The Mimivirus 1.2 Mb dsDNA genome is elegantly organized into a nuclear-like weapon

Chantal Abergel (✉ chantal.abergel@igs.cnrs-mrs.fr)

French National Centre for Scientific Research <https://orcid.org/0000-0003-1875-4049>

Alejandro Villalta Casares

French National Centre for Scientific Research <https://orcid.org/0000-0002-7857-7067>

Emmanuelle Quemin

University of Hamburg

Alain Schmitt

French National Centre for Scientific Research

Jean-Marie Alempic

French National Centre for Scientific Research

Audrey Lartigue

French National Centre for Scientific Research

Vojta Prazak

University of Oxford

Daven Vasishtan

Oxford

Agathe Colmant

French National Centre for Scientific Research

Flora Honore

French National Centre for Scientific Research <https://orcid.org/0000-0002-0390-8730>

Yohann Coute

University Grenoble Alpes, CEA <https://orcid.org/0000-0003-3896-6196>

Kay Gruenewald

University of Oxford <https://orcid.org/0000-0002-4788-2691>

Lucid Belmudes

Univ. Grenoble Alpes, CEA, INSERM, IRIG, BGE

Biological Sciences - Article

Keywords: Mimivirus 1.2 Mb dsDNA, viral genome, organization, RNA polymerase subunits

Posted Date: February 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-83682/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Mimivirus 1.2 Mb genome is elegantly organized into a nuclear-like weapon

Alejandro Villalta^{a#}, Emmanuelle R. J. Quemin^{b#}, Alain Schmitt^{a#}, Jean-Marie Alempic^a, Audrey Lartigue^a, Vojtěch Pražák^c, Lucid Belmudes^d, Daven Vasishtan^c, Agathe M. G. Colmant^a, Flora A. Honoré^a, Yohann Couté^d, Kay Grünewald^{b,c}, Chantal Abergel^{a*}

^aAix–Marseille University, Centre National de la Recherche Scientifique, Information Génomique & Structurale, Unité Mixte de Recherche 7256 (Institut de Microbiologie de la Méditerranée, FR3479), 13288 Marseille Cedex 9, France.

^bCentre for Structural Systems Biology, Heinrich-Pette-Institute, Leibniz Institute for Experimental Virology, University of Hamburg, D-22607 Hamburg, Germany

^cDivision of Structural Biology, Wellcome Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, United Kingdom

^dUniv. Grenoble Alpes, CEA, INSERM, IRIG, BGE, 38000 Grenoble, France

[#]These authors contributed equally to the work

Summary

Mimivirus is the prototype of the Mimiviridae family of giant dsDNA viruses, initially isolated in *Acanthamoeba*¹. Little is known about the organization of the viral genome inside the membrane limited nucleoid² and whether unpacking or other rearrangements are required prior to transcription and replication. Here we show that opening of its large icosahedral capsid *in vitro* leads to the release of electron dense, 30 nm diameter rod-shaped objects that appear to be expelled from the particles and unwinding. We developed a purification procedure and characterized the detailed structure at various stages of decompaction using cryo-electron microscopy single particle analyses and its composition by proteomics. This revealed that the viral genome is encased into a helical protein shell surprisingly made of the two GMC-type oxydoreductases that are also the major components of the glycosylated fibrils surrounding the capsid³. The 1.2 Mb genome is folded to follow a 5- or 6-start left-handed helix, depending on the nature of the GMC-oxydoreductase, with each helical strand lining the interior of the protein shell. Proteomic analyses of the purified genomic fibre revealed the presence of several RNA polymerase subunits as well as additional proteins involved in genome compaction that can fit into the central channel of the protein shield. Such an elegant supramolecular organization represents a remarkable evolutionary solution for packaging

while protecting the viral genome, in a state ready for immediate transcription upon unwinding in the host cytoplasm. We expect that a dedicated energy-driven machinery is required for the assembly of this rod-shaped giant viral chromosome and its further compaction in the membrane limited electron dense nucleoid, characteristic of the mature Mimivirus particles^{2,4,5}. The parsimonious implication of the same protein in two functionally unrelated substructures of the virion is also unexpected for a giant virus with a thousand genes at its disposal.

Main

Giant viruses were discovered with the isolation of Mimivirus from *Acanthamoeba* species^{1,6}. These viruses now represent a highly diverse group of dsDNA viruses infecting unicellular eukaryotes⁷ which play important role in the environment⁸. They also challenge the canonical definitions of viruses^{9,10} as they can encode central translation components as well as a complete glycosylation machinery¹¹ among other unique features.

Mimivirus has been the most extensively studied giant virus over the years¹². The virions are 0.75 μm wide and are made of 0.45 μm icosahedral capsids surrounded by a dense layer of fibrils¹. Due to the size of the virions, structural analyses have provided some insights into their icosahedral capsid structure^{2,13–15} but little is known about the packaging of the 1.2 Mb dsDNA genome¹⁶. Inside the capsids, a lipidic membrane delineates an internal compartment (~350 nm in diameter), we refer to as the nucleoid, containing the viral genome, together with all the proteins necessary to initiate the replicative cycle within the host cytoplasm^{17–19}.

Acanthamoeba cells engulf Mimivirus particles, fooled by their bacteria-like size and the sweet taste of the heavily glycosylated decorating fibrils^{1,6}. Once in the vacuole, the Stargate portal located at one specific vertex of the icosahedron opens up²⁰, enabling the viral membrane to fuse with that of the phagosome to deliver the nucleoid into the host cytoplasm^{17,19}. Subsequently, the nucleoid starts losing its electron dense appearance, transcription begins and the early viral factory is formed^{18,21,22}. Previous AFM studies of Mimivirus infectious cycle evidenced a highly condensed nucleoprotein mass enclosing the organized DNA into the nucleoid². We have developed an *in vitro* protocol for particle opening that leads to the release of ~ 30 nm-wide rod-shaped fibres of micrometric lengths. The structure appears to be expelled from the icosahedral capsids, similarly to party whistle, as if packed under high pressure (Fig. 1).

Figure 1: The Mimivirus genomic fibre A] Micrograph of an ultrathin section of Mimivirus capsids filled by dense DNA strings. The thickness of the strings is enhanced by the dehydration caused by the inclusion protocol. B] Negative staining micrograph of an open Mimivirus capsid where the relaxing genomic fibre is still encased into the membrane limited compartment. C] Multiple strands of genomic fibre expelled from the capsid by the proteolytic treatment. D] Negative staining micrograph of purified fibres, E] Cryo-EM micrograph of purified fibres. F] Slices through a cryo-tomogram focused on a long and broken fibre. The plane through which the fibre is viewed is displayed on the cross-sections on the left with blue dashed lines: (from top to bottom) protein shell in the first and fifth slices; the helically folded DNA in the second and fourth slices; the lumen of the fibre in the third slice. Thickness of the slices is 1.1 nm. Distance between the tomographic slices is 4.4 nm between first and second and fourth and fifth slices, and 6.6 nm between the second, third and fourth slices. Multiple states of unwinding can be observed in D, E and F. Scale bars 100 nm.

Various unwinding states of the fibre were observed, ranging from the most compact rod-shaped structures to unfolded ribbons (Fig. 1). After optimizing the extraction on different strains of group-A Mimiviruses, we focused on an isolate from La Réunion Island (Mimivirus reunion) with high yields of fibres that were purified on a sucrose gradient (Extended data Methods). We first confirmed the presence of DNA in the structure by electrophoresis on agarose gels (Extended data Fig. 1) and performed cryo-electron microscopy (cryo-EM) to prospect its location by bubblegram acquisitions^{23,24}. Surprisingly, the sample could sustain very high electron doses of at least 600 e-/Å² for unwinding helices and up to 825 e-/Å² for long compact ones before bubbles appear as a sign of hydrogen gas trapped inside the DNA-protein complex upon protein damage²⁵ (Extended data Fig. 2). No bubbles could be detected

in unfolded ribbons suggesting that in addition to compact the Mimivirus genome, the rod-shaped structure could also protect it inside the proteinaceous shell. Assuming that each turn of DNA in the B-form helix progresses of $\sim 34 \text{ \AA}$ for 10 bp, the Mimivirus 1.2×10^6 bp linear genome would extend over $\sim 400 \text{ }\mu\text{m}$ corresponding to a volume of $1.3 \times 10^6 \text{ nm}^3$ ($\sim 300 \text{ }\mu\text{m}$ for $\sim 1 \times 10^6 \text{ nm}^3$ if in the A-form)²⁶. The spherical nucleoid ($\sim 350 \text{ nm}$ in diameter) corresponds to an approximate volume of $2.2 \times 10^7 \text{ nm}^3$ that can accommodate over a dozen genomes in its linear, unprotected form. However, it can only hold $\sim 30 \text{ }\mu\text{m}$ of a 30-nm-wide flexible cylindrical genomic fibre, which directly implies that the Mimivirus genome cannot be simply fitted linearly in the centre of the fibre, but must be folded back into the structure. Several DNA compaction solutions to this geometrical problem have been described. For instance, the DNA of filamentous viruses infecting archaea^{27,28} is wrapped with proteins to form a ribbon which in turn folds into a helical rod around a hollow lumen. In contrast, the chromatin of cellular eukaryotes consists of DNA wrapped around histone complexes²⁹. The later organization would be consistent with previous evolutionary hypotheses linking giant DNA viruses with the emergence of the eukaryotic nucleus^{30–33}. In order to shed light on Mimivirus packaging strategy, we performed a structural analysis of the purified genomic fibre using cryo-EM single-particle analysis and sub-volume averaging. The different stages of decompaction initially observed by negative staining and cryo-EM (Fig. 1) resulted in a highly heterogeneous dataset for single particle analysis. After pre-processing and manual picking, 2D classes obtained in Relion^{34,35} (Fig. 2, Extended data Fig. 3 & 4) were subjected to cluster analysis using the widths of the helical segments and the correlations between the layer line patterns of the Fourier transforms (Fig. 2, Extended data Methods and Fig. 4) to sort the different conformations. Three independent clusters (CI) could be separated, corresponding to the most (CI1), intermediate (CI2) and least compact (CI3) fibre forms prior to complete unfolding (19,526; 2,357 and 8,648 overlapping segments respectively, Fig. 1, Fig. 2, Extended data Methods and Fig. 4).

Figure 2: Iterative helical 3D reconstruction based on optimized theoretical models for each compaction state of the genomic fibre. 1- A] Automatic sorting of the 2D classes using the fibre width W1 and pairwise correlations of re-oriented 2D classes resulting into 3 main clusters (compact, C11 in blue; intermediate, C12 in cyan and unwound, C13 in orange). B] A theoretical model was then iteratively built (Extended data Fig. 4 & 5) to compute theoretical 2D classes further compared to the experimental ones. C] Two orthogonal views of the refined maps of the compact (C11) and relaxed (C13) forms of the Mimivirus 5- and 6-start helices obtained from Relion³⁵ (Extended data Methods). Each start of the maps has a different colour and images were produced using ChimeraX³⁶.

The least populated intermediate conformation (Fig. 2, C12) was not analysed further and we focused on the most (C11) and the less compact (C13) datasets for independent 3D reconstructions. After 3D classification, we identified two independent structures with approximatively the same width, both made of a ~8 nm-thick proteinaceous external shell, for

each cluster (C11, C13). The most compact conformation (C11) converged to two ~29 nm left-handed helices in which the dsDNA strands are internally lining the protein shell and leave a hollow channel. The most abundant 3D class (13,252 particles) corresponds to a 5-start helix with 5 dsDNA strands following the helical symmetry, located 6.6 Å away from the protein shell, with a ~9 nm wide hollow channel. The second 3D class (12,294 particles) is a 6-start helix with 6 dsDNA strands lining the proteinaceous shell (5.8 Å spacing) and a ~11 nm hollow channel. In C13, the least compact form, the two structures correspond to a ~33.2 nm-wide left-handed helices with either a 5- or a 6-start (5,036 and 6,320 particles respectively), in which the DNA is no longer visible inside the ~17 nm-wide central channel (Fig. 2, Fig. 3, Extended data Table 1).

Figure 3: Structures of the compact form of the Mimivirus 5-start helix: A] Perspective view of the compact form. The map (solid) and cartoon representation of the qu_946 dimers are colour coded according to each start of the 5-start helix. To facilitate the visualisation, the protein shell is presented alone in the front view of the fibre, while it is embedded into the map in the back, expanding the size of the fibre. B] Orthographic view of the sliced structure to visualize the 5 DNA strands lining the protein shell. C] Map of the best 3D reconstruction obtained after refinement in Relion^{34,35}. D] Orthogonal view of the clipped map (transparent) in which was fitted the qu_946 model (solid) and the 5 DNA strands. E] Final cryo-EM map coloured by local resolution from 4.3 Å (blue) to 13.19 Å (red) with Euler angle distribution of the particles used for the 3D reconstruction in red. The Orthogonal view of the colour coded map is at the bottom. The resolution ranges between 4.3 and 8 Å for the protein shell and ~9 Å for the DNA. Images were produced using ChimeraX³⁶, the length of the fibre was defined to present the 5 starts of the helix.

Mass spectrometry-based proteomic analyses identified the homologs of two GMC-oxydoreductases qu_143 and qu_946 (R135 and L894/93 in Mimivirus) as the main components of the purified genomic fibre (Extended data Table 2). This was confirmed by successfully fitting the Mimivirus R135 GMC-oxydoreductase dimeric structure³ (PDB 4Z24, lacking the 50 amino acid long cysteine-rich N-terminal domain) into the ~8 nm external proteinaceous shell after 3D refinement of the well resolved 5-start structure from C11 (Fig. 3

& 4). Quite unexpectedly, GMC-oxydoreductases were already known to compose the fibrils surrounding Mimivirus capsids³⁷. Interestingly, the proteomic analyses measured different sequence coverages for the GMC-oxydoreductases depending whether they were associated to the peripheral fibrils or the genomic fibre, with the striking absence of the N-terminal domain in the genomic fibre (Extended data Fig. 6). There is uninterpreted electronic density in the compact 5-start form in which 5 additional N-terminal residues of the qu_946 sequence could be built for each monomer, compared to the reference X-ray structure. This strikingly brings the cysteines of each monomer (C51 in qu_946) close enough to make a disulphide bridge, directly after the 50 amino acid domain non covered in the proteomic analysis of the genomic fibre (Fig. 4, Extended data Fig. 6 & 8). Inspection of individual fibres in the tomograms and sub-volume averaging confirmed the co-existence of 5- and 6-start helices containing DNA while some intermediates in structure decompaction were also observed. In unwinding or broken fibres, masses that could correspond to proteins inside the lumen were sometime visible as well as directly dissociated DNA fragments, either in the central channel or at the breakage points of the fibres in its periphery. Densities corresponding to the dimer composing the protein shell were also commonly observed on free DNA strands (Extended data. Fig. 7).

In the 3D reconstructions, each dimer is in contact with two independent dsDNA strands through each monomer. As a result, the DNA is interspersed between two dimers, each corresponding to one start of the helix. In the 5-start structure, these contacts involve one aspartate (D82 relative to the N-terminal Methionine in qu_946), one glutamate (E321), three lysines (K84, K344, K685), one arginine (R324) and a histidine (H343) (Fig. 4). The resolution (up to 4.3 Å for the protein shell, Fig. 3E) did not allow the identification of potential ions bridging the DNA phosphates to the protein.

Figure 4: Fitting of the qu_946 crystal structure into the electron density map of the 5-start compact genomic fibre A] Periodic contacts between the dsDNA and the protein shell in the 5-start map. B] Bond

representation of the qu_946 dimer fitted into the 5-start electron density map (up to 4.3 Å resolution) highlighting the interaction between each monomer and one dsDNA strands. The manually built N-terminal residues are visible with the C51 residues making the terminal disulphide bridge. The image was built using the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

We postulated that the distance between the dsDNA strands is defined by each GMC-oxydoreductase with different residues being exposed to interact with the dsDNA. On one hand, while there are no clashes between the dimers in the 5-start 3D reconstruction with either GMC oxydoreductases, there is uninterpreted electron density using the qu_143 protein model uniquely lacking 12 C-terminal residues compared to others (Extended data Fig. 8). On the other hand, when positioning the two GMC oxydoreductases into the best 6-start 3D reconstruction it appeared that only qu_143 could be fitted into the map due to steric hindrance from the longer C-terminal tail in qu_946. We also noticed that in the 6 start helix part of the qu_143 model needed to be reoriented to avoid clashes with adjacent molecules. Even though we did not attempt to optimize the model due to the lack of resolution of the 6 start helix map (6 to 8 Å resolution) (Fig. 2), we were able to compare the spacing between the dsDNA strands (39.5 Å in the 5-start, 36.4 Å in the 6-start) and identified two lysine residue (K319, K342), an arginine (R322) and a glutamine (Q649) in the 6-start qu_143 structure which would be ideally positioned to interact with the DNA. In both 5- and 6-start helices, the genome would need to be further compacted in A-form DNA to fit the genomic fibre into the ~350 nm wide nucleoid. This is consistent with our 3D reconstructions, as only a compact A-form dsDNA would be properly positioned to produce the observed periodic contacts between the protein shell and the dsDNA all along the fibre (Fig. 4A). Thus despite the conformational heterogeneity and the flexibility of the rod-shaped structure we were able to compute an atomic resolution model of the Mimivirus genomic fibre within a map with a 4.3 Å maximum resolution (Fig. 3).

The proteomic analysis revealed the presence of additional proteins including several RNA polymerase subunits: Rpb1 and Rpb2 (qu_530/532 and qu_261/259/257/255), Rpb3/11 (qu_493), Rpb5 (qu_245), and Rpb6 (qu_220), in addition to the mRNA capping enzyme (qu_404), a putative regulator of chromosome condensation (qu_366), the core protein (qu_431), several oxidative stress proteins, a 93 amino-acid-long basic protein of unknown function (qu_544) and additional hypothetical proteins (Extended data Table 2). Upon inspection of the negative staining micrographs, macromolecules strikingly resembling the characteristic groovy structure of the Poxviruses RNA polymerase⁴⁰ were frequently noticed scattered around the unwinding fibre, sometimes sitting on DNA (Extended data Fig. 9). The

hollow lumen of the fibre is large enough to accommodate the Mimivirus RNA-polymerase, most likely sitting on the highly conserved promoter sequence of early genes⁴¹. This central position supports the involvement of the RNA polymerase in genome packaging into the nucleoid and could determine the channel width through its anchoring on the genome.

According to this scenario, the RNA polymerase would be sterically locked inside the compact form of the genomic fibre and then could start moving and transcribing upon helix unwinding, initiating the replicative cycle and the building of the viral factory. This would make the Mimivirus genomic fibre a nuclear-like weapon where the genome and the transcription machinery are compacted into a proteinaceous shield, ready for transcription upon decompaction.

Such an organization is conceptually reminiscent of icosahedral and filamentous dsRNA viruses which package and protect their genomes together with the replicative RNA polymerase into an inner core⁴²⁻⁴⁴. As a result, replication and transcription take place within the protein shield and viral genomes remain protected during their entire infectious cycle. In the case of dsDNA viruses however, the double helix must additionally open up to allow transcription to proceed.

The structure of Mimivirus genomic fibre supports a complex assembly process where the DNA must first be folded on itself prior being packaged, a step which could involve the repeat containing regulator of chromosome condensation (qu_366). As discussed earlier, the composition of the proteinaceous shell, through different contacting residues between the dsDNA phosphates and the two GMC-oxydoreductases, determines the number of dsDNA strands folded into the structure before loading into the nucleoid. Then the RNA polymerase could define the helix diameter through its anchoring onto the dsDNA. Additional proteins such as the 10 kDa basic qu_544 protein with a central hydrophobic segment could serve as linkers between the genome and the protein shell. Finally, in addition to their structural roles, the two GMC-oxydoreductases making the proteinaceous shield, could, together with the other oxidative stress proteins identified in the fibre, alleviate the oxidative stress to which the virions are exposed while entering the cell by phagocytosis. The apparent complexity of the flexible genomic fibre, with various unwinding states and two different helical symmetries, requires additional work to stabilize these complexes in order to identify the contributions of the various proteins composing the fibre and to progress towards 5- and 6-start higher resolution structures. Alternatively, we could perform the structural study of the genomic fibre of another member of the family encoding a unique GC-oxydoreductase.

The genomic fibre is additionally folded into the membrane limited nucleoid² and since it is expelled from the nucleoid as a flexible straight structure, we suspect that an active, energy dependent process is required to bundle it into the nucleoid. To our knowledge, the structure of the genomic fibre used by Mimivirus to packaged and protect its genome is unique in the dsDNA viral world and represents the first description of the genome organization of a giant virus. While herpesviruses^{45,46}, bacteriophages^{47,48} or some archaeal viruses⁴⁹ simply package their dsDNA genome as coils or naked helices, other archaeal viruses tightly wrap it with proteins to fold it into a helix^{27,28,49}. Such an efficient structure is most likely shared by other members of the Mimiviridae family infecting *Acanthamoeba* and could be used by other dsDNA viruses with exclusively cytoplasmic infectious cycles to immediately express early genes upon transfer into the infected cell, like poxviruses⁵⁰. Finally, the parsimonious use of moonlighting GMC-oxydoreductases to play a central role in two functionally unrelated substructures of the Mimivirus particle: i) as a component of the heavily glycosylated peripheral fibril layer, ii) as a proteinaceous shield to package the dsDNA into the genomic fibre, questions the evolutionary incentive leading to such an organization for a virus encoding close to a thousand proteins.

Data availability:

Mimivirus reunion genome has been deposited under the following accession number: BankIt2382307 Seq1 MW004169. 3D reconstruction maps of the 5-start compact and the corresponding PDB have been deposited to EMDB (Deposition number: D_1292110813, 7AIM). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD021585 and 10.6019/PXD021585.

References:

1. Raoult, D. *et al.* The 1.2-megabase genome sequence of Mimivirus. *Science* **306**, 1344–1350 (2004).
2. Kuznetsov, Y. G., Klose, T., Rossmann, M. & McPherson, A. Morphogenesis of mimivirus and its viral factories: an atomic force microscopy study of infected cells. *J. Virol.* **87**, 11200–11213 (2013).
3. Klose, T. *et al.* A Mimivirus Enzyme that Participates in Viral Entry. *Struct. Lond. Engl.* **23**, 1058–1065 (2015).
4. Claverie, J.-M. & Abergel, C. Mimivirus and its virophage. *Annu. Rev. Genet.* **43**, 49–66 (2009).
5. Mutsafi, Y., Zauberman, N., Sabanay, I. & Minsky, A. Vaccinia-like cytoplasmic replication of the giant Mimivirus. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 5978–5982 (2010).
6. La Scola, B. *et al.* A giant virus in amoebae. *Science* **299**, 2033 (2003).
7. Abergel, C., Legendre, M. & Claverie, J.-M. The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. *FEMS Microbiol. Rev.* **39**, 779–796 (2015).
8. Schulz, F. *et al.* Giant virus diversity and host interactions through global metagenomics. *Nature* **578**, 432–436 (2020).
9. Forterre, P. Giant viruses: conflicts in revisiting the virus concept. *Intervirology* **53**, 362–378 (2010).
10. Claverie, J.-M. & Abergel, C. Mimivirus: the emerging paradox of quasi-autonomous viruses. *Trends Genet.* **26**, 431–437 (2010).
11. Piacente, F. *et al.* The rare sugar N-acetylated viosamine is a major component of Mimivirus fibers. *J. Biol. Chem.* **292**, 7385–7394 (2017).
12. Colson, P., La Scola, B., Levasseur, A., Caetano-Anollés, G. & Raoult, D. Mimivirus: leading the way in the discovery of giant viruses of amoebae. *Nat. Rev. Microbiol.* **15**, 243–254 (2017).
13. Xiao, C. *et al.* Structural Studies of the Giant Mimivirus. *PLoS Biol.* **7**, e1000092 (2009).

14. Klose, T. *et al.* The three-dimensional structure of Mimivirus. *Intervirology* **53**, 268–273 (2010).
15. Ekeberg, T. *et al.* Three-dimensional reconstruction of the giant mimivirus particle with an x-ray free-electron laser. *Phys. Rev. Lett.* **114**, 098102 (2015).
16. Chelikani, V., Ranjan, T. & Kondabagil, K. Revisiting the genome packaging in viruses with lessons from the “Giants”. *Virology* doi:10.1016/j.virol.2014.06.022.
17. Claverie, J.-M. *et al.* Mimivirus and Mimiviridae: giant viruses with an increasing number of potential hosts, including corals and sponges. *J. Invertebr. Pathol.* **101**, 172–180 (2009).
18. Arslan, D., Legendre, M., Seltzer, V., Abergel, C. & Claverie, J.-M. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 17486–17491 (2011).
19. Schrad, J. R., Abrahão, J. S., Cortines, J. R. & Parent, K. N. Structural and Proteomic Characterization of the Initiation of Giant Virus Infection. *Cell* **181**, 1046-1061.e6 (2020).
20. Zauberman, N. *et al.* Distinct DNA exit and packaging portals in the virus Acanthamoeba polyphaga mimivirus. *PLoS Biol.* **6**, e114 (2008).
21. Suzan-Monti, M., Scola, B. L., Barrassi, L., Espinosa, L. & Raoult, D. Ultrastructural Characterization of the Giant Volcano-like Virus Factory of Acanthamoeba polyphaga Mimivirus. *PLoS ONE* **2**, e328 (2007).
22. Mutsafi, Y., Fridmann-Sirkis, Y., Milrot, E., Hevroni, L. & Minsky, A. Infection cycles of large DNA viruses: emerging themes and underlying questions. *Virology* **466–467**, 3–14 (2014).
23. Wu, W., Thomas, J. A., Cheng, N., Black, L. W. & Steven, A. C. Bubblegrams Reveal the Inner Body of Bacteriophage KZ. *Science* **335**, 182–182 (2012).
24. Cheng, N., Wu, W., Watts, N. R. & Steven, A. C. Exploiting radiation damage to map proteins in nucleoprotein complexes: The internal structure of bacteriophage T7. *J. Struct. Biol.* **185**, 250–256 (2014).

25. Mishyna, M. *et al.* Effects of radiation damage in studies of protein-DNA complexes by cryo-EM. *Micron* **96**, 57–64 (2017).
26. Li, S., Olson, W. K. & Lu, X.-J. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res.* **47**, W26–W34 (2019).
27. DiMaio, F. *et al.* Virology. A virus that infects a hyperthermophile encapsidates A-form DNA. *Science* **348**, 914–917 (2015).
28. Wang, F. *et al.* Structures of filamentous viruses infecting hyperthermophilic archaea explain DNA stabilization in extreme environments. *Proc. Natl. Acad. Sci.* 202011125 (2020)
doi:10.1073/pnas.2011125117.
29. Robinson, P. J. J., Fairall, L., Huynh, V. A. T. & Rhodes, D. EM measurements define the dimensions of the ‘30-nm’ chromatin fiber: evidence for a compact, interdigitated structure. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 6506–6511 (2006).
30. Takemura, M. Poxviruses and the origin of the eukaryotic nucleus. *J. Mol. Evol.* **52**, 419–425 (2001).
31. Bell, P. J. Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? *J. Mol. Evol.* **53**, 251–256 (2001).
32. Claverie, J.-M. Viruses take center stage in cellular evolution. *Genome Biol.* **7**, 110 (2006).
33. Chaikerasitak, V. *et al.* Assembly of a nucleus-like structure during viral replication in bacteria. *Science* **355**, 194–197 (2017).
34. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
35. He, S. & Scheres, S. H. W. Helical reconstruction in RELION. *J. Struct. Biol.* **198**, 163–176 (2017).
36. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis: UCSF ChimeraX Visualization System. *Protein Sci.* **27**, 14–25 (2018).

37. Boyer, M. *et al.* Mimivirus shows dramatic genome reduction after intraamoebal culture. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 10296–10301 (2011).
38. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation: PROTEIN SCIENCE.ORG. *Protein Sci.* **27**, 293–315 (2018).
39. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallogr. Sect. Struct. Biol.* **75**, 861–877 (2019).
40. Grimm, C. *et al.* Structural Basis of Poxvirus Transcription: Vaccinia RNA Polymerase Complexes. *Cell* **179**, 1537–1550.e19 (2019).
41. Suhre, K., Audic, S. & Claverie, J.-M. Mimivirus gene promoters exhibit an unprecedented conservation among all eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 14689–14693 (2005).
42. Toriyama, S. An RNA-dependent RNA Polymerase Associated with the Filamentous Nucleoproteins of Rice Stripe Virus. *J. Gen. Virol.* **67**, 1247–1255 (1986).
43. Collier, A. M. *et al.* Initiation of RNA Polymerization and Polymerase Encapsidation by a Small dsRNA Virus. *PLOS Pathog.* **12**, e1005523 (2016).
44. Ding, K. *et al.* In situ structures of rotavirus polymerase in action and mechanism of mRNA transcription and release. *Nat. Commun.* **10**, 2216 (2019).
45. Gong, D. *et al.* DNA-Packing Portal and Capsid-Associated Tegument Complexes in the Tumor Herpesvirus KSHV. *Cell* **178**, 1329–1343.e12 (2019).
46. Liu, Y.-T., Jih, J., Dai, X., Bi, G.-Q. & Zhou, Z. H. Cryo-EM structures of herpes simplex virus type 1 portal vertex and packaged genome. *Nature* **570**, 257–261 (2019).
47. Sun, L. *et al.* Cryo-EM structure of the bacteriophage T4 portal protein assembly at near-atomic resolution. *Nat. Commun.* **6**, 7548 (2015).
48. Rao, V. B. & Feiss, M. Mechanisms of DNA Packaging by Large Double-Stranded DNA Viruses. *Annu. Rev. Virol.* **2**, 351–378 (2015).
49. Ptchelkine, D. *et al.* Unique architecture of thermophilic archaeal virus APBV1 and its genome packaging. *Nat. Commun.* **8**, 1436 (2017).

50. Malkin, A. J., McPherson, A. & Gershon, P. D. Structure of Intracellular Mature Vaccinia Virus Visualized by In Situ Atomic Force Microscopy. *J. Virol.* **77**, 6332–6340 (2003).

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 832601). This work was also partially supported by the French National Research Agency ANR-16-CE11-0033-01. Proteomic experiments were partly supported by the ProFI grant (ANR-10-INBS-08-01). The cryo-EM work was performed at the multi-user Cryo-EM Facility at CSSB, supported by UHH and DFG grants (INST 152/772-1|152/774-1|152/775-1|152/776-1). We thank Dr. Carolin Seuring for support and technical help. Processing was performed on the DESY Maxwell cluster and we thank Wolfgang Lugmayr for assistance. Work in the laboratory of Pr. Kay Grünewald is funded by the Wellcome Trust (107806/Z/15/Z), the Leibniz Society, the City of Hamburg and the BMBF (05K18BHA). We also gratefully acknowledge support from the Alexander von Humboldt foundation to Emmanuelle Quemin (individual research fellowship No. FRA 1200789 HFST-P). We thank Pr. Jean-Michel Claverie for his comments on the manuscript and discussions all along the project and Drs. Eric Durand and Juan Reguera for their helpful support on structural work. The preliminary electron microscopy experiments were performed on the PiCSL-FBI core facility (Nicolas Brouilly et/ou Fabrice Richard et/ou Aïcha Aouane, IBDM, AMU-Marseille), member of the France-BioImaging national research infrastructure (ANR-10-INBS-04) and on the IMM imaging platform (Dr. Artemis Kosta). YC and LB thank the support of the bottom-up platform and informatics group of EDyP.

Contributions: C.A. conceived and designed the research; A.V., E.Q., A.S., J-M.A, A.L., L.B., A.C. F.H., Y.C. and C.A. performed research; K.G. granted access to acquisition and analytic tools; A.V., E.Q., A.S, V.P., D.V., K.G. and C.A analysed the data, A.V., A.Q., A.S., K.G. and C.A. wrote the paper.

Competing Financial Interests: None

Figures

Figure 1

The Mimivirus genomic fibre A] Micrograph of an ultrathin section of Mimivirus capsids filled by dense DNA strings. The thickness of the strings is enhanced by the dehydration caused by the inclusion protocol. B] Negative staining micrograph of an open Mimivirus capsid where the relaxing genomic fibre is still encased into the membrane limited compartment. C] Multiple strands of genomic fibre expelled from the capsid by the proteolytic treatment. D] Negative staining micrograph of purified fibres, E] Cryo-EM micrograph of purified fibres. F] Slices through a cryo-tomogram focused on a long and broken fibre. The plane through which the fibre is viewed is displayed on the cross-sections on the left with blue dashed lines: (from top to bottom) protein shell in the first and fifth slices; the helically folded DNA in the second and fourth slices; the lumen of the fibre in the third slice. Thickness of the slices is 1.1 nm. Distance between the tomographic slices is 4.4 nm between first and second and fourth and fifth slices, and 6.6 nm between the second, third and fourth slices. Multiple states of unwinding can be observed in D, E and F. Scale bars 100 nm.

Figure 2

Iterative helical 3D reconstruction based on optimized theoretical models for each compaction state of the genomic fibre. 1- A] Automatic sorting of the 2D classes using the fibre width W1 and pairwise correlations of re-oriented 2D classes resulting into to 3 main clusters (compact, C1 in blue; intermediate, C2 in cyan and unwound, C3 in orange). B] A theoretical model was then iteratively built (Extended data Fig. 4 & 5) to compute theoretical 2D classes further compared to the experimental ones. C] Two orthogonal views of the refined maps of the compact (C1) and relaxed (C3) forms of the Mimivirus 5-

and 6-start helices obtained from Relion35 (Extended data Methods). Each start of the maps has a different colour and images were produced using ChimeraX36.

Figure 3

Structures of the compact form of the Mimivirus 5-start helix: A] Perspective view of the compact form. The map (solid) and cartoon representation of the qu_946 dimers are colour coded according to each start of the 5-start helix. To facilitate the visualisation, the protein shell is presented alone in the front view of the fibre, while it is embedded into the map in the back, expanding the size of the fibre. B] Orthographic view of the sliced structure to visualize the 5 DNA strands lining the protein shell. C] Map of the best 3D reconstruction obtained after refinement in Relion34,35. D] Orthogonal view of the clipped map (transparent) in which was fitted the qu_946 model (solid) and the 5 DNA strands. E] Final cryo-EM map coloured by local resolution from 4.3 Å (blue) to 13.19 Å (red) with Euler angle distribution of the particles used for the 3D reconstruction in red. The Orthogonal view of the colour coded map is at the bottom. The resolution ranges between 4.3 and 8 Å for the protein shell and ~9 Å for the DNA. Images were produced using ChimeraX36, the length of the fibre was defined to present the 5 starts of the helix.

Figure 4

Fitting of the qu_946 crystal structure into the electron density map of the 5-start compact genomic fibre
A] Periodic contacts between the dsDNA and the protein shell in the 5-start map. B] Bond highlighting the interaction between each monomer and one dsDNA strands. The manually built N-terminal residues are visible with the C51 residues making the terminal disulphide bridge. The image was built using the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInformationFinal.pdf](#)