



HAL
open science

The giant mimivirus 1.2 Mb genome is elegantly organized into a 30 nm diameter helical protein shield

Alejandro Villalta, Alain Schmitt, Leandro F Estrozi, Emmanuelle Rj Quemin, Jean-Marie Alempic, Audrey Lartigue, Vojtěch Pražák, Lucid Belmudes, Daven Vasishtan, Agathe Mg Colmant, et al.

► To cite this version:

Alejandro Villalta, Alain Schmitt, Leandro F Estrozi, Emmanuelle Rj Quemin, Jean-Marie Alempic, et al.. The giant mimivirus 1.2 Mb genome is elegantly organized into a 30 nm diameter helical protein shield. *eLife*, 2022, 11, 10.7554/eLife.77607 . hal-03740274

HAL Id: hal-03740274

<https://amu.hal.science/hal-03740274>

Submitted on 29 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

1 **Title: The giant mimivirus 1.2 Mb genome is elegantly organized into a 30 nm diameter**
2 **helical protein shield**

3 **Authors:** Alejandro Villalta^{1†}, Alain Schmitt^{1†}, Leandro F. Estrozi^{2†}, Emmanuelle R. J.
4 Queminn^{3#}, Jean-Marie Alempic¹, Audrey Lartigue¹, Vojtěch Pražák^{3,4}, Lucid Belmudes⁵, Da-
5 ven Vasishthan^{3,4}, Agathe M. G. Colmant¹, Flora A. Honoré¹, Yohann Couté⁵, Kay Grüne-
6 wald^{3,4}, Chantal Abergel^{1*}

7 **Affiliations:**

8 ¹Aix–Marseille University, Centre National de la Recherche Scientifique, Information Génomique & Structurale, Unité Mixte de Recherche 7256 (Institut de Microbiologie de la Méditerranée, FR3479, IM2B), 13288 Marseille Cedex 9, France.

11 ²Univ. Grenoble Alpes, CNRS, CEA, Institut de Biologie Structurale (IBS), 38000 Grenoble, France.

13 ³Centre for Structural Systems Biology, Leibniz Institute for Experimental Virology (HPI), University of Hamburg, D-22607 Hamburg, Germany.

15 ⁴Division of Structural Biology, Wellcome Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, United Kingdom.

17 ⁵Univ. Grenoble Alpes, CEA, INSERM, IRIG, BGE, 38000 Grenoble, France

18 *Corresponding author. Email: Chantal.Abergel@igs.cnrs-mrs.fr

19 †These authors contributed equally to the work

20 # present address: Department of Virology, Institute for Integrative Biology of the Cell (I2BC), Centre National de la Recherche Scientifique UMR9198, 1, avenue de la Ter-rasse, 91198 Gif-sur-Yvette Cedex, France.

23 **One-Sentence Summary:** Mimivirus genome organization in the icosahedral virion.

24 **Subject Area:** Virology, Cro-electron microscopy

25 **Keywords:** Mimivirus, giant virus, dsDNA, viral genome organization, genomic fiber, cryo-EM, cryo-ET, helical reconstruction, proteomics.

28 **Abstract**

29 Mimivirus is the prototype of the *Mimiviridae* family of giant dsDNA viruses. Little is known
30 about the organization of the 1.2 Mb genome inside the membrane-limited nucleoid filling the
31 ~0.5 µm icosahedral capsids. Cryo-electron microscopy, cryo-electron tomography and
32 proteomics revealed that it is encased into a ~30 nm diameter helical protein shell surprisingly
33 composed of two GMC-type oxidoreductases, which also form the glycosylated fibrils
34 decorating the capsid. The genome is arranged in 5- or 6-start left-handed super-helices, with
35 each DNA-strand lining the central channel. This luminal channel of the nucleoprotein fiber is
36 wide enough to accommodate oxidative stress proteins and RNA polymerase subunits
37 identified by proteomics. Such elegant supramolecular organization would represent a
38 remarkable evolutionary strategy for packaging and protecting the genome, in a state ready
39 for immediate transcription upon unwinding in the host cytoplasm. The parsimonious use of
40 the same protein in two unrelated substructures of the virion is unexpected for a giant virus
41 with thousand genes at its disposal.

43 **Introduction**

44 *Acanthamoeba* infecting giant viruses were discovered with the isolation of mimivirus (1, 2).
45 Giant viruses now represent a highly diverse group of dsDNA viruses infecting unicellular
46 eukaryotes (3) which play important roles in the environment (4–6). They also challenge the
47 canonical definitions of viruses (7, 8) as they can encode central translation components (2, 9)
48 as well as a complete glycosylation machinery (10, 11) among other unique features.

49 Mimivirus has been the most extensively studied giant virus infecting *Acanthamoeba* (12).
50 The virions are 0.75 μm wide and consist of icosahedral capsids of 0.45 μm diameter
51 surrounded by a dense layer of radially arranged fibrils (2). Structural analyses of the virions
52 have provided some insights into the capsid structure (13–17) but given the size of the
53 icosahedral particles (and hence the sample thickness), accessing the internal organization of
54 the core of the virions remains challenging. Consequently, little is known about the packaging
55 of the 1.2 Mb dsDNA genome (18). Inside the capsids, a lipid membrane delineates an
56 internal compartment (~340 nm in diameter, Fig 1A, Figure 1-figure supplement 1), referred
57 to as the nucleoid, which contains the viral genome, together with all proteins necessary to
58 initiate the replicative cycle within the host cytoplasm (17, 19–21). *Acanthamoeba* cells
59 engulf mimivirus particles, fooled by their bacteria-like size and the heavily glycosylated
60 decorating fibrils (1, 2, 11). Once in the phagosome, the Stargate portal located at one specific
61 vertex of the icosahedron opens up (22), enabling the viral membrane to fuse with that of the
62 host vacuole to deliver the nucleoid into the host cytoplasm (17, 20). EM studies have shown
63 that next the nucleoid gradually loses its electron dense appearance, transcription begins and
64 the early viral factory is formed (21, 23, 24). Previous Atomic Force Microscopy (AFM)
65 studies of the mimivirus infectious cycle suggested that the DNA forms a highly condensed
66 nucleoprotein complex enclosed within the nucleoid (13). Here we show that opening of the
67 large icosahedral capsid *in vitro* led to the release of rod-shaped structures of about 30 nm
68 width. These structures were further purified and the various conformations characterized
69 using cryo-electron microscopy (cryo-EM), tomography and MS-based proteomics.

70

71 **Results**

72 **Capsid opening induces the release of a ~ 30 nm-wide rod-shaped structure that contains** 73 **the dsDNA genome**

74 We developed an *in vitro* protocol for particle opening that led to the release of ~ 30 nm-wide
75 rod-shaped fibers of several microns in length (Fig. 1, Fig. 1-figure supplement 1). We coined
76 this structure the mimivirus genomic fiber. Complete expulsion of the opened nucleoid
77 content produced bundled fibers resembling a “ball of yarn” (13) (Fig. 1C). The capsid
78 opening procedure involves limited proteolysis and avoids harsh conditions, as we found that
79 the structure becomes completely denatured by heat (95°C) and is also sensitive to acidic
80 treatment, thus preventing its detection in such conditions (17). Various conformations of the
81 genomic fiber were observed, sometimes even on the same fiber (Fig. 1D), ranging from the
82 most compact rod-shaped structures (Fig. 1D (left), E-F (top)) to more relaxed structures
83 where DNA strands begin to dissociate (Fig. 1D (right), E-F (bottom)). After optimizing the
84 *in vitro* extraction on different strains of group-A mimiviruses, we focused on an isolate from
85 La Réunion Island (mimivirus reunion), as more capsids were opened by our protocol, leading
86 to higher yields of genomic fibers that were subsequently purified on sucrose gradient. All
87 opened capsids released genomic fibers (Fig. 1C, Fig. 1-figure supplement 1).

88 The first confirmation of the presence of DNA in the genomic fiber was obtained by agarose
89 gel electrophoresis (Fig. 1-figure supplement 3). Electron cryo microscopy (cryo-EM)

90 bubblegram analysis (25, 26) gave a further indication that the nucleic acid is located in the
91 fiber lumen. Alike other nucleoprotein complexes, fibers are expected to be more susceptible
92 to radiation damage than pure proteinaceous structures. Surprisingly, the specimen could
93 sustain higher electron irradiation before the appearance of bubbles compared to other studies
94 (27): 600 e-/Å² for relaxed helices and up to 900 e-/Å² for long compact ones, while no
95 bubbles could be detected in unfolded ribbons (Fig. 1-figure supplement 4). For comparison,
96 bacteriophage capsids containing free DNA, i.e. not in the form of nucleoproteins, show
97 bubbling for doses of ~30-40 e-/Å² (27).

98 **Cryo-EM single particle analysis of the different compaction states of the mimivirus** 99 **genomic fiber**

100 In order to shed light on the mimivirus genome packaging strategy and to determine the
101 structure of the purified genomic fibers, we performed cryo-EM single-particle analysis. The
102 different conformations of the genomic fiber initially observed by negative staining (Fig. 1)
103 and cryo-EM resulted in a highly heterogeneous dataset for single-particle analysis. In order
104 to separate different conformations, *in silico* sorting through 2D classification (using Relion
105 (28, 29)) was performed. Next, we carried out cluster analysis relying on the widths of the
106 helical segments and correlations (real and reciprocal spaces) between the experimental 2D
107 class patterns (Fig. 1-figures supplement 5 & 6). Three independent clusters (C1) could be
108 distinguished, corresponding to the compact (C11), intermediate (C12) and relaxed (C13) fiber
109 conformations (Fig. 1-figure supplement 6) with the latter being the widest. For each cluster,
110 we determined their helical symmetry parameters by image power spectra analyses and
111 performed structure determination and refinement (Fig. 2, Fig. 2-figures supplement 1-3).

112 For both C11a and C13a conformations, after 3D-refinement, we obtained helical structures of
113 3.7 Å resolution (FSC threshold 0.5, masked), corresponding to 5-start left-handed helices
114 made of a ~8 nm-thick proteinaceous external shell (Fig. 2-figures supplement 1-2). For the
115 most compact conformation (C11a) 5 dsDNA strands were lining the interior of the protein
116 shell leaving a ~9 nm wide central channel (Fig. 2-3, supplementary file 1). The dsDNA
117 strands appear as curved cylinders in the helical structure, the characteristic shape of the DNA
118 (minor and major groove) becoming only visible after focused refinement of a single strand of
119 dsDNA (Fig. 3, Fig. 3-figures supplement 1-2). In the relaxed subcluster C13a, the DNA
120 strands at the interface to the ~17 nm-wide central channel are not clearly recognizable (Fig.
121 2, supplementary file 1 & Fig. 2-figure supplement 2), most likely because they are at least
122 partially detached inside the broken expanded fiber. The breaks after relaxation of the helix
123 might be the result of the extraction and purification treatment, while DNA will remain in the
124 central channel, at least in the early phase of *Acanthamoeba* infection.

125 Finally, the 4 Å resolution C12 map obtained after 3D refinement (Fig. 2-figure supplement 3)
126 corresponds to a 6-start left-handed helix made of a ~8 nm-thick proteinaceous external shell,
127 with 6 dsDNA strands lining the shell interior and leaving a ~12 nm wide inner channel (Fig.
128 2, supplementary file 1).

129

130 **The most abundant proteins in the genomic fiber are GMC-oxidoreductases, the same** 131 **that compose the fibrils decorating mimivirus capsid**

132 MS-based proteomic analyses performed on three biological replicates identified two GMC-
133 oxidoreductases as the main components of the purified genomic fiber (qu_946 and qu_143 in
134 mimivirus reunion corresponding to L894/93 and R135 respectively in mimivirus prototype)
135 (supplementary file 2). The two mimivirus reunion proteins share 69% identity (81%
136 similarity). The available mimivirus R135 GMC-oxidoreductase dimeric structure (30) (PDB

137 4Z24, lacking the 50 amino acid long cysteine-rich N-terminal domain) was fitted into the
138 EM maps (Figs. 2-3). This is quite unexpected, since GMC-oxidoreductases are already
139 known to compose the fibrils surrounding mimivirus capsids (11, 31). The corresponding
140 genes are highly expressed during the late phase of the infection cycle at the time of virion
141 assembly. Notably, the proteomic analyses provided different sequence coverages for the
142 GMC-oxidoreductases depending on whether samples were intact virions or purified genomic
143 fiber preparations, with substantial under-representation of the N-terminal domain in the
144 genomic fiber (Fig. 2-figure supplement 4). Accordingly, the maturation of the GMC-
145 oxidoreductases involved in genome packaging must be mediated by one of the many
146 proteases encoded by the virus or the host cell. Interestingly, mimivirus M4 (31), a laboratory
147 strain having lost the genes responsible for the synthesis of the two polysaccharides
148 decorating mimivirus fibrils (11) also lacks the GMC-oxidoreductase genes. Additional
149 studies on this specific variant will be key to establish if it exhibits a similar genomic fiber,
150 and if yes, which proteins are composing it.

151 **Analysis of the genomic fiber structure**

152 The EM maps and FSC curves of C11a are shown in Fig. 2-figure supplement 1. An additional
153 step of refinement focused on the asymmetric unit further improved the local resolution to 3.3
154 Å as indicated by the corresponding FSC (Fig. 3 and Fig. 2-figure supplement 1). After fitting
155 the most abundant GMC-oxidoreductase qu_946 (SWISS-MODEL model (32)) in the final
156 map of C11a, five additional N-terminal residues in each monomer were manually built using
157 the uninterpreted density available. This strikingly brings the cysteines of each monomer
158 (C51 in qu_946) close enough to allow a disulfide bridge, directly after the 50 amino acids
159 domain not covered in the proteomic analysis of the genomic fiber (Fig. 3G, Fig. 2-figure
160 supplement 4). The N-terminal chain, being more disordered than the rest of the structure, it is
161 absent in the focused refined map, and also absent in the C13a map of the relaxed helix,
162 suggesting that a break of the disulfide bridge could be involved in the observed unwinding
163 process. Models of the three helical assemblies and asymmetric unit were further refined
164 using the real-space refinement program in PHENIX 1.18.2 (33). In the 3.3 Å resolution map
165 of the asymmetric unit, most side chains and notably the FAD co-factor are accommodated by
166 density suggesting that the oxidoreductase enzyme could be active (Fig. 3D). Density that can
167 be attributed to the FAD cofactor is also present in the C12 and C13a maps. The atomic models
168 of C11a and C13a dimers are superimposable with a core RMSD of 0.68 Å based on C α atoms.

169

170 Inspection of individual genomic fibers in the tomograms confirmed the co-existence of both
171 5- and 6-start left-handed helices containing DNA (Fig. 1-figure supplement 2 & Fig. 3-
172 animation 1, Fig.1-video 1). Further, some intermediate and relaxed structures were also
173 observed in which the DNA segments appeared detached from the protein shell and
174 sometimes completely absent from the central channel of the broken fibers. Both GMC-
175 oxidoreductases (qu_946 and qu_143) can be fitted in the 5- and 6-start maps.

176 In relaxed or broken fibers, large electron dense structures that might correspond to proteins
177 inside the lumen were sometimes visible (Fig. 1-figure supplement 2C & E & Fig.1-video 3)
178 as well as dissociating DNA fragments, either in the central channel (Fig. 1-figure supplement
179 2B & Fig.1-video 2) or at the breakage points of the fibers in its periphery (Fig. 1-figure
180 supplement 2D & Fig.1-video 4). Densities corresponding to the dimer subunits composing
181 the protein shell were also commonly observed on dissociated DNA strands (Fig. 1E-F, Fig.
182 1-figure supplement 2 and Fig.1-video 1-4).

183 In the C11a and C12 helices, the monomers in each dimer are interacting with two different
184 dsDNA strands. As a result, the DNA strands are interspersed between two dimers, each also

185 corresponding to a different strand of the protein shell helix (Fig. 3). Based on the periodic
186 contacts between protein shell and DNA strands, these interactions might involve, in the case
187 of the C11a helix, one aspartate (D82 relative to the N-terminal Methionine in qu_946), one
188 glutamate (E321), two lysines (K344, K685), one arginine (R324) and a histidine (H343) or
189 one asparagine (N80), two lysines (K319, K342), one arginine (R322) and one tyrosine
190 (Y687), in the case of the qu_143 (Fig. 3E-F, supplementary file 3). Intra- and inter-strands
191 contacts between each dimer are presented in supplementary file 3 for qu_143 and qu_946 in
192 C11a, C12 and C13 maps.

193 Despite the conformational heterogeneity and the flexibility of the rod-shaped structure, we
194 were able to build three atomic models of the mimivirus genomic fiber, in compact (5- and 6-
195 start) and relaxed (5-start) states. Higher resolution data would still be needed to determine
196 the precise structure of the dsDNA corresponding to the viral genome (Fig. 3B, H), however,
197 the lower resolution for this part of the map even in focused refinement runs (Fig. 3H) might
198 also mean that the DNA does not always bind in the same orientation.

199 **Rough estimation of genome compaction to fit into the nucleoid**

200 Since there is a mixture of 5 and 6 strands of DNA in the genomic fiber, this could correspond
201 to 5 or 6 genomes per fiber or to a single folded genome. Assuming that the length of DNA in
202 B-form is $\sim 34 \text{ \AA}$ for 10 bp, the mimivirus linear genome of 1.2×10^6 bp would extend over
203 $\sim 400 \text{ \mu m}$ and occupy a volume of $1.3 \times 10^6 \text{ nm}^3$ ($\sim 300 \text{ \mu m}$ and $\sim 1 \times 10^6 \text{ nm}^3$ if in A-form) (34).
204 The volume of the nucleoid (19) ($\sim 340 \text{ nm}$ in diameter) is approximately $2.1 \times 10^7 \text{ nm}^3$ and
205 could accommodate over 12 copies of viral genomes in a naked state, but only 40 \mu m of the
206 $\sim 30 \text{ nm}$ wide flexible genomic fiber. Obviously, the mimivirus genome cannot be simply
207 arranged linearly in the genomic fiber and must undergo further compaction to accommodate
208 the 1.2 Mb genome in a $\sim 40 \text{ \mu m}$ long genomic fiber. As a result, the complete mimivirus
209 genome, folded at least 5 times, fits into the helical shell. This structure surprisingly
210 resembles a nucleocapsid, such as the archaea infecting APBV1 nucleocapsid (35).

211 **Additional proteins, including RNA polymerase subunits, are enriched in the genomic 212 fiber**

213 The proteomic analysis of fiber preparations revealed the presence of additional proteins
214 including several RNA polymerase subunits: Rpb1 and Rpb2 (qu_530/532 and
215 qu_261/259/257/255), Rpb3/11 (qu_493), Rpb5 (qu_245), RpbN (qu_379) and Rpb9
216 (qu_219), in addition to a kinesin (qu_313), a regulator of chromosome condensation
217 (qu_366), a helicase (qu_572), to be possibly associated with the genome (supplementary file
218 2). In addition to the two GMC-oxidoreductases, at least three oxidative stress proteins were
219 also identified together with hypothetical proteins (supplementary file 2). RNA polymerase
220 subunits start being expressed 1 h post infection with a peak after 5h and are expressed until
221 the end the infection cycle. GMC-oxidoreductases, kinesin, regulator of chromosome
222 condensation are all expressed after 5 hours of infection until the end of the cycle.

223 As expected, the core protein (qu_431) composing the nucleoid and the major capsid proteins
224 (MCP, qu_446) were significantly decreased in the genomic fiber proteome compared to
225 intact virions. In fact, qu_431 and qu_446 represent respectively 4.5% and 9.4% of the total
226 protein abundance in virions whereas they only account for 0.4% and 0.7% of the total protein
227 abundance in the genomic fiber, suggesting that they could be contaminants in this
228 preparation. On the contrary, we calculated enrichment factors of more than five hundred
229 (qu_946) and twenty-six (qu_143) in the genomic fiber samples compared to the intact virion.
230 Finally, the most abundant RNA polymerase subunit (qu_245) is increased by a factor of
231 eight in the genomic fiber compared to intact virion (if the six different subunits identified are
232 used, the global enrichment is seven-fold). Furthermore, upon inspection of the negative

233 staining micrographs, macromolecules strikingly resembling the characteristic structure of the
234 poxviruses RNA polymerase (36) were frequently observed scattered around the unwinding
235 fiber and sometimes sitting on DNA strands near broken fibers (Fig. 4). Together with the
236 tomograms showing large electron dense structures in the lumen, some RNA polymerase
237 units could occupy the center of the genomic fiber.

238 **Discussion**

239 Several DNA compaction solutions have been described. For instance, the DNA of
240 filamentous viruses infecting archaea is wrapped by proteins to form a ribbon which in turn
241 folds into a helical rod forming a cavity in its lumen (37, 38). In contrast, the chromatin of
242 cellular eukaryotes consists of DNA wrapped around histone complexes (39). It was recently
243 shown that the virally encoded histone doublets of the *Marseilleviridae* can form nucleosomes
244 (40, 41) and such organization would be consistent with previous evolutionary hypotheses
245 linking giant DNA viruses with the emergence of the eukaryotic nucleus (42–46).
246 Herpesviruses (47, 48), bacteriophages (49, 50) and APBV1 archaeal virus (35) package their
247 dsDNA genome as naked helices or coils. Yet, APBV1 nucleocapsid structure strikingly
248 resemble the mimivirus genomic fiber with a proteinaceous shell enclosing the folded dsDNA
249 genome. Consequently, mimivirus genomic fiber is a nucleocapsid further bundled as a ball of
250 yarn into the nucleoid, itself encased in the large icosahedral capsids. The structure of the
251 mimivirus genomic fiber described herein supports a complex assembly process where the
252 DNA must be folded into 5 or 6 strands prior to or concomitant with packaging, a step that
253 may involve the repeat containing regulator of chromosome condensation (qu_366) identified
254 in the proteomic analysis of the genomic fiber. The proteinaceous shell, *via* contacting
255 residues between the dsDNA and the GMC-oxidoreductases, would guide the folding of the
256 dsDNA strands into the structure prior loading into the nucleoid. The lumen of the fiber being
257 large enough to accommodate the mimivirus RNA polymerase, we hypothesize that it could
258 be sitting on the highly conserved promoter sequence of early genes (51). This central
259 position would support the involvement of the RNA polymerase in genome packaging into
260 the nucleoid and could determine the channel width *via* its anchoring on the genome (Fig.
261 4D). According to this scenario, the available space (although tight) for the RNA polymerase
262 inside the genomic fiber lumen suggests it could be sterically locked inside the compact form
263 of the genomic fiber and could start moving and transcribing upon helix relaxation, initiating
264 the replicative cycle and the establishment of the cytoplasmic viral factory. The genome and
265 the transcription machinery would thus be compacted together into a proteinaceous shield,
266 ready for transcription upon relaxation (Fig. 2-video 1). This organization would represent a
267 remarkable evolutionary strategy for packaging and protecting the viral genome, in a state
268 ready for immediate transcription upon unwinding in the host cytoplasm. This is conceptually
269 reminiscent of icosahedral and filamentous dsRNA viruses which pack and protect their
270 genomes together with the replicative RNA polymerase into an inner core (52–54). As a
271 result, replication and transcription take place within the protein shield and viral genomes
272 remain protected during their entire infectious cycle. In the case of dsDNA viruses however,
273 the double helix must additionally open up to allow transcription to proceed, possibly
274 involving the helicase identified in our proteomic study (supplementary file 2). Finally, in
275 addition to their structural roles, the FAD containing GMC-oxidoreductases making the
276 proteinaceous shield, together with other oxidative stress proteins (supplementary file 2),
277 could alleviate the oxidative stress to which the virions are exposed while entering the cell by
278 phagocytosis.

279 Mimivirus virion thus appears as a Russian doll, with its icosahedral capsids covered with
280 heavily glycosylated fibrils, two internal membranes, one lining the capsid shell, the other
281 encasing the nucleoid, in which the genomic fiber is finally folded. To our knowledge, the

282 structure of the genomic fiber used by mimivirus to package and protect its genome in the
283 nucleoid represents the first description of the genome organization of a giant virus. Since the
284 genomic fiber appears to be expelled from the nucleoid as a flexible and subsequently straight
285 structure starting decompaction upon release, we suspect that an active, energy-dependent,
286 process is required to bundle it into the nucleoid during virion assembly. Such an efficient
287 structure is most likely shared by other members of the Mimiviridae family infecting
288 *Acanthamoeba* and could be used by other dsDNA viruses relying on exclusively cytoplasmic
289 replication like poxviruses to immediately express early genes upon entry into the infected
290 cell (55, 56). Finally, the parsimonious use of moonlighting GMC-oxidoreductases playing a
291 central role in two functionally unrelated substructures of the mimivirus particle - i) as a
292 component of the heavily glycosylated peripheral fibril layer and ii) as a proteinaceous shield
293 to package the dsDNA into the genomic fibers questions the evolutionary incentive leading to
294 such an organization for a virus encoding close to a thousand proteins.

295 **Materials and Methods**

296 *Nucleoid extraction*

297 Mimivirus reunion virions defibrillated, as described previously (11, 19), were centrifuged at
298 10,000 x g for 10 min, resuspended in 40 mM TES pH 2 and incubated for 1h at 30°C to
299 extract the nucleoid from the opened capsids.

300 *Extraction and purification of the mimivirus genomic fiber*

301 The genomic fiber was extracted from 12 mL of purified mimivirus reunion virions at 1.5×10^{10}
302 particles/mL, split into 12x1 mL samples processed in parallel. Trypsin (Sigma T8003)
303 in 40 mM Tris-HCl pH 7.5 buffer was added at a final concentration of 50 µg/mL and the
304 virus-enzyme mix was incubated for 2h at 30°C in a heating dry block (Grant Bio PCH-1).
305 DTT was then added at a final concentration of 10 mM and incubated at 30°C for 16h.
306 Finally, 0.001% Triton X-100 was added to the mix and incubated for 4h at 30°C. Each tube
307 was vortexed for 20 s with 1.5 mm diameter stainless steel beads (CIMAP) to separate the
308 fibers from the viral particles and centrifuged at 5,000 x g for 15 min to pellet the opened
309 capsids. The supernatant was recovered, and the fibers were concentrated by centrifugation at
310 15,000 x g for 4h at 4°C. Most of the supernatant was discarded leaving 12x~200 µL of
311 concentrated fibers that were pooled and layered on top of ultracentrifuge tubes of 4 mL
312 (polypropylene centrifuge tubes, Beckman Coulter) containing a discontinuous sucrose
313 gradient (40%, 50%, 60%, 70% w/v in 40 mM Tris-HCl pH 7.5 buffer). The gradients were
314 centrifuged at 200,000 x g for 16h at 4 °C. Since no visible band was observed, successive 0.5
315 mL fractions were recovered from the bottom of the tube, the first one supposedly
316 corresponding to 70% sucrose. Each fraction was dialyzed using 20 kDa Slide-A-Lyzers
317 (ThermoFisher) against 40 mM Tris-HCl pH 7.5 to remove the sucrose. These fractions were
318 further concentrated by centrifugation at 15,000 x g, at 4°C for 4h and most of the supernatant
319 was removed, leaving ~100 µL of sample at the bottom of each tube. At each step of the
320 extraction procedure the sample was imaged by negative staining transmission electron
321 microscopy (TEM) to assess the integrity of the genomic fiber (Fig. 1-figure supplement 1).
322 Each fraction of the gradient was finally controlled by negative staining TEM. For proteomic
323 analysis, an additional step of concentration was performed by speedvac (Savant
324 SPD131DDA, Thermo Scientific).

325 *Negative stain TEM*

326 300 mesh ultra-thin carbon-coated copper grids (Electron Microscopy Sciences, EMS) were
327 prepared for negative staining by adsorbing 4-7 µL of the sample for 3 min., followed by two
328 washes with water before staining for 2 min in 2% uranyl acetate. The grids were imaged

329 either on a FEI Tecnai G2 microscope operated at 200 keV and equipped with an Olympus
330 Veleta 2k camera (IBDM microscopy platform, Marseille, France); a FEI Tecnai G2
331 microscope operated at 200 keV and equipped with a Gatan OneView camera (IMM,
332 microscopy platform, France) or a FEI Talos L120c operated at 120 keV and equipped with a
333 Ceta 16M camera (CSSB multi-user cryo-EM facility, Germany, Fig. 1, Fig. 1-figure
334 supplement 1).

335 *Cryo-electron tomography*

336 **Sample preparation**

337 For cryo-ET of the mimivirus genomic fiber, samples were prepared as described above for
338 single-particle analysis except that 5 nm colloidal gold fiducial markers (UMC, Utrecht) were
339 added to the sample right before plunge freezing at a ratio of 1:2 (sample: fiducial markers).

340 **Data acquisition**

341 Tilt series were acquired using SerialEM (57) on a Titan Krios (Thermo Scientific)
342 microscope operated at 300 keV and equipped with a K3 direct electron detector and a GIF
343 BioQuantum energy filter (Gatan). We used the dose-symmetric tilt-scheme (58) starting at 0°
344 with a 3° increment to +/-60° at a nominal magnification of 64,000x, a pixel size of 1.4 Å and
345 a total fluence of 150 e-/Å² over the 41 tilts (*i.e.* ~3.7 e-/Å²/tilt for an exposure time of 0.8 s
346 fractionated into 0.2 s frames (supplementary file 4).

347 **Data processing**

348 Tilt series were aligned and reconstructed using the IMOD (59). For visualization purposes,
349 we applied a binning of 8 and SIRT-like filtering from IMOD (59) as well as a bandpass
350 filter bsoft (60). The tomograms have been deposited on EMPIAR, accession number 1131
351 and videos were prepared with Fiji (Fig. 1E-F, Fig. 1-figure supplement 2 & Fig.1-video 1-
352 4).

353 *Agarose gel electrophoresis and DNA dosage to assess the presence of DNA into the fiber*

354 Genomic DNA was extracted from 10¹⁰ virus particles using the PureLink™ Genomic DNA
355 mini kit (Invitrogen) according to the manufacturer's protocol. Purified genomic fiber was
356 obtained following the method described above. The purified fiber was treated by adding
357 proteinase K (PK) (Takara ST 0341) to 20 µL of sample (200 ng as estimated by dsDNA
358 Qubit fluorometric quantification) at a final concentration of 1 mg/mL and incubating the
359 reaction mix at 55°C for 30 min. DNase treatment was done by adding DNase (Sigma
360 10104159001) and MgCl₂ to a final concentration of 0.18 mg/mL and 5 mM, respectively, in
361 20 µL of sample and incubated at 37°C for 30 min prior to PK treatment. For RNase
362 treatment, RNase (Sigma SLBW2866) was added to 20 µL (200 ng) of sample solution to a
363 final concentration of 1 mg/mL and incubated at 37°C for 30 min prior to PK treatment. All
364 the samples were then loaded on a 1 % agarose gel and stained with ethidium bromide after
365 migration. The bands above the 20 kbp marker correspond to the stacked dsDNA fragments
366 of various lengths compatible with the negative staining images of the broken fibers where
367 long DNA fragment are still attached to the helical structure. The mimivirus purified genomic
368 DNA used as a control migrates at the same position (Fig. 1-figure supplement 3).

369 *Cryo-EM bubblegram analysis*

370 Samples were prepared as described for single-particle analysis. Dose series were acquired on
371 a Titan Krios (Thermo Scientific) microscope operated at 300 keV and equipped with a K3
372 direct electron detector and a GIF BioQuantum (Gatan) energy filter. Micrographs were
373 recorded using SerialEM (57) at a nominal magnification of 81,000x, a pixel size of 1.09 Å

374 and a rate of 15 e⁻/pixel/s (Fig. 1-figure supplement 4 & supplementary file 4). Dose series
375 were acquired by successive exposures of 6 s, resulting in an irradiation of 75 e⁻/Å² per
376 exposure. Micrographs were acquired with 0.1 s frames and aligned in SerialEM (57). In a
377 typical bubblegram experiment, 12 to 15 successive exposures were acquired in an area of
378 interest with cumulative irradiations of 900 to 1,125 e⁻/Å² total (Fig. 1-figure supplement 4).

379 *Single-particle analysis by cryo-EM*

380 *Sample preparation*

381 For single-particle analysis, 3 μL of the purified sample were applied to glow-discharged
382 Quantifoil R 2/1 Cu grids, blotted for 2 s using a Vitrobot Mk IV (Thermo Scientific) and
383 applying the following parameters: 4°C, 100% humidity, blotting force 0, and plunge frozen
384 in liquid ethane/propane cooled to liquid nitrogen temperature.

385 *Data acquisition*

386 Grids were imaged using a Titan Krios (Thermo Scientific) microscope operated at 300 keV
387 and equipped with a K2 direct electron detector and a GIF BioQuantum energy filter
388 (Gatan). 7,656 movie frames were collected using the EPU software (Thermo Scientific) at a
389 nominal magnification of 130,000x with a pixel size of 1.09 Å and a defocus range of -1 to -
390 3 μm. Micrographs were acquired using EPU (Thermo Scientific) with 8 s exposure time,
391 fractionated into 40 frames and 7.5 e⁻/pixel/s (total fluence of 50.5 e⁻/Å²)
392 (supplementary file 4).

393 *2D classification and clustering of 2D classes*

394 All movie frames were aligned using MotionCor2 (61) and used for contrast transfer function
395 (CTF) estimation with CTFFIND-4.1 (62). Helical segments of the purified genomic fibers,
396 manually picked with Relion 3.0 (28, 29)(28, 29)(28, 29)(28, 29)(28, 29), were initially
397 extracted with different box sizes, 400 pixels for 3D reconstructions, 500 pixels for initial 2D
398 classifications and clustering and 700 pixels to estimate the initial values of the helical
399 parameters. Particles were subjected to reference-free 2D classification in Relion 3.1.0 (28,
400 29), where multiple conformations of the fiber were identified (Fig. 1-figures supplement 5 &
401 6).

402 We then performed additional cluster analysis of the 194 initial 2D classes provided by Relion
403 (Fig. 1-figure supplement 5) to aim for more homogeneous clusters, eventually corresponding
404 to different states (Fig. 1-figure supplement 6). A custom 2-step clustering script was written
405 in python with the use of Numpy(63) and Scikit-learn (64) libraries. First, a few main clusters
406 were identified by applying a DBSCAN (65) clustering algorithm on the previously estimated
407 fiber external width values (W1). The widths values, estimated by adjusting a parameterized
408 cross-section model on each 2D stack, range from roughly 280 Å to 340 Å. The cross-section
409 model fitting process is based on adjusting a section profile (S) described in Equation 1 on the
410 longitudinally integrated 2D-class profile (Fig. 1-figure supplement 6A). The cross-section
411 model is composed of a positive cosine component, (parameterized by the center position μ,
412 its width σ₁ and amplitude a₁) associated with the fiber external shell, a negative cosine
413 component (parameterized by the center position μ, its width σ₂ and amplitude a₂) associated
414 with the central hollow lumen (W2), and a constant p₀, accounting for the background level,
415 as

$$S(\mu, \sigma_1, a_1, \sigma_2, a_2, p_0, x) = a_1 \cos \frac{x - \mu}{\sigma_1} - a_2 \cos \frac{x - \mu}{\sigma_2} + p_0 \quad \text{Equation 1}$$

416 Then, as a second step, each main cluster was subdivided into several sub-clusters by

417 applying a KMEANS(63) clustering algorithm on a pairwise similarity matrix. This similarity
418 metric was based on a 2D image cross-correlation scheme, invariant to image shifts, and
419 mirroring (67). The number of sub-clusters was manually chosen by visual inspection. For the
420 most populated 2D classes corresponding to the most compact conformations, the number of
421 sub-clusters was small: n=2 sub-clusters, C11a and C11b, and n=1 sub-cluster for the
422 intermediate class, C12. However, the number of sub-clusters was higher (n=5) for the relaxed
423 conformation (C13 from green to purple in Fig. 1-figure supplement 6), highlighting the
424 overall heterogeneity of our dataset with compact, intermediate, relaxed states and even loss
425 of one protein strands (C13b) and unwound ribbons.

426 *Identification of candidate helical parameters*

427 Fourier transform analysis methods have been used to identify helical parameters candidates
428 (68–70) for the C11a, C12 and C13a clusters. For each cluster, we first estimated the repeat
429 distance by applying the method consisting in a longitudinal autocorrelation with a windowed
430 segment of the real space 2D class of fixed size (100 Å) (69). Then, a precise identification of
431 the power spectrum maxima could be achieved on a high signal-to-noise ratio power
432 spectrum, obtained by averaging all the constituting segments in the Fourier domain, which
433 helps lowering the noise, and fill in the CTF zeros regions. The best candidates were validated
434 with Helixplorer (<http://rico.ibs.fr/helixplorer/>).

435 For the C11a cluster, the parameters of a 1-start helix have been identified with a rise of 7.937
436 Å and a twist of 221.05°. For the C12 cluster, the candidate parameters are a rise of 20.383 Å
437 and a twist of 49.49° and C3 cyclic symmetry. For the C13a cluster, the candidate parameters
438 correspond to a rise of 31.159 Å and a twist of 24° and D5 symmetry (Fig. 1-figure
439 supplement 6C-E).

440 *Cryo-EM data processing and 3D reconstruction*

441- C11a-C13a

442 After helical parameters determination, two last 2D classifications were performed on
443 segments extracted with a box size of 400 pixels (decimated to 200 pixels) using the proper
444 rises for the most compact C11a (7.93 Å, 113,026 segments, overlap ~98.2%) and the relaxed
445 C13a (31.16 Å, 16,831 segments, overlap ~92.9%). Values of the helical parameters (rise and
446 twist) were then used for Relion 3D classification (28, 29), with a +/-10% freedom search
447 range, using a featureless cylinder as initial reference (diameter of 300 Å for the compact
448 particles C11a and 340 Å for the relaxed particles C13a). The superimposable 3D classes
449 (same helical parameters, same helix orientation) were then selected, reducing the data set to
450 95,722 segments for the compact fiber (C11a) and to 15,289 segments for the relaxed fiber
451 (C13a). After re-extraction of the selected segments without scaling, further 3D refinement
452 was performed with a 3D classification output low pass filtered to 15 Å as reference. With
453 this, the maps were resolved enough (C11a: 4.4 Å, C13a: 4.8; FSC threshold 0.5) to identify
454 secondary structure elements (with visible cylinders corresponding to the helices) (Fig. 2-
455 figures supplement 1-2).

456- C12

457 The twelve 2D classes corresponding to segments of the C12 conformation, were extracted
458 with a box size of 400 pixels (decimated to 200 pixels, rise 20.4 Å, 5,775 segments, overlap
459 ~94.9%). They were used in Relion for 3D refinement with the helical parameters values
460 identified previously, with a +/-10% freedom search range, using a 330 Å large featureless
461 cylinder as initial reference and resulted in a 7.1 Å map (FSC threshold 0.5) and C3 cyclic
462 symmetry (Fig. 2-figure supplement 3).

463 *Focused refinement of a single oxidoreductase dimer*

464 The C11a 3D map was used to make a mask corresponding to a single dimer of
465 oxidoreductases through the segmentation module (71) of the program Chimera (72) (this
466 mask was deposited as part of the EMDB D_1292117739). With Relion and this mask we
467 performed partial signal subtraction (73) to remove the information of the other dimers from
468 the experimental images generating a new stack of subtracted images. This was followed by
469 focused refinement with an initial model created (command "relion_reconstruct") from the
470 subtracted dataset and the corresponding orientation/centering parameters from the partial
471 signal subtraction. CTF parameters refinement was performed followed by a last 3D
472 refinement step and post-processing (B-factor applied -45). This led to the best resolved 3.3 Å
473 3D map (FSC threshold 0.5, masked, Fig. 2-figure supplement 1) that was used to build the
474 atomic model of the GMC-oxidoreductase and its ligands (supplementary file 1, Fig. 3). As
475 expected, densities corresponding to DNA were still visible in the final 3D map because the
476 3D density used during partial signal subtraction contained no high-resolution information
477 about the DNA structure. A Relion project directory summarizing all the steps to perform the
478 focused refinement of the asymmetric unit (oxidoreductase dimer) is available under the link:

479 https://src.koda.cnrs.fr/igs/genfiber_c11a_focusrefine_relion_pipeline

480 *DNA focused refinement*

481 The 3D EM maps of C11a and C12 were used to make masks for each DNA strand separately
482 with ChimeraX (74). Each strand was used in a partial signal subtraction removing
483 information from all proteins keeping only information near the presumed DNA regions on
484 the corresponding experimental images. The subtracted (DNA only) datasets were merged
485 and subjected to 2D classification to allow visual assessment of the quality of the subtracted
486 images (Fig. 3-figure supplement 1). 3D refinement was performed followed by 3D
487 classification and final refinement of the best 3D classes (Fig. 3E, Fig. 3-figure supplement
488 2).

489 For the C13a compaction state (where some barely visible densities could correspond to
490 remaining DNA) a cylinder with a diameter corresponding to the lumen of the relaxed
491 structure (140 Å) was used as mask for partial signal subtraction in an attempt to enhance the
492 information from the presumed DNA regions. However, the signal remained too weak to be
493 recognized as such.

494

495 *Automatic picking of the three different conformations (C11a, C12, C13a)*

496 Projections from the different helical maps were used as new input for automatic picking of
497 each C11a, C12 and C13a clusters to get more homogeneous datasets for each conformation
498 (75). For each dataset a final round of extraction (box size C11a: 380 pixels, 121,429
499 segments; C12: 400 pixels, 8479 segments; C13a: 400 pixels, 11,958 segments) and 3D
500 refinement with solvent flattening (Central z length 30%) was performed using the
501 appropriate helical parameters and additional symmetries (none for C11a, C3 for C12 and D5
502 for C13a). This led to the improved maps presented in Figs. 2-3, Fig. 2-figures supplement 1-3
503 (C11a: 3.7 Å; C12: 4.0 Å; C13a: 3.7 Å; FSC threshold 0.5, masked). Post-processing was
504 performed with B-factor -80 (supplementary file 1, Fig. 2-figures supplement 1-3).

505 *Structures refinement*

506 The resolution of the EM map enabled to fit the R135 dimeric structure (30) (PDB 4Z24) into
507 the maps using UCSF Chimera 1.13.1 (72). The qu_143 and qu_946 models were obtained
508 using SWISS-MODEL (76) (closest PDB homologue: 4Z24). It is only at that stage that the

509 best fitted qu_946 model was manually inspected and additional N-terminal residues built
510 using the extra density available in the cryo-EM map of the 5-start compact reconstruction
511 (Fig. 3). The entire protein shell built using the corresponding helical parameters were finally
512 fitted into the C11a and C13a maps and were further refined against the map using the real-
513 space refinement program in PHENIX 1.18.2 (33) (Fig. 2-figures supplement 1-2). The
514 qu_143 and qu_946 models were also used to build the entire shell using the C12 helical
515 parameters and symmetry and were further refined using the real-space refinement program in
516 PHENIX 1.18.2 (33) (Fig. 2-figure supplement 3). Validations were also performed into
517 PHENIX 1.18.2 (33) using the comprehensive validation program and statistics in
518 supplementary file 1 correspond to the qu_143 model. The qu_946 model was manually
519 corrected and ultimately refined and validated into PHENIX 1.18.2 (33) using the highest
520 resolution focused refined C11a map. In that map the 2 first amino acid are disordered
521 including the cysteine.

522 **Mass spectrometry-based proteomic analysis of mimivirus virion and genomic fiber**

523 Proteins extracted from total virions and purified fiber were solubilized with Laemmli buffer
524 (4 volumes of sample with 1 volume of Laemmli 5X - 125 mM Tris-HCl pH 6.8, 10% SDS,
525 20% glycerol, 25% β -mercaptoethanol and traces of bromophenol blue) and heated for 10 min
526 at 95 °C. Three independent infections using 3 different batches of virions were performed
527 and the genomic fiber was extracted from the resulting viral particles to analyze three
528 biological replicates. Extracted proteins were stacked in the top of a SDS-PAGE gel (4-12%
529 NuPAGE, Life Technologies), stained with Coomassie blue R-250 (Bio-Rad) before in-gel
530 digestion using modified trypsin (Promega, sequencing grade) as previously described (77).
531 Resulting peptides were analyzed by online nanoliquid chromatography coupled to tandem
532 MS (UltiMate 3000 RSLCnano and Q-Exactive Plus, Thermo Scientific). Peptides were
533 sampled on a 300 μ m x 5 mm PepMap C18 precolumn and separated on a 75 μ m x 250 mm
534 C18 column (Reprosil-Pur 120 C18-AQ, 1.9 μ m, Dr. Maisch) using a 60-min gradient for
535 fiber preparations and a 140-min gradient for virion. MS and MS/MS data were acquired
536 using Xcalibur (Thermo Scientific). Peptides and proteins were identified using Mascot
537 (version 2.7.0) through concomitant searches against mimivirus reunion, classical
538 contaminant databases (homemade) and the corresponding reversed databases. The Proline
539 software (78) was used to filter the results: conservation of rank 1 peptides, peptide score \geq
540 25, peptide length \geq 6, peptide-spectrum-match identification false discovery rate $<$ 1% as
541 calculated on scores by employing the reverse database strategy, and minimum of 1 specific
542 peptide per identified protein group. Proline was then used to perform a compilation and
543 MS1-based quantification of the identified protein groups. Intensity-based absolute
544 quantification (iBAQ) (79) values were calculated from MS intensities of identified peptides.
545 The viral proteins detected in a minimum of two replicates are reported with their molecular
546 weight, number of identified peptides, sequence coverage and iBAQ values in each replicate.
547 The number of copies per fiber for each protein was calculated according to their iBAQ value
548 based on the GMC-oxidoreductases iBAQ value (see below, supplementary file 2). Mapping
549 of the identified peptides for the GMC-oxidoreductases are presented in Fig. 2-figure
550 supplement 4.

551 **Protein/DNA ratio validation**

552 To compare the theoretical composition of the genomic fiber accommodating a complete
553 genome with the experimental concentrations in protein and DNA of the sample, we
554 performed DNA and protein quantification (Qubit fluorometric quantification, Thermo
555 Fischer Scientific) on two independent samples of the purified genomic fiber. This returned a
556 concentration of 1.5 ng/ μ L for the dsDNA and 24 ng/ μ L for the proteins in one sample and 10

557 ng/ μ L for the dsDNA and 150 ng/ μ L for the protein in the second sample (deposited on
558 agarose gel, Fig. 1-figure supplement 3). Considering that all proteins correspond to the
559 GMC-oxidoreductases subunits (~71 kDa), in the sample there is a molecular ratio of 7.2
560 dsDNA base pairs per GMC-oxidoreductases subunit in the first sample and 7.68 in the
561 second. Based on our model, a 5-start genomic fiber containing the complete mimivirus
562 genome (1,196,989 bp) should be composed of ~95,000 GMC-oxidoreductases subunits
563 (~93,000 for a 6-start). This gives a molecular ratio of 12.5 dsDNA base pairs per GMC-
564 oxidoreductase subunit, which would be more than experimentally measured. However, in the
565 cryo-EM dataset, there are at most 50% of genomic fibers containing the DNA genome
566 (mostly C11), while some DNA strands can be observed attached to the relaxed genomic fiber
567 C13 but the vast majority of released DNA was lost during purification of the genomic fiber.
568 Applying an estimated loss of 50 % to the total DNA compared to the measured values, we
569 obtain a ratio, which is in the same order of magnitude of the ones measured in the purified
570 genomic fiber samples.

571 **Model visualization**

572 Molecular graphics and analyses were performed with UCSF Chimera 1.13.1 (72) and UCSF
573 ChimeraX 1.1 (74), developed by the Resource for Biocomputing, Visualization, and
574 Informatics at the University of California, San Francisco, with support from National
575 Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and
576 Computational Biology, National Institute of Allergy and Infectious Diseases.

577 **Analysis of macromolecules in NS-TEM images of mimivirus genomic fiber**

578 Manually picked positions of the particles of interest have been used to automatically extract
579 193 Å square area from the micrographs (Fig. 4C (E)). Then, manual clipping of the particle
580 from its noisy background has been achieved using GIMP's intelligent scissors and a smooth
581 transparent to black mask, producing the clean extraction image (Fig. 4C (CE)). This clipping
582 step was applied in order to improve the semi-automatic identification of the closest
583 orientation in all RNA polymerase projections. The projection views of the different RNA
584 polymerase were produced by converting the PDB model into a volume density through
585 EMAN2's e2pdb2mrc software, and converting the volume into a 2D projection using a
586 dedicated python script, ultimately applying a gaussian blur filter ($\sigma=2.9\text{\AA}$) in order to
587 roughly simulate the whole imaging transfer function. The projections dataset of all
588 orientations of the vaccinia virus RNA polymerase structure was produced with a 5° rotation
589 step in all angles (PDB: 6RIC, equivalent subunits identified by our MS-based proteomics of
590 purified mimivirus genomic fibers). Preferred projections (Fig. 4C (P)) orientations were
591 manually assessed.

592

593 **Acknowledgements**

594 The cryo-EM work was performed at the multi-user Cryo-EM Facility at CSSB. We thank
595 Jean-Michel Claverie for his comments on the manuscript and discussions all along the
596 project. We thank Irina Gutsche, Ambroise Desfosses, Eric Durand and Juan Reguera for their
597 helpful support on structural work. We thank Carolin Seuring for support and technical help.
598 Processing was performed on the DESY Maxwell cluster, at IBS and IGS. We thank
599 Wolfgang Lugmayr for assistance and Sebastien Santini and Guy Schoehn for support. The
600 preliminary electron microscopy experiments were performed on the PiCSL-FBI core facility
601 (Nicolas Brouilly, Fabrice Richard and Aïcha Aouane, IBDM, AMU-Marseille), member of
602 the France-BioImaging national research infrastructure and on the IMM imaging platform
603 (Artemis Kosta).

604

605 **Competing interests:** Authors declare that they have no competing interests

606

607 **References**

- 608 1. B. La Scola, S. Audic, C. Robert, L. Jungang, X. de Lamballerie, M. Drancourt, R. Birtles, J.-M.
609 Claverie, D. Raoult, A giant virus in amoebae. *Science*. **299**, 2033 (2003).
- 610 2. D. Raoult, S. Audic, C. Robert, C. Abergel, P. Renesto, H. Ogata, B. La Scola, M. Suzan, J.-M.
611 Claverie, The 1.2-megabase genome sequence of Mimivirus. *Science*. **306**, 1344–1350 (2004).
- 612 3. C. Abergel, M. Legendre, J.-M. Claverie, The rapidly expanding universe of giant viruses:
613 Mimivirus, Pandoravirus, Pithovirus and Mollivirus. *FEMS Microbiol. Rev.* **39**, 779–796 (2015).
- 614 4. F. Schulz, S. Roux, D. Paez-Espino, S. Jungbluth, D. A. Walsh, V. J. Denef, K. D. McMahon, K. T.
615 Konstantinidis, E. A. Eloe-Fadrosh, N. C. Kyrpides, T. Woyke, Giant virus diversity and host
616 interactions through global metagenomics. *Nature*. **578**, 432–436 (2020).
- 617 5. M. Moniruzzaman, A. R. Weinheimer, C. A. Martinez-Gutierrez, F. O. Aylward, Widespread
618 endogenization of giant viruses shapes genomes of green algae. *Nature*. **588**, 141–145 (2020).
- 619 6. H. Kaneko, R. Blanc-Mathieu, H. Endo, S. Chaffron, T. O. Delmont, M. Gaia, N. Henry, R.
620 Hernández-Velázquez, C. H. Nguyen, H. Mamitsuka, P. Forterre, O. Jaillon, C. de Vargas, M. B.
621 Sullivan, C. A. Suttle, L. Guidi, H. Ogata, Eukaryotic virus composition can predict the efficiency
622 of carbon export in the global ocean. *iScience*. **24**, 102002 (2021).
- 623 7. P. Forterre, Giant viruses: conflicts in revisiting the virus concept. *Intervirology*. **53**, 362–378
624 (2010).
- 625 8. J.-M. Claverie, C. Abergel, Mimivirus: the emerging paradox of quasi-autonomous viruses.
626 *Trends Genet.* **26**, 431–437 (2010).
- 627 9. C. Abergel, J. Rudinger-Thirion, R. Giegé, J.-M. Claverie, Virus-encoded aminoacyl-tRNA
628 synthetases: structural and functional characterization of mimivirus TyrRS and MetRS. *J. Virol.*
629 **81**, 12406–12417 (2007).
- 630 10. F. Piacente, C. De Castro, S. Jeudy, M. Gaglianone, M. E. Laugieri, A. Notaro, A. Salis, G.
631 Damonte, C. Abergel, M. G. Tonetti, The rare sugar N-acetylated viosamine is a major
632 component of Mimivirus fibers. *J. Biol. Chem.* **292**, 7385–7394 (2017).
- 633 11. A. Notaro, Y. Couté, L. Belmudes, M. E. Laugeri, A. Salis, G. Damonte, A. Molinaro, M. G.
634 Tonetti, C. Abergel, C. De Castro, Expanding the Occurrence of Polysaccharides to the Viral
635 World: The Case of Mimivirus. *Angew. Chem. Int. Ed.* **60**, 19897–19904 (2021).
- 636 12. P. Colson, B. La Scola, A. Levasseur, G. Caetano-Anollés, D. Raoult, Mimivirus: leading the way in
637 the discovery of giant viruses of amoebae. *Nat. Rev. Microbiol.* **15**, 243–254 (2017).
- 638 13. Y. G. Kuznetsov, T. Klose, M. Rossmann, A. McPherson, Morphogenesis of mimivirus and its
639 viral factories: an atomic force microscopy study of infected cells. *J. Virol.* **87**, 11200–11213
640 (2013).

- 641 14. C. Xiao, Y. G. Kuznetsov, S. Sun, S. L. Hafenstein, V. A. Kostyuchenko, P. R. Chipman, M. Suzan-
642 Monti, D. Raoult, A. McPherson, M. G. Rossmann, Structural Studies of the Giant Mimivirus.
643 *PLoS Biol.* **7**, e1000092 (2009).
- 644 15. T. Klose, Y. G. Kuznetsov, C. Xiao, S. Sun, A. McPherson, M. G. Rossmann, The three-
645 dimensional structure of Mimivirus. *Intervirology.* **53**, 268–273 (2010).
- 646 16. T. Ekeberg, M. Svenda, C. Abergel, F. R. N. C. Maia, V. Seltzer, J.-M. Claverie, M. Hantke, O.
647 Jönsson, C. Nettelblad, G. van der Schot, M. Liang, D. P. DePonte, A. Barty, M. M. Seibert, B.
648 Iwan, I. Andersson, N. D. Loh, A. V. Martin, H. Chapman, C. Bostedt, J. D. Bozek, K. R. Ferguson,
649 J. Krzywinski, S. W. Epp, D. Rolles, A. Rudenko, R. Hartmann, N. Kimmel, J. Hajdu, Three-
650 dimensional reconstruction of the giant mimivirus particle with an x-ray free-electron laser.
651 *Phys. Rev. Lett.* **114**, 098102 (2015).
- 652 17. J. R. Schrad, J. S. Abrahão, J. R. Cortines, K. N. Parent, Structural and Proteomic Characterization
653 of the Initiation of Giant Virus Infection. *Cell.* **181**, 1046-1061.e6 (2020).
- 654 18. V. Chelikani, T. Ranjan, K. Kondabagil, Revisiting the genome packaging in viruses with lessons
655 from the “Giants.” *Virology.* **466–467**, 15–26 (2014).
- 656 19. Y. G. Kuznetsov, C. Xiao, S. Sun, D. Raoult, M. Rossmann, A. McPherson, Atomic force
657 microscopy investigation of the giant mimivirus. *Virology.* **404**, 127–137 (2010).
- 658 20. J.-M. Claverie, R. Grzela, A. Lartigue, A. Bernadac, S. Nitsche, J. Vacelet, H. Ogata, C. Abergel,
659 Mimivirus and Mimiviridae: giant viruses with an increasing number of potential hosts,
660 including corals and sponges. *J. Invertebr. Pathol.* **101**, 172–180 (2009).
- 661 21. D. Arslan, M. Legendre, V. Seltzer, C. Abergel, J.-M. Claverie, Distant Mimivirus relative with a
662 larger genome highlights the fundamental features of Megaviridae. *Proc. Natl. Acad. Sci. U. S.*
663 *A.* **108**, 17486–17491 (2011).
- 664 22. N. Zauberman, Y. Mutsafi, D. B. Halevy, E. Shimoni, E. Klein, C. Xiao, S. Sun, A. Minsky, Distinct
665 DNA exit and packaging portals in the virus *Acanthamoeba polyphaga* mimivirus. *PLoS Biol.* **6**,
666 e114 (2008).
- 667 23. M. Suzan-Monti, B. L. Scola, L. Barrassi, L. Espinosa, D. Raoult, Ultrastructural Characterization
668 of the Giant Volcano-like Virus Factory of *Acanthamoeba polyphaga* Mimivirus. *PLoS ONE.* **2**,
669 e328 (2007).
- 670 24. Y. Mutsafi, Y. Fridmann-Sirkis, E. Milrot, L. Hevroni, A. Minsky, Infection cycles of large DNA
671 viruses: emerging themes and underlying questions. *Virology.* **466–467**, 3–14 (2014).
- 672 25. W. Wu, J. A. Thomas, N. Cheng, L. W. Black, A. C. Steven, Bubblegrams Reveal the Inner Body of
673 Bacteriophage KZ. *Science.* **335**, 182–182 (2012).
- 674 26. N. Cheng, W. Wu, N. R. Watts, A. C. Steven, Exploiting radiation damage to map proteins in
675 nucleoprotein complexes: The internal structure of bacteriophage T7. *J. Struct. Biol.* **185**, 250–
676 256 (2014).
- 677 27. M. Mishyna, O. Volokh, Ya. Danilova, N. Gerasimova, E. Pechnikova, O. S. Sokolova, Effects of
678 radiation damage in studies of protein-DNA complexes by cryo-EM. *Micron.* **96**, 57–64 (2017).
- 679 28. S. He, S. H. W. Scheres, Helical reconstruction in RELION. *J. Struct. Biol.* **198**, 163–176 (2017).

- 680 29. S. H. W. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure
681 determination. *J. Struct. Biol.* **180**, 519–530 (2012).
- 682 30. T. Klose, D. A. Herbst, H. Zhu, J. P. Max, H. I. Kenttämaa, M. G. Rossmann, A Mimivirus Enzyme
683 that Participates in Viral Entry. *Struct. Lond. Engl.* **1993**, 23, 1058–1065 (2015).
- 684 31. M. Boyer, S. Azza, L. Barrassi, T. Klose, A. Campocasso, I. Pagnier, G. Fournous, A. Borg, C.
685 Robert, X. Zhang, C. Desnues, B. Henrissat, M. G. Rossmann, B. La Scola, D. Raoult, Mimivirus
686 shows dramatic genome reduction after intraamoebal culture. *Proc. Natl. Acad. Sci. U. S. A.*
687 **108**, 10296–10301 (2011).
- 688 32. A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P.
689 de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: homology modelling of
690 protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
- 691 33. D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L.-W.
692 Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J.
693 S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger,
694 A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, Macromolecular structure
695 determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta*
696 *Crystallogr. Sect. Struct. Biol.* **75**, 861–877 (2019).
- 697 34. S. Li, W. K. Olson, X.-J. Lu, Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D
698 nucleic acid structures. *Nucleic Acids Res.* **47**, W26–W34 (2019).
- 699 35. D. Ptchelkine, A. Gillum, T. Mochizuki, S. Lucas-Staat, Y. Liu, M. Krupovic, S. E. V. Phillips, D.
700 Prangishvili, J. T. Huiskonen, Unique architecture of thermophilic archaeal virus APBV1 and its
701 genome packaging. *Nat. Commun.* **8**, 1436 (2017).
- 702 36. C. Grimm, H. S. Hillen, K. Bedenk, J. Bartuli, S. Neyer, Q. Zhang, A. Hüttenhofer, M. Erlacher, C.
703 Dienemann, A. Schlosser, H. Urlaub, B. Böttcher, A. A. Szalay, P. Cramer, U. Fischer, Structural
704 Basis of Poxvirus Transcription: Vaccinia RNA Polymerase Complexes. *Cell.* **179**, 1537–1550.e19
705 (2019).
- 706 37. F. DiMaio, X. Yu, E. Rensen, M. Krupovic, D. Prangishvili, E. H. Egelman, Virology. A virus that
707 infects a hyperthermophile encapsidates A-form DNA. *Science.* **348**, 914–917 (2015).
- 708 38. F. Wang, D. P. Baquero, L. C. Beltran, Z. Su, T. Osinski, W. Zheng, D. Prangishvili, M. Krupovic, E.
709 H. Egelman, Structures of filamentous viruses infecting hyperthermophilic archaea explain DNA
710 stabilization in extreme environments. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 19643–19652 (2020).
- 711 39. P. J. J. Robinson, L. Fairall, V. A. T. Huynh, D. Rhodes, EM measurements define the dimensions
712 of the “30-nm” chromatin fiber: evidence for a compact, interdigitated structure. *Proc. Natl.*
713 *Acad. Sci. U. S. A.* **103**, 6506–6511 (2006).
- 714 40. Y. Liu, H. Bisio, C. M. Toner, S. Jeudy, N. Philippe, K. Zhou, S. Bowerman, A. White, G. Edwards,
715 C. Abergel, K. Luger, Virus-encoded histone doublets are essential and form nucleosome-like
716 structures. *Cell.* **184**, 4237–4250.e19 (2021).
- 717 41. M. I. Valencia-Sánchez, S. Abini-Agbomson, M. Wang, R. Lee, N. Vasilyev, J. Zhang, P. De
718 loannes, B. La Scola, P. Talbert, S. Henikoff, E. Nudler, A. Erives, K.-J. Armache, The structure of
719 a virus-encoded nucleosome. *Nat. Struct. Mol. Biol.* **28**, 413–417 (2021).

- 720 42. P. J. Bell, Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? *J. Mol.*
721 *Evol.* **53**, 251–256 (2001).
- 722 43. P. J. L. Bell, Evidence supporting a viral origin of the eukaryotic nucleus. *Virus Res.* **289**, 198168
723 (2020).
- 724 44. V. Chaikerasitak, K. Nguyen, K. Khanna, A. F. Brilot, M. L. Erb, J. K. C. Coker, A. Vavilina, G. L.
725 Newton, R. Buschauer, K. Pogliano, E. Villa, D. A. Agard, J. Pogliano, Assembly of a nucleus-like
726 structure during viral replication in bacteria. *Science.* **355**, 194–197 (2017).
- 727 45. J.-M. Claverie, Viruses take center stage in cellular evolution. *Genome Biol.* **7**, 110 (2006).
- 728 46. M. Takemura, Poxviruses and the origin of the eukaryotic nucleus. *J. Mol. Evol.* **52**, 419–425
729 (2001).
- 730 47. D. Gong, X. Dai, J. Jih, Y.-T. Liu, G.-Q. Bi, R. Sun, Z. H. Zhou, DNA-Packing Portal and Capsid-
731 Associated Tegument Complexes in the Tumor Herpesvirus KSHV. *Cell.* **178**, 1329-1343.e12
732 (2019).
- 733 48. Y.-T. Liu, J. Jih, X. Dai, G.-Q. Bi, Z. H. Zhou, Cryo-EM structures of herpes simplex virus type 1
734 portal vertex and packaged genome. *Nature.* **570**, 257–261 (2019).
- 735 49. L. Sun, X. Zhang, S. Gao, P. A. Rao, V. Padilla-Sanchez, Z. Chen, S. Sun, Y. Xiang, S. Subramaniam,
736 V. B. Rao, M. G. Rossmann, Cryo-EM structure of the bacteriophage T4 portal protein assembly
737 at near-atomic resolution. *Nat. Commun.* **6**, 7548 (2015).
- 738 50. V. B. Rao, M. Feiss, Mechanisms of DNA Packaging by Large Double-Stranded DNA Viruses.
739 *Annu. Rev. Virol.* **2**, 351–378 (2015).
- 740 51. K. Suhre, S. Audic, J.-M. Claverie, Mimivirus gene promoters exhibit an unprecedented
741 conservation among all eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 14689–14693 (2005).
- 742 52. S. Toriyama, An RNA-dependent RNA Polymerase Associated with the Filamentous
743 Nucleoproteins of Rice Stripe Virus. *J. Gen. Virol.* **67**, 1247–1255 (1986).
- 744 53. A. M. Collier, O. L. Lyytinen, Y. R. Guo, Y. Toh, M. M. Poranen, Y. J. Tao, Initiation of RNA
745 Polymerization and Polymerase Encapsidation by a Small dsRNA Virus. *PLOS Pathog.* **12**,
746 e1005523 (2016).
- 747 54. K. Ding, C. C. Celma, X. Zhang, T. Chang, W. Shen, I. Atanasov, P. Roy, Z. H. Zhou, In situ
748 structures of rotavirus polymerase in action and mechanism of mRNA transcription and release.
749 *Nat. Commun.* **10**, 2216 (2019).
- 750 55. A. J. Malkin, A. McPherson, P. D. Gershon, Structure of Intracellular Mature Vaccinia Virus
751 Visualized by In Situ Atomic Force Microscopy. *J. Virol.* **77**, 6332–6340 (2003).
- 752 56. R. Blanc-Mathieu, H. Dahle, A. Hofgaard, D. Brandt, H. Ban, J. Kalinowski, H. Ogata, R.-A.
753 Sandaa, A Persistent Giant Algal Virus, with a Unique Morphology, Encodes an Unprecedented
754 Number of Genes Involved in Energy Metabolism. *J. Virol.* **95**, e02446-20 (2021).
- 755 57. D. N. Mastrorade, Automated electron microscope tomography using robust prediction of
756 specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

- 757 58. W. J. H. Hagen, W. Wan, J. A. G. Briggs, Implementation of a cryo-electron tomography tilt-
758 scheme optimized for high resolution subtomogram averaging. *J. Struct. Biol.* **197**, 191–198
759 (2017).
- 760 59. J. R. Kremer, D. N. Mastronarde, J. R. McIntosh, Computer visualization of three-dimensional
761 image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).
- 762 60. J. B. Heymann, D. M. Belnap, Bsoft: image processing and molecular modeling for electron
763 microscopy. *J. Struct. Biol.* **157**, 3–18 (2007).
- 764 61. S. Q. Zheng, E. Palovcak, J.-P. Armache, K. A. Verba, Y. Cheng, D. A. Agard, MotionCor2:
765 anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat.*
766 *Methods.* **14**, 331–332 (2017).
- 767 62. A. Rohou, N. Grigorieff, CTFIND4: Fast and accurate defocus estimation from electron
768 micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
- 769 63. C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E.
770 Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett,
771 A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W.
772 Weckesser, H. Abbasi, C. Gohlke, T. E. Oliphant, Array programming with NumPy. *Nature.* **585**,
773 357–362 (2020).
- 774 64. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P.
775 Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M.
776 Perrot, É. Duchesnay, Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–
777 2830 (2011).
- 778 65. M. Hahsler, M. Piekenbrock, D. Doran, **dbscan** : Fast Density-Based Clustering with R. *J. Stat.*
779 *Softw.* **91** (2019), doi:10.18637/jss.v091.i01.
- 780 66. S. Mannor, X. Jin, J. Han, X. Jin, J. Han, X. Jin, J. Han, X. Zhang, "K-Means Clustering" in
781 *Encyclopedia of Machine Learning*, C. Sammut, G. I. Webb, Eds. (Springer US, Boston, MA, 2011;
782 http://link.springer.com/10.1007/978-0-387-30164-8_425), pp. 563–564.
- 783 67. M. Guizar-Sicairos, S. T. Thurman, J. R. Fienup, Efficient subpixel image registration algorithms.
784 *Opt. Lett.* **33**, 156 (2008).
- 785 68. N. Coudray, R. Lasala, Z. Zhang, K. M. Clark, M. E. Dumont, D. L. Stokes, Deducing the symmetry
786 of helical assemblies: Applications to membrane proteins. *J. Struct. Biol.* **195**, 167–178 (2016).
- 787 69. R. Diaz, W. J. Rice, D. L. Stokes, Fourier-Bessel reconstruction of helical assemblies. *Methods*
788 *Enzymol.* **482**, 131–165 (2010).
- 789 70. C. Sachse, Single-particle based helical reconstruction—how to make the most of real and
790 Fourier space. *AIMS Biophys.* **2**, 219–244 (2015).
- 791 71. G. D. Pintilie, J. Zhang, T. D. Goddard, W. Chiu, D. C. Gossard, Quantitative analysis of cryo-EM
792 density map segmentation by watershed and scale-space filtering, and fitting of structures by
793 alignment to regions. *J. Struct. Biol.* **170**, 427–438 (2010).

- 794 72. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E.
795 Ferrin, UCSF Chimera-A visualization system for exploratory research and analysis. *J. Comput.*
796 *Chem.* **25**, 1605–1612 (2004).
- 797 73. X. Bai, E. Rajendra, G. Yang, Y. Shi, S. H. Scheres, Sampling the conformational space of the
798 catalytic subunit of human γ -secretase. *eLife*. **4**, e11182 (2015).
- 799 74. T. D. Goddard, C. C. Huang, E. C. Meng, E. F. Pettersen, G. S. Couch, J. H. Morris, T. E. Ferrin,
800 UCSF ChimeraX: Meeting modern challenges in visualization and analysis: UCSF ChimeraX
801 Visualization System. *Protein Sci.* **27**, 14–25 (2018).
- 802 75. L. F. Estrozi, J. Navaza, Fast projection matching for cryo-electron microscopy image
803 reconstruction. *J. Struct. Biol.* **162**, 324–334 (2008).
- 804 76. A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P.
805 de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: homology modelling of
806 protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
- 807 77. M. G. Casabona, Y. Vandenbrouck, I. Attree, Y. Couté, Proteomic characterization of
808 *Pseudomonas aeruginosa* PAO1 inner membrane. *Proteomics*. **13**, 2419–2423 (2013).
- 809 78. D. Bouyssié, A.-M. Hesse, E. Mouton-Barbosa, M. Rompais, C. Macron, C. Carapito, A. Gonzalez
810 de Peredo, Y. Couté, V. Dupierris, A. Burel, J.-P. Menetrey, A. Kalaitzakis, J. Poisat, A. Romdhani,
811 O. Burlet-Schiltz, S. Cianférani, J. Garin, C. Bruley, Proline: an efficient and user-friendly
812 software suite for large-scale proteomics. *Bioinforma. Oxf. Engl.* **36**, 3148–3155 (2020).
- 813 79. B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, M. Selbach,
814 Global quantification of mammalian gene expression control. *Nature*. **473**, 337–342 (2011).

815

816 **Funding:**

817 This project has received funding from the European Research Council (ERC) under the
818 European Union’s Horizon 2020 research and innovation program (grant agreement No
819 832601).

820 This work was also partially supported by the French National Research Agency ANR-16-
821 CE11-0033-01.

822 Proteomic experiments were partly supported by ProFI (ANR-10-INBS-08-01) and GRAL, a
823 program from the Chemistry Biology Health (CBH) Graduate School of University Grenoble
824 Alpes (ANR-17-EURE-0003).

825 Cryo-EM data collection was supported by DFG grants (INST 152/772-1|152/774-1|152/775-
826 1|152/776-1).

827 Work in the laboratory of Kay Grunewald is funded by the Wellcome Trust (107806/Z/15/Z),
828 the Leibniz Society, the City of Hamburg and the BMBF (05K18BHA).

829 Emmanuelle Quemini received support from the Alexander von Humboldt foundation to
830 (individual research fellowship No. FRA 1200789 HFST-P).

831 France-BioImaging national research infrastructure (ANR-10-INBS-04).

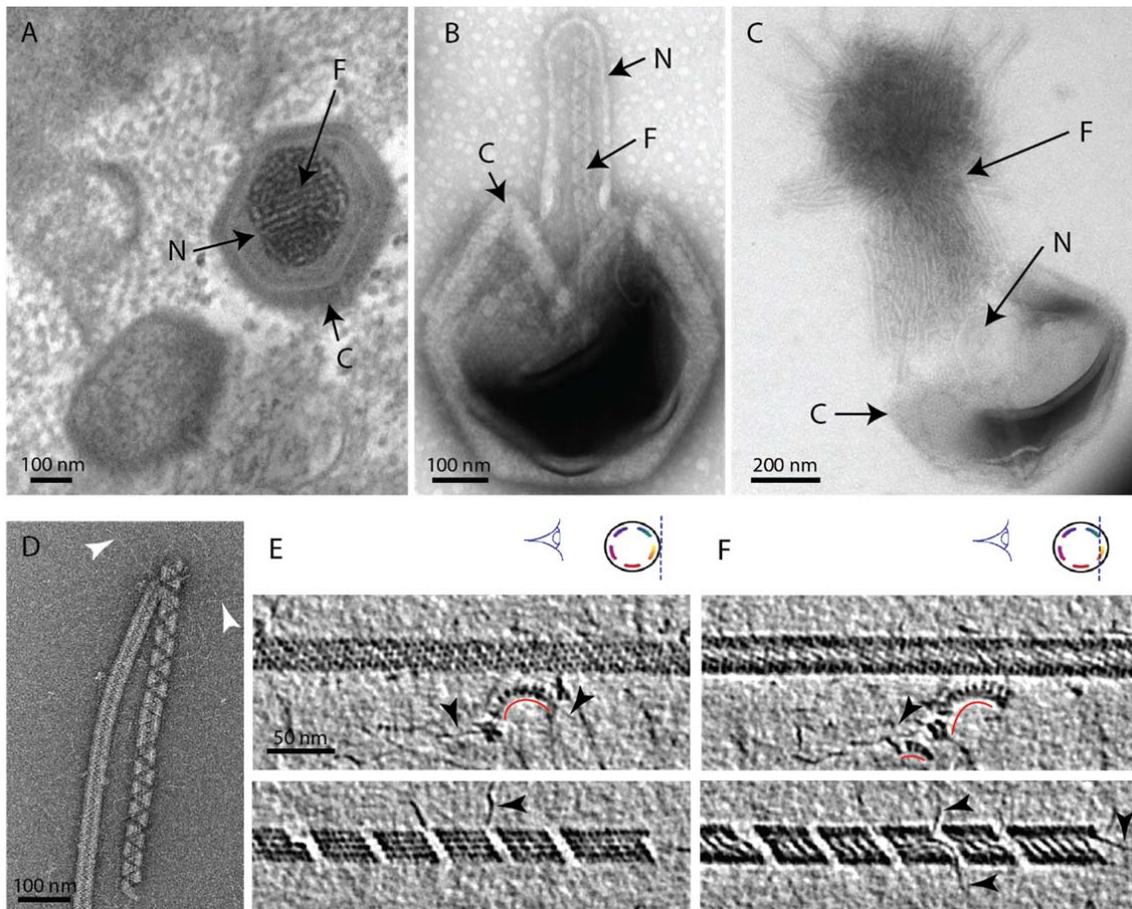
832

833 **Data and materials availability:** Mimivirus reunion genome has been deposited under the
 834 following accession number: BankIt2382307 Seq1 MW004169. 3D reconstruction maps and
 835 the corresponding PDB have been deposited to EMDB (Deposition number C11a: 7YX4,
 836 EMD-14354; C11a focused refined: 7PTV, EMD-13641; C13a: 7YX5, EMD-14355; C12:
 837 7YX3, EMD-14353). The mass spectrometry proteomics data have been deposited to the
 838 ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier
 839 PXD021585 and 10.6019/PXD021585. The tomograms have been deposited in EMPIAR
 840 under the accession number 1131.

841

842 **Figures**

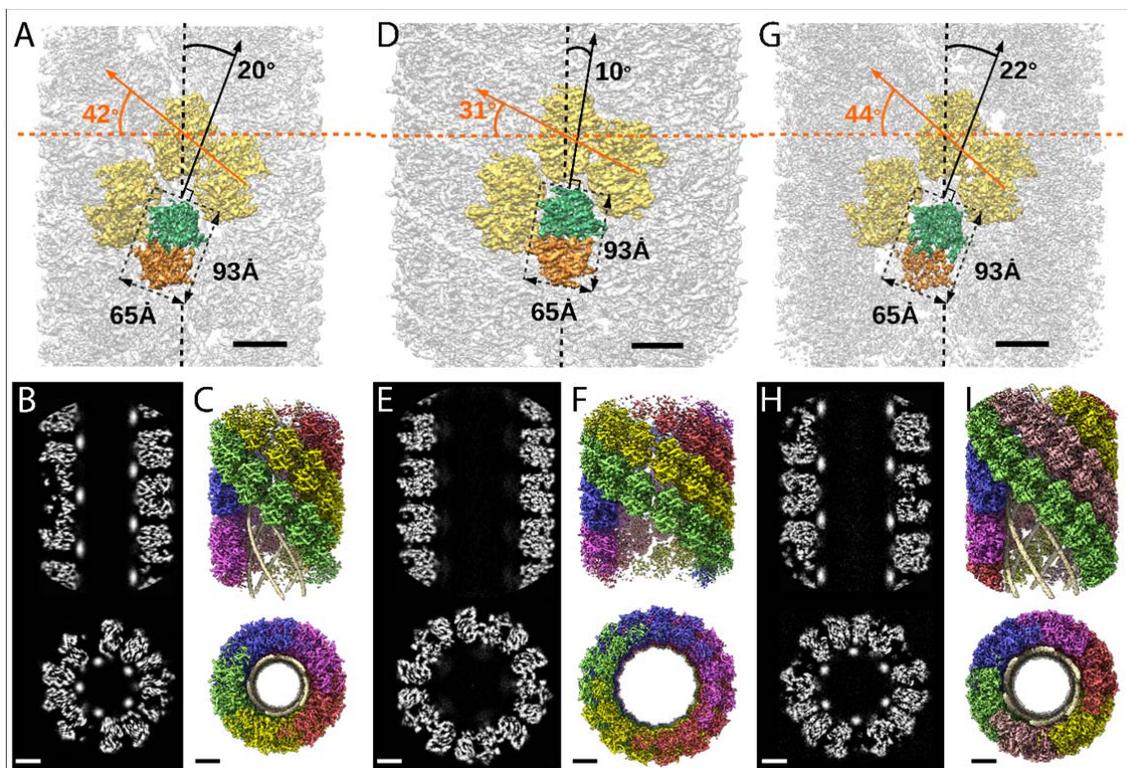
843



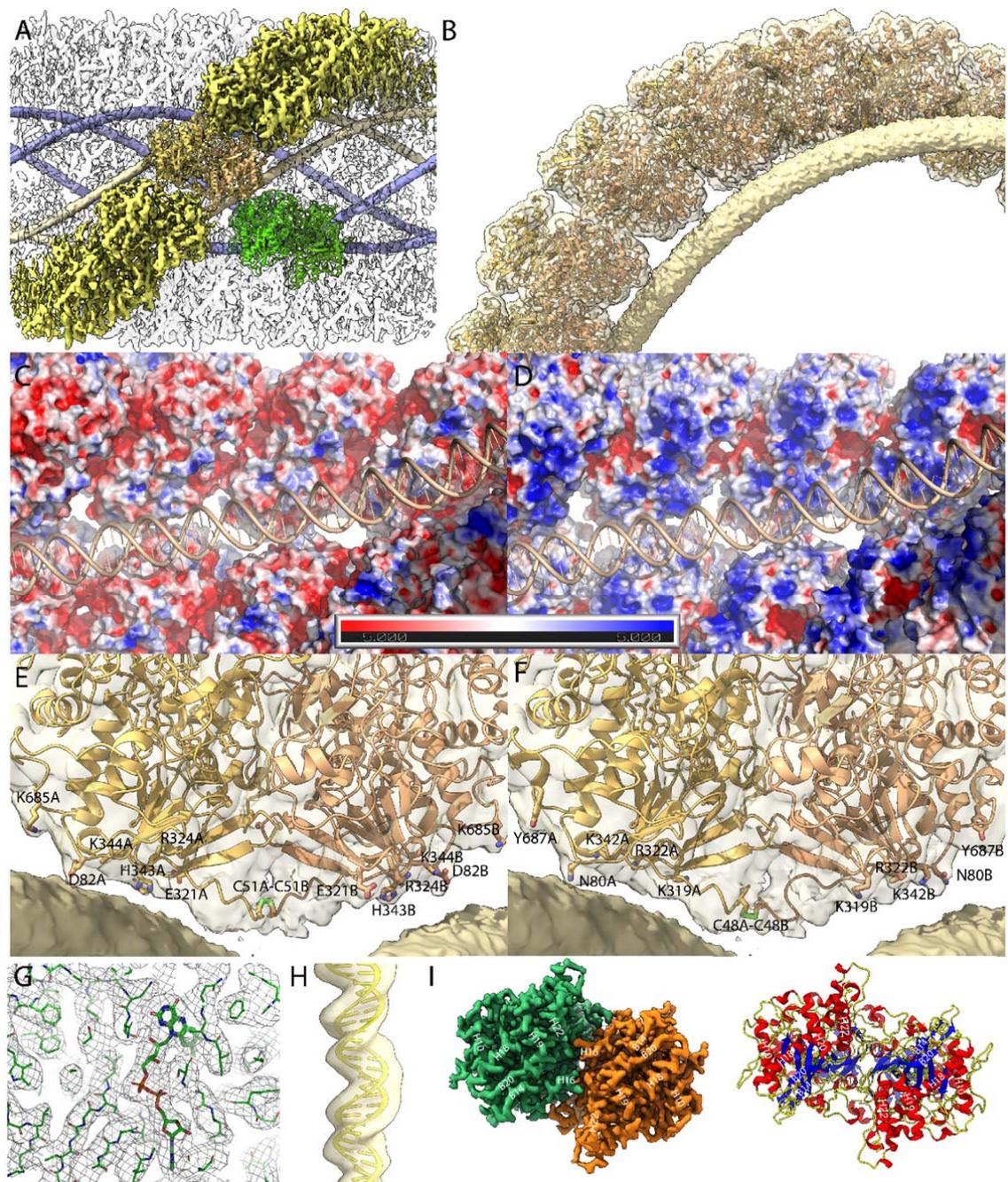
844

845 **Figure 1: The mimivirus genomic fiber** A] Micrograph of an ultrathin section of resin-
 846 embedded infected cells showing the DNA tightly packed inside mimivirus capsids (C) with
 847 electron dense material inside the nucleoid (N). The string-like features, most likely enhanced
 848 by the dehydration caused by the fixation and embedding protocol, correspond to the genomic
 849 fiber (F) packed into the nucleoid. B] Micrograph of negative stained mimivirus capsid (C)
 850 opened *in vitro* with the genomic fiber (F) still being encased into the membrane limited
 851 nucleoid (N). C] Multiple strands of the flexible genomic fiber (F) are released from the
 852 capsid (C) upon proteolytic treatment. D] Micrograph of negative stained purified mimivirus
 853 genomic fibers showing two conformations (the right fiber resembling the one in [B] and free
 854 DNA strands (white arrowheads). E] Slices through two electron cryo tomograms of the
 855 isolated helical protein shell of the purified genomic fibers in compact (top) or relaxed

856 conformation (bottom) in the process of losing one protein strand (Fig. 1-figure supplement 2,
 857 Fig.1-video 1 & 4). F] Different slices through the two tomograms shown in [E] reveal DNA
 858 strands lining the helical protein shell of the purified genomic fibers in compact (top) or
 859 relaxed conformation (bottom). Examples of DNA strands extending out at the breaking
 860 points of the genomic fiber are marked by black arrowheads. Note in the top panel, individual
 861 DNA strands coated by proteins (red arc). The slicing planes at which the mimivirus genomic
 862 fibers were viewed are indicated on diagrams on the top right corner as blue dashed lines and
 863 the internal colored segments correspond to DNA strands lining the protein shell. The
 864 thickness of the tomographic slices is 1.1 nm and the distance between tomographic slices in
 865 panels [E] and [F] is 4.4 nm. Scale bars as indicated.

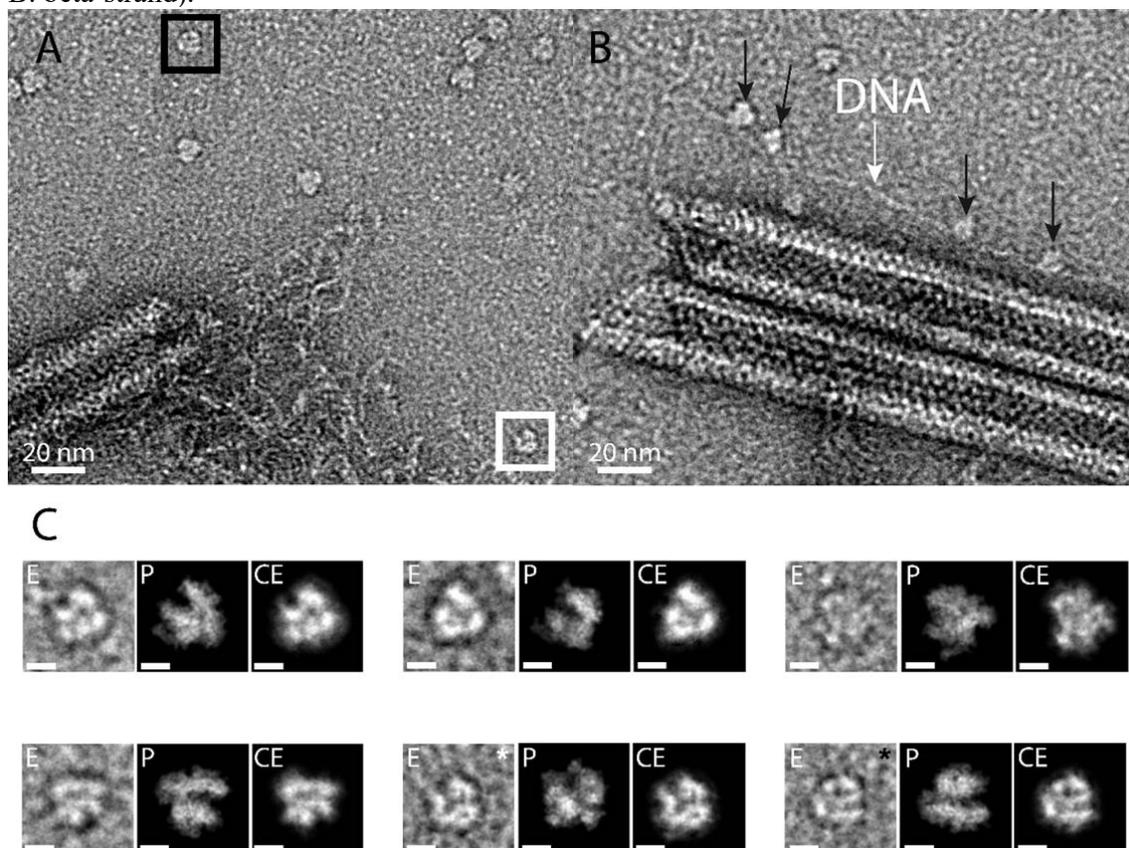


866 **Figure 2: Structures of the mimivirus genomic fiber for C11a [A-C], C13a [D-F], and C12**
 867 **[G-I]:** A, D, G] Electron microscopy (EM) maps of C11a (A) and C13a (D) are shown with
 868 each monomer of one GMC-oxidoreductase dimer colored in green and orange and 3 adjacent
 869 dimers in yellow, to indicate the large conformational change taking place between the two
 870 fiber states. The transition from C11a to C13a (5-start helix) corresponds to a rotation of each
 871 individual unit (corresponding to a GMC- oxidoreductase dimer) by $\sim -10^\circ$ relative to the
 872 fiber longitudinal axis and a change in the steepness of the helical rise by $\sim -11^\circ$. Scale bars
 873 50 Å. Compared to C11a, the C12 6-start helix (G) shows a difference of $\sim 2^\circ$ relative to the
 874 fiber longitudinal axis and $\sim 2^\circ$ in the steepness of the helical rise. Scale bars 50 Å. B, E, H]
 875 Cross-sectional (bottom) and longitudinal (top) sections through the middle of final C11a (B)
 876 C13a (E) and C12 (H) EM maps. Scale bars 50 Å. C, F, I] Longitudinal (top) and orthogonal
 877 (bottom) views of final C11a (C) and C13a (F) and C12 (I) EM maps color-coded according to
 878 each start of the 5-start helix. Densities for some asymmetric units in the front have been
 879 removed on the side view map to show the 5 DNA strands lining the protein shell interior.
 880 Scale bars 50 Å.
 881



882
 883 **Figure 3: Maps of the compact (C11a and C12) genomic fiber structures:** A] C11a EM map
 884 corresponding to the protein shell prior focused refinement is shown as a transparent surface
 885 and the 5 DNA strands as solid surface. One protein dimer strand is shown yellow except for
 886 one asymmetric unit (transparent yellow) to illustrate the dimer fit. The position of a second
 887 dimer (green) from the adjacent dimer strand is shown to emphasize that the DNA strand
 888 (gold) is lining the interface between two dimers. B] Cartoon representation of GMC-
 889 oxidoreductase qu_946 dimers fitted into one of the C11a 5-start helix strands in the 3.7 Å
 890 resolution map. The map is shown at a threshold highlighting the periodicity of contacts
 891 between the dsDNA and the protein shell. C, D] Charged distribution on surface
 892 representation of C11a protein shell made of qu_946 dimers (C) or qu_143 dimers (D). E, F]
 893 Cartoon representation of qu_946 (E) and qu_143 (F) fitted into the C11a cryo-EM maps

894 highlighting the interacting residues (given as stick models) between each monomer and one
 895 dsDNA strand. The isosurface threshold chosen allows visualization of density for the
 896 manually built N-terminal residues, including terminal cysteines (stick model), of two
 897 neighboring monomers that could form a terminal disulfide bridge. G] Zoom into the 3.3 Å
 898 resolution focused refined C11a map illustrating the fit of the side chains and the FAD ligand
 899 (Fig. 3-video 1). H] Cartoon representation of the DNA fitted in the focused refined DNA
 900 only map (Fig. 3-figure supplement 2). I] Focused refined C11a map colored by monomer,
 901 next to a cartoon representation of the qu_946 dimer (α -helices in red, β -strands in blue and
 902 coils in yellow). Secondary structure elements are annotated in both representations (H: helix,
 903 B: beta-strand).



904
 905 **Figure 4: RNA polymerase could be associated to the genomic fiber.** A] Micrograph of
 906 negative stained fiber with released DNA still being connected to a relaxed and broken fiber
 907 and adjacent scattered macromolecular complexes that might resemble RNA polymerases. B]
 908 Strikingly, some of them (black arrows) appear to sit on a DNA strand (white arrow). C] E,
 909 particle extracted from the NS-TEM image, P, projections of vaccinia virus RNA polymerase
 910 (6RIC and 6RUI (36)) structure in preferred orientation, CE, clean extraction (see Material
 911 and methods section). White and black boxed corresponds to images with white and black
 912 asterisk (CE), respectively. Scale bar 50 Å. Negative staining imaging may dehydrate the
 913 objects and change macromolecules volumes.

914

915 Supplementary Figures legend

916 **Fig. 1-figure supplement 1: Negative staining micrographs of opened Mimivirus reunion**
 917 **virions before purification of the genomic fiber.** (A) Spherical nucleoids (N) released from

918 the capsids upon specific treatment (see methods). **(B)** Dissolved nucleoids expelled from the
919 opened capsids after regular treatment (the region marked with dashed lines is enlarged in Fig
920 1C), all of the opened capsids (>25%) are releasing the condensed genomic fiber. **(C)**
921 Multiple broken genomic fibers can be seen next to debris prior to purification and an isolated
922 Stargate (S). **(D-E)** Multiple segments of the flexible genomic fiber are released upon
923 proteolytic treatment of the capsids. **(F)** Example of a long, flexible, genomic fiber released
924 from an open capsid (C).

925 **Fig. 1-figure supplement 2: Cryo-ET of Mimivirus genomic fiber.** (Upper panel) Slices
926 through tomograms exhibiting different features of compact, unwinding and broken genomic
927 fibers of Mimivirus. **(A)** Slices through a cryo-tomogram focused on a long and broken
928 fiber. Features of the protein shell are visible in tangential slices (top and bottom), while
929 filamentous densities corresponding to DNA strands become visible when slicing deeper
930 inside the fiber. Scale bar, 100 nm. **(B)** A straight filament running along the fiber symmetry
931 axis has been observed in several cases (blue rectangle), which may correspond to a single or
932 a bundle of DNA strands dissociated from the protein shell. **(C)** Large amorphous electron
933 dense structures have been observed inside the lumen of the fiber (highlighted with purple
934 circles). **(D)** DNA strands are often observed emanating from breaks along the
935 genomic fiber, partially disassembling (indicated with yellow arrowheads). **(E)** Detail of the
936 central slices through the genomic fibers showed in the third depicted slice in **A-D**.
937 Thickness of the slices is 1.1 nm. Distance between the tomographic slices (from top to
938 bottom) is 4.4 nm between the first and second and the fourth and fifth, and 6.6 nm between
939 the third and second or fourth. The scheme on the left represents a cross-section of a cartoon
940 representation of the 5-start genomic fiber depicted as a cylinder containing 5 strands of
941 coloured DNA internally lining the helical protein shell. It also indicates the plane through
942 which the genomic fiber is viewed as 2D slices extracted from the tomograms in Fig.1-video
943 1 **(A)**, Fig.1-video 2 **(B)**, Fig.1-video 3 **(C)**, and Fig.1-video 4 **(D)**. Scale bars **(B, C, D)**, 50
944 nm.

945 **Fig. 1-figure supplement 3: Agarose gel electrophoresis of the purified genomic fiber**
946 **compared to the viral genomic DNA:** M: Molecular weight markers (1 kbp DNA Ladder
947 Plus, Euromedex). Lane 1, 100 ng of Mimivirus genomic DNA, Lane 2: 200 ng of untreated
948 genomic fiber, Lane 3: 200 ng of Proteinase K-(PK) treated genomic fiber, Lane 4: 200 ng of
949 DNase- and PK-treated genomic fiber, Lane 5: 200 ng of RNase- and PK-treated genomic
950 fiber.

951 **Fig. 1-figure supplement 4: Bubblegrams on the Mimivirus genomic fiber.** Field of view
952 for genomic fibers either (from top to bottom): long and bent, long and straight, long and
953 broken, or a mix of short and broken, subjected to a dose series. The total electron flux
954 applied is specified for each field of view. First column: fibers showed after the first exposure
955 of $75 \text{ e}^-/\text{\AA}^2$. Second column: representative micrographs indicating the point at which the first
956 sign of radiation damage as “bubbles” inside the fiber (red arrowheads) were detected.
957 Nucleoproteins with high nucleic acid content are often more susceptible to bubbling than
958 pure proteinaceous structures. In unfolded ribbons (black arrowhead), no bubbles are
959 detected at an accumulated flux of up to $1,125 \text{ e}^-/\text{\AA}^2$. Scale bars, 100 nm. Third column:
960 enlarged view where the effect of radiation damage as trapped “bubbles” is more clearly
961 visible.

962 **Fig. 1-figure supplement 5.** 200 2D classes (6 empty) were obtained after reference-free 2D
963 classification of fibers acquired (see methods) for single-particle analysis and extracted with a
964 box size of 500 pixels in Relion 3.1.0 after motion correction, CTF estimation and manual

965 picking. The 2D classes are representative of the different relaxation states of the Mimivirus
966 genomic fiber observed in our highly heterogeneous dataset.

967 **Fig. 1-figure supplement 6: Clustering analysis of the 2D classes.** (A) Cross-section model
968 adjusted to one 2D class, representative of cluster C11b, to estimate the parameters of the
969 widths W1 (external) and W2 (internal). (B) Automatic sorting of the 2D classes using the
970 fiber width W1 and pairwise correlations of the 2D classes resulting in 3 main clusters
971 (compact, C11 in dark blue; intermediate, C12 in cyan and relaxed, C13 in green and orange).
972 **Power spectra indexing with Helixplorer** (<http://rico.ibs.fr/helixplorer/>) for the candidate
973 helical parameters estimation. (C) C11a: 5-start, cyclic symmetry C1
974 (<https://tinyurl.com/mrxnyjwn>), (D) C12: 6-start, cyclic symmetry C3
975 (<https://tinyurl.com/yckmk79b>) et (E) C13a: 5-start, dihedral symmetry D5
976 (<https://tinyurl.com/38v5f37m>).

977 **Fig. 2-figure supplement 1: Workflow of the C11a compact helix reconstruction process
978 and fitting.** Segment extraction was performed with a box size of 400 pixels (pix) binned
979 twice (box size 200 pix, 2.18 Å/pix). The distance between consecutive boxes was equal to
980 the axial rise calculated by indexation on the power spectrum. From 2D classification with
981 200 classes, 30 classes were selected for C11a after clustering (see methods and Fig. S5).
982 First, 3D classification was carried out using the segments from 30 2D classes, helical
983 symmetry parameters from the power spectrum indexation and a 300 Å featureless cylinder as
984 3D reference. 3D-refinement of the 4 boxed 3D-classes was achieved using one low pass
985 filtered 3D class as reference on the unbinned segments. One step of focused refinement (with
986 solvent flattening) was performed using a low pass filtered 3D reconstruction (before solvent
987 flattening) as reference. 3D-refinement was then performed. The resulting map was then used
988 as model for automatic picking. A last 3D-refinement cycle was performed using as reference
989 a featureless cylinder. The bottom right graph presents the Fourier shell correlation (FSC)
990 curves for the final 3D reconstructions (helical assembly and asymmetric unit).

991 **Fig. 2-figure supplement 2: Workflow of the C13a relaxed helix reconstruction process
992 and fitting.** Segment extraction was performed with a box size of 400 pixels (pix) binned
993 twice (box size 200 pix, 2.18 Å/pix) using the rise obtained from the power spectrum
994 indexation (see methods). From 2D classifications with 200 classes, 42 classes were selected
995 for C13a after clustering (see methods and Fig. S5). First, 3D classification was carried out
996 using the 42 2D classes, with D5 symmetry, rise and twist from the power spectrum
997 indexation and a featureless cylinder of 340 Å as reference. 3D-refinement of the two 3D-
998 classes (red boxes) was achieved using one low pass filtered 3D class as reference on the
999 unbinned segments. The resulting map was used as model for automatic picking and segment
1000 extraction. Final 3D-refinement was performed on unbinned segments using as initial
1001 reference a featureless cylinder with a last step of solvent flattening.

1002 **Fig. 2-figure supplement 3: Workflow of the C12 helix reconstruction process and fitting.**
1003 Segment extraction was performed with a box size of 400 pixels (pix) binned twice (box size
1004 200 pix, 2.18 Å/pix) using the rise obtained from the power spectrum indexation (see
1005 methods). From 2D classifications with 200 classes, 12 classes were selected for C12 based on
1006 correlation (see methods and Fig. S5). First, 3D-refinements combining 12 2D classes with
1007 solvent flattening, were performed with the C3 symmetry, rise and twist from the power
1008 spectrum indexation and a 330 Å featureless cylinder as reference. The resulting map was
1009 used as model for automatic picking and segment extraction on the unbinned segments. Final
1010 3D-refinement was performed on unbinned segments using as initial reference a featureless
1011 cylinder with a last step of solvent flattening.

1012 **Fig. 2-figure supplement 4: Comparison of sequence coverages** for qu_946 (A) and

1013 qu_143 (B) obtained by MS-based proteomic analysis of genomic fibers and total virions.
1014 Identified peptides are highlighted in red, blue and grey when identified respectively in three,
1015 two and one replicate of the purified genomic fiber.

1016 **Fig. 3-figure supplement 1: 2D Classification of subtracted segments.** 50 2D classes (2
1017 empty) were obtained after reference-free 2D classification in Relion 3.1.0 of segments (box
1018 size of 400 pixels) of the C11a genomic fibers after signal subtraction to keep only the
1019 information of the DNA strands inside the fiber. The 2D classes show the enhancement of the
1020 signal corresponding to the DNA.

1021 **Fig. 3-figure supplement 2: Workflow of the DNA strand focused refinement process.**
1022 One strand segmented from the 5-start compacted map (C11a) was low-pass filtered at 20 Å
1023 and used as reference for 3D-refinement of the dataset corresponding to the merged dataset of
1024 the individually subtracted 5 strands. The result of the refinement was low-pass filtered at 30
1025 Å and used as reference for 3D-classification. The best 3D class was further refined leading to
1026 a focused refined map of the DNA strand.

1027 **Supplementary file 1: Mimivirus Reunion genomic fibers data statistics (33)**

1028 **Supplementary file 2: Mass spectrometry-based proteomic analysis** of A] three
1029 independent preparations of mimivirus genomic fiber and B] one sample of purified
1030 mimivirus virions. (ND: Not detected). RNA polymerase subunits are marked in red.

1031 **Supplementary file 3: Contacting residues for each GMC-oxidoreductases in the**
1032 **different maps.** Conserved or divergent amino-acids are color coded in green or in red,
1033 respectively.

1034 **Supplementary file 4: Data acquisition parameters for Cryo-EM**

1035 **Supplementary Material**

1036 Figures supplement (Fig. 1-supplement figure 1-6, Fig. 2-supplement figure 1-4, Fig. 3-
1037 supplement figure 1-2)

1038 Supplementary files 1-4

1039 Animation supplement (Fig.3-animation 1)

1040 Video supplement (Fig.1-video 1-4, Fig.2-video 1, Fig. 3-video 1)

1041