gag lattice and radial position in the particle (Fig. 2b,c). In both the *in vitro* particle and in the immature virion, the peak corresponding to the (1,0) reflection is found at the same radius as the density previously inferred to be the outer (N-terminal) domain of CA¹. In the *in vitro* particles the 6.9 nm⁻¹ reflection is strongest at a radius of 50 ± 2 nm. The surface area of this layer is 31,000 ± 2,000 nm², and it would contain 570 ± 40 copies of the 8.0-nm unit cell. If the unit cell contains the expected six proteins, this corresponds to 3,400 ± 200 copies of Gag per *in vitro*-assembled particle.

Scanning transmission EM

To confirm this stoichiometry, we used STEM to measure the number of Gag proteins in *in vitro*-assembled HIV Gag particles. This technique permits the measurement of the masses of individual freeze-dried particles^{14,24}. Four independent preparations of particles were mixed with the rod-shaped tobacco mosaic virus (TMV) as an internal standard^{14,24} and freeze-dried onto multiple EM grids for STEM measurement (Fig. 3a). Mean masses and standard deviations for each grid were measured to produce a plot of mean TMV mass per unit length versus mean particle mass (Fig. 3b). The best first-order fit was calculated, taking into account the errors in both variables, from which the value and error for particle mass at the known TMV mass per unit length could be read. The mass distribution for the *in vitro* particles was 165 ± 29 MDa. As the HIV Gag protein used for assembly has a mass of 41 kDa, and 5–10 % of the particle mass is contributed by nucleic acid, the *in vitro*-assembled particles contain 3,800 ± 700 molecules of Gag. This is fully consistent with our conclusions from direct observation of the lattice, and confirms that there are six proteins in the unit cell.

DISCUSSION

HIV-1 virions have a broad range of diameters. The number of unit cells is proportional to surface area and varies markedly among virions of different sizes. In the immature virion, the N-terminal domain of CA is found 13 nm from the outside of the particle. The 8.0-nm lattice is found at a radius of ~60 nm in an average immature 145-nm

particle⁴. A complete spherical lattice of this size would contain ~820 unit cells, corresponding to ~4,900 copies of Gag. The calculated stoichiometry for HIV is three times that of RSV as determined by STEM¹⁴, indicating considerable variation among retroviruses. About 95% of HIV-1 virions have diameters between 119 nm and 207 nm, corresponding to 3,000 and 11,000 Gag proteins, respectively. Revision of the number Gag proteins per virion directly affects the estimated numbers of minor virion components, as the stoichiometries of HIV accessory proteins and cellular components have been estimated by comparison with CA. Thus, assuming it has a completely closed spherical lattice, an average 145-nm immature HIV-1 virion would contain ~700 molecules of Vpr (1:7 compared with Gag²⁵), 500 molecules of cyclophilin A (1:10 compared with Gag²⁶) and 250 molecules of reverse transcriptase and integrase (1:20 compared with Gag).

The early view of retroviral maturation held that after cleavage of Gag by the viral protease, the liberated CA condensed to form the mature viral core (reviewed in refs. 9,27). Here we show that the unit cell of the CA domain of the Gag lattice is smaller in the immature virion than the unit cell of the CA lattice in the mature core, and by implication the CA domain forms a more densely packed lattice before maturation than after maturation. Because the mature HIV-1 core is formed from ~1,000–1,500 molecules of CA^{2,4}, only a fraction of the Gag molecules gives rise to CA molecules that are used for core formation. This confirms that core formation requires a second assembly step, and is not simply a condensation of the CA portion of the Gag lattice. Furthermore, a consequence of maturation is the creation of a large pool of free CA in the virion. These conclusions are consistent with other observations on HIV-1. First, about one-third of mature HIV virions are known to contain two cores⁴. Double-cored particles (average diameter of 159 nm, which would correspond to 6,200 Gag proteins) are larger than single-cored particles (134 nm; 4,100 Gag proteins) and concentration estimates suggest that virions may contain more CA than is required for core formation⁴. Second, it has been noted that NC is substantially enriched relative to CA in preparations of purified viral cores^{28,29}. Third, Prevelige and co-workers have provided evidence (this issue) for two states of the CA protein in mature HIV-1 virions using mass spectrometry³⁰.

In summary, our data show that the average immature HIV particle contains ~5,000 copies of Gag, and that less than one-third of the total CA from Gag contributes to the mature viral core. Furthermore, CA is more densely packed as part of the Gag lattice before maturation than as the CA lattice in the mature core shell. The nature of the difference in protein-protein contacts in the two lattices remains to be elucidated.

METHODS

Image processing of particle centers. Cryo-EM was conducted as described^{1,31} using a Philips CM200FEG operated at 200 kV at a magnification of 38,000×. Micrographs were digitized on a UMAX 3000 scanner at 8.3-μm step size to give a pixel size of 0.22 nm. Image analysis was carried out using IMAGIC (Image Science Software). Soft-edged (5 pixel drop-off) circles of 250 pixels (55 nm) in diameter were excised from the centers of particle images and padded into 1,024² boxes before Fourier transformation and conversion to powers. The rotationally averaged sum of all power spectra exhibited a maximum at 32 transform units from the center for immature virus and at 33 transform units for *in vitro*-assembled particles. The s.d. in the position of this maximum between independent preparations was <1 transform unit, and therefore data from separate preparations were combined.

All power spectra were rotationally aligned against a single white pixel at the radius of the maximum in the rotationally averaged sum of all power spectra. The aligned power spectra were further aligned by four iterations of summation and rotational alignment against the sum. The sum of the aligned power

spectra showed a hexagonal pattern of peaks. The brightest reflections were at the position of the pixel used as a template for the first round of alignment. The s.d. in the positions of the three independent peaks in each case is <0.5 transform units.

Image processing of edges of particles. Analysis of the edge striations in particles assembled *in vitro* was carried out after centering the images by iterative alignment against a rotationally averaged sum of all particles, and discarding poorly aligned particles (13%). Images were low-pass-filtered to 0.88 nm. Centered rings of incrementally decreasing radius were extracted and linearized using MATLAB (Mathworks). The mean density along each ring was calculated and plotted against radius to generate a density profile. Data from the particles were radially aligned using these profiles. Immature HIV-1 virions were processed similarly, but rings for linearization were calculated based on a trace of the edge of the particle. Twenty immature virions were analyzed. Peaks in the averaged density profile were assigned based on previously published deletion analysis¹. To identify the radius within the particle from which lattice reflections arise, linearized rings were 1D Fourier transformed and arranged to produce a plot with radial distance along the *x*-axis and the power of the 1D Fourier transform along the *y*-axis. The radius of the lattice is that at which the power of the 6.9-nm reflection is greatest. The s.d. of the shifts required to align the radial density profiles provides an estimate of error of the radial position of the lattice.

STEM. STEM was carried out essentially as described¹⁴. The instrument was calibrated to give an absolute mass for individual particles, but variation from experiment to experiment can occur. A sample of known mass, TMV, is therefore included as a control. Mean particle mass was plotted against mean TMV mass per unit length for each grid analyzed. Where the masses of <10 Gag particles or <10 TMV rods could be measured from a single grid, the results from that grid were discarded. The best first-order fit was calculated using the Numerical Recipes *fitexy* program³². The fit takes into account the s.e.m. in both variables. The corrected mass of the *in vitro* particles was identified from the value at the known TMV mass per unit length of 13.1 kDa Å⁻¹. To obtain an estimate of the s.e.m. using *fitexy*, the s.e.m. values for each point were scaled by multiplication by the square root of the mean number of data points before fitting. Preparations of *in vitro*-assembled RSV Gag particles were analyzed in parallel and gave masses consistent with published experiments³³ (data not shown).

ACKNOWLEDGMENTS

We thank D.I. Stuart, E.Y. Jones, R. Matadeen, R.L. Kingston and R.J. Hurrellbrink for critical readings of the manuscript, D.I. Stuart and E.Y. Jones for discussions of diffraction analysis and K.V. Fernando for help with programming. The BNL STEM is a US National Institutes of Health (NIH) supported resource center, NIH P41-RR01777, with additional support provided by the US Department of Energy, Office of Biological and Environmental Research. This work was supported by a Wellcome Trust Programme Grant to S.D.F., a US Public Health Service grant to V.M.V. and a Deutsche Forschungsgemeinschaft grant to H.-G.K. J.A.G.B. holds a Wellcome Trust structural biology studentship. S.D.F. is a Wellcome Trust principal research fellow.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 29 January; accepted 19 April 2004

Published online at <http://www.nature.com/nsmb/>

1. Wilk, T. *et al.* Organization of immature human immunodeficiency virus type 1. *J. Virol.* **75**, 759–771 (2001).
2. Li, S., Hill, C.P., Sundquist, W.I. & Finch, J.T. Image reconstructions of helical assemblies of the HIV-1 CA protein. *Nature* **407**, 409–413 (2000).
3. Ganser, B.K., Li, S., Klishko, V.Y., Finch, J.T. & Sundquist, W.I. Assembly and analysis of conical models for the HIV-1 core. *Science* **283**, 80–83 (1999).

4. Briggs, J.A.G., Wilk, T., Welker, R., Krausslich, H.G. & Fuller, S.D. Structural organization of authentic, mature HIV-1 virions and cores. *EMBO J.* **22**, 1707–1715 (2003).
5. Mayo, K. *et al.* Retrovirus capsid protein assembly arrangements. *J. Mol. Biol.* **325**, 225–237 (2003).
6. Nermut, M.V. *et al.* Further evidence for hexagonal organization of HIV gag protein in prebudding assemblies and immature virus like particles. *J. Struct. Biol.* **123**, 143–149 (1998).
7. Cimarelli, A. & Darlix, J.L. Assembling the human immunodeficiency virus type 1. *Cell Mol. Life Sci.* **59**, 1166–1184 (2002).
8. Frankel, A.D. & Young, J.A. HIV-1: fifteen proteins and an RNA. *Annu. Rev. Biochem.* **67**, 1–25 (1998).
9. Wilk, T. & Fuller, S.D. Towards the structure of the human immunodeficiency virus: divide and conquer. *Curr. Opin. Struct. Biol.* **9**, 231–243 (1999).
10. Turner, B.G. & Summers, M.F. Structural biology of HIV. *J. Mol. Biol.* **285**, 1–32 (1999).
11. Piatak, M. *et al.* High-levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. *Science* **259**, 1749–1754 (1993).
12. Layne, S.P. *et al.* Factors underlying spontaneous inactivation and susceptibility to neutralization of human-immunodeficiency-virus. *Virology* **189**, 695–714 (1992).
13. Zhu, P. *et al.* Electron tomography analysis of envelope glycoprotein trimers on HIV and simian immunodeficiency virus virions. *Proc. Natl. Acad. Sci. USA* **100**, 15812–15817 (2003).
14. Vogt, V.M. & Simon, M.N. Mass determination of rous sarcoma virus virions by scanning transmission electron microscopy. *J. Virol.* **73**, 7050–7055 (1999).
15. Wilk, T., Gowen, B.E. & Fuller, S.D. Actin associates with the nucleocapsid domain of the Gag polyprotein in the human immunodeficiency virus (HIV-1). *J. Virol.* **73**, 1931–1940 (1999).
16. Accola, M.A., Strack, B. & Gottlinger, H.G. Efficient particle production by minimal gag constructs which retain the carboxy-terminal domain of human immunodeficiency virus type 1 capsid-p2 and a late assembly domain. *J. Virol.* **74**, 5395–5402 (2000).
17. Johnson, M.C., Scobie, H.M., Ma, Y.M. & Vogt, V.M. Nucleic acid-independent retrovirus assembly can be driven by dimerization. *J. Virol.* **76**, 11177–11185 (2002).
18. Zhang, Y.Q., Qian, H.Y., Love, Z. & Barklis, E. Analysis of the assembly function of the human immunodeficiency virus type 1 gag protein nucleocapsid domain. *J. Virol.* **72**, 1782–1789 (1998).
19. Gamble, T.R. *et al.* Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science* **278**, 849–853 (1997).
20. del Alamo, M., Neira, J.L. & Mateu, M.G. Thermodynamic dissection of a low affinity protein-protein interface involved in human immunodeficiency virus assembly. *J. Biol. Chem.* **278**, 27923–27929 (2003).
21. 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).
22. Gross, I. *et al.* A conformational switch controlling HIV-1 morphogenesis. *EMBO J.* **19**, 103–113 (2000).
23. Campbell, S. & Rein, A. *In vitro* assembly properties of human immunodeficiency virus type 1 Gag protein lacking the p6 domain. *J. Virol.* **73**, 2270–2279 (1999).
24. Wall, J.S., Hainfeld, J.F. & Simon, M.N. Scanning transmission electron microscopy of nuclear structures. *Methods Cell Biol.* **53**, 139–164 (1998).
25. Muller, B., Tessmer, U., Schubert, U. & Krausslich, H.G. Human immunodeficiency virus type 1 Vpr protein is incorporated into the virion in significantly smaller amounts than Gag and is phosphorylated in infected cells. *J. Virol.* **74**, 9727–9731 (2000).
26. Franke, E.K., Yuan, H.E.H. & Luban, J. Specific incorporation of cyclophilin-A into HIV-1 virions. *Nature* **372**, 359–362 (1994).
27. Freed, E.O. HIV-1 Gag protein: diverse functions in the virus life cycle. *Virology* **251**, 1–15 (1998).
28. Wilk, T., Hohenberg, H., Tessmer, U., Huckhagel, C. & Krausslich, H.G. Biochemical and structural analysis of isolated mature cores of human immunodeficiency virus type 1. *J. Virol.* **74**, 1168–1177 (2000).
29. Forshey, B.M. & Aiken, C. Disassembly of human immunodeficiency virus type 1 cores *in vitro* reveals association of Nef with the subviral ribonucleoprotein complex. *J. Virol.* **77**, 4409–4414 (2003).
30. Lanman, J. *et al.* Key interactions in HIV-1 maturation identified by mass spectrometry based H/D exchange. *Nat. Struct. Mol. Biol.* **11**, 676–677 (2004).
31. Fuller, S.D., Wilk, T., Gowen, B.E., Krausslich, H.G. & Vogt, V.M. Cryo-electron microscopy reveals ordered domains in the immature HIV-1 particle. *Curr. Biol.* **7**, 729–738 (1997).
32. Press, W.H., Flannery, B.P., Teukolsky, S.A. & Vetterling, W.T. *Numerical Recipes in Fortran 77* (Cambridge Univ. Press, Cambridge, UK, 1992).
33. Yu, F. *et al.* Characterization of Rous sarcoma virus Gag particles assembled *in vitro*. *J. Virol.* **75**, 2753–2764 (2001).