



**HAL**  
open science

## Functional redundancy revealed by the deletion of the mimivirus GMC-oxidoreductase genes

Jean-Marie Alempic, Hugo Bisio, Alejandro Villalta, Sébastien Santini, Audrey Lartigue, Alain Schmitt, Claire Bugnot, Anna Notaro, Lucid Belmudes, Annie Adrait, et al.

### ► To cite this version:

Jean-Marie Alempic, Hugo Bisio, Alejandro Villalta, Sébastien Santini, Audrey Lartigue, et al.. Functional redundancy revealed by the deletion of the mimivirus GMC-oxidoreductase genes. *microLife*, 2024, 5, 10.1093/femsml/uqae006 . hal-04571293

**HAL Id: hal-04571293**

**<https://amu.hal.science/hal-04571293>**

Submitted on 7 May 2024

**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.

1 **Functional redundancy revealed by the deletion of the mimivirus GMC-oxidoreductase**  
2 **genes**

3 **Authors:** Jean-Marie Alempic<sup>1†</sup>, Hugo Bisio<sup>1†</sup>, Alejandro Villalta<sup>1†</sup>, Sébastien Santini<sup>1</sup>,  
4 Audrey Lartigue<sup>1</sup>, Alain Schmitt<sup>1</sup>, Claire Bugnot<sup>1#</sup>, Anna Notaro<sup>2</sup>, Lucid Belmudes<sup>3</sup>, Annie  
5 Adrait<sup>3</sup>, Olivier Poirot<sup>1</sup>, Denis Ptchelkine<sup>4</sup>, Cristina De Castro<sup>2</sup>, Yohann Couté<sup>3</sup>, Chantal Ab-  
6 ergel<sup>1\*</sup>

7 DOI: [10.1093/femsml/uqae006](https://doi.org/10.1093/femsml/uqae006)

8 **Affiliations :**

9 <sup>1</sup>Aix–Marseille University, Centre National de la Recherche Scientifique, Information Généo-  
10 mique & Structurale, Unité Mixte de Recherche 7256 (Institut de Microbiologie de la Médi-  
11 terranée, FR3479, IM2B, IOM), 13288 Marseille Cedex 9, France.

12 <sup>2</sup>Department of Agricultural Sciences, University of Naples Federico II, Via Università 100,  
13 80055, Portici, Italy

14 <sup>3</sup>Univ. Grenoble Alpes, CEA, INSERM, UA13 BGE, CNRS, CEA, FR2048, 38000 Grenoble,  
15 France

16 <sup>4</sup>Aix–Marseille University, Centre National de la Recherche Scientifique, Architecture et  
17 Fonction des Macromolécules Biologiques, Unité Mixte de Recherche 7257 (IM2B), 13288  
18 Marseille Cedex 9, France.

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

20 †These authors contributed equally to the work

21 # Present address: Sorbonne Université, Centre National de la Recherche Scientifique, BIOM  
22 UMR7232, Observatoire Océanologique, 1 avenue Pierre Fabre, 66650 Banyuls sur mer,  
23 France.

24 **One-Sentence Summary:** Functional redundancy preserves mimivirus genomic fiber and layer of  
25 fibrils formation.

26 **Subject Area:** Biological sciences/Microbiology/Virology/Viral evolution; Biological  
27 sciences/Microbiology/Virology/Virus structures

28 **Keywords:** Mimivirus, giant virus, genomic fiber, cryo-EM, helical reconstruction, layer of fibrils,  
29 glycosylation, MS-based proteomics.

30

## 31 **Abstract**

32 The mimivirus 1.2Mb genome was shown to be organized into a nucleocapsid-like genomic  
33 fiber encased in the nucleoid compartment inside the icosahedral capsid (1). The genomic  
34 fiber protein shell is composed of a mixture of two GMC-oxidoreductase paralogs, one of  
35 them being the main component of the glycosylated layer of fibrils at the surface of the virion  
36 (2). In this study, we determined the effect of the deletion of each of the corresponding genes  
37 on the genomic fiber and the layer of surface fibrils. First, we deleted the GMC-  
38 oxidoreductase the most abundant in the genomic fiber, and determined its structure and  
39 composition in the mutant. As expected, it was composed of the second GMC-oxidoreductase  
40 and contained 5- and 6-start helices similar to the wild-type fiber. This result led us to propose  
41 a model explaining their coexistence. Then, we deleted the GMC-oxidoreductase the most  
42 abundant in the layer of fibrils to analyze its protein composition in the mutant. Second, we  
43 showed that the fitness of single mutants and the double mutant were not decreased compared  
44 to the wild-type viruses in laboratory conditions. Third, we determined that deleting the  
45 GMC-oxidoreductase genes did not impact the glycosylation or the glycan composition of the  
46 layer of surface fibrils, despite modifying their protein composition. Since the glycosylation  
47 machinery and glycan composition of members of different clades are different (3, 4), we  
48 expanded the analysis of the protein composition of the layer of fibrils to members of the B  
49 and C clades and showed that it was different among the three clades and even among isolates  
50 within the same clade. Taken together, the results obtained on two distinct central processes  
51 (genome packaging and virion coating) illustrate an unexpected functional redundancy in  
52 members of the family *Mimiviridae*, suggesting this may be the major evolutionary force  
53 behind their giant genomes.

## 54 **Introduction**

55 Mimivirus is the inaugural member of the family *Mimiviridae* part of the *Nucleocytoviricota*  
56 phylum encompassing large and giant DNA viruses infecting eukaryotes (5). Members of the  
57 family *Mimiviridae* infecting amoeba have dsDNA genomes up to 1.5 Mb encoding over  
58 1000 proteins, including a complete glycosylation machinery (3, 4, 6–9). Mimivirus virion  
59 penetrate the cell through phagocytosis, and the acidic vacuole mediates opening of the  
60 stargate structure at one vertex of its icosahedral capsid (10, 11). The internal membrane  
61 unwraps and fuses with the phagosome membrane, allowing transfer of the nucleoid  
62 compartment into the host cytoplasm, while empty capsids remain in the vacuole (6, 10, 12).

63 The infectious cycle occurs in the cytoplasm where a large viral factory is developed (6, 12–  
64 14). At the late stage of the cycle, neo-synthesized virions bud at the periphery of the viral  
65 factory where they are filled with the genome. Finally, a glycosylated layer of fibrils  
66 composed of proteins and two large polysaccharides synthesized by the virally-encoded  
67 machinery, is added to the capsids (4, 15, 16). As a result, the 750 nm-diameter virions  
68 resemble Russian dolls made of the external layer of reticulated glycosylated fibrils (referred  
69 to as the “layer of fibrils”) decorating the surface of the icosahedral capsids. Underneath the  
70 capsid shell, the nucleoid compartment encases the 1.2 Mb dsDNA genome organized into a  
71 30 nm large nucleocapsid-like structure (referred to as the “genomic fiber”). The genomic  
72 fiber is made of a protein shell internally lined by the folded DNA and a central channel that  
73 can accommodate large proteins such as the viral RNA polymerase (1). Three independent  
74 genomic fiber structures have been determined by cryo-electron microscopy (cryo-EM): two  
75 compact 5- and 6-start DNA-containing helices, and a 5-start relaxed helix, without DNA (1).  
76 Unexpectedly, the protein shell was found to be composed of two glucose–methanol–choline  
77 (GMC) oxidoreductases sharing 69% sequence identity, with a ratio of 5 between qu\_946 and  
78 qu\_143 according to the protein composition of the purified genomic fiber (1). The resolution  
79 of the reconstructions (3.7 and 4.4 Å) prevented us from determining whether each helix  
80 contained a single paralog or a mixture of both. Interestingly, one of the two GMC-  
81 oxidoreductases, qu\_143 (R135 in mimivirus prototype), was known to compose the external  
82 layer of fibrils at the surface of mimivirus capsids and it was hypothesized that it was the  
83 major target for glycosylation (4, 9, 17, 18). However, despite its involvement in both the  
84 mimivirus genomic fiber inside the nucleoid, and the layer of fibrils at the periphery of the  
85 icosahedral capsid, GMC-oxidoreductase homologs are absent in the laboratory-evolved  
86 mimivirus M4 strain (18). M4 also lacks the glycosylation machinery, described for members  
87 of different clades of the subfamily *Megamimivirinae* and responsible for synthesizing and  
88 branching the polysaccharides on the capsids (3, 4). In the present study, we used in-house  
89 developed tools (19) to delete mimivirus GMC-oxidoreductase genes. We then assessed the  
90 fitness cost associated to these deletions and investigated their impact on the formation of the  
91 genomic fiber and the protein and glycan composition of the layer of fibrils. Cryo-EM was  
92 used to determine the structure of the KO\_946 genomic fiber made of qu\_143, the less  
93 abundant in the wild-type (wt) genomic fiber. Nuclear magnetic resonance (NMR) and Gas  
94 Chromatography Mass Spectrometry (GC-MS) were used to analyze the compositions in  
95 glycans and their structures for each mutant and to compare them with the wt layer of fibrils.  
96 We used Mass spectrometry (MS)-based proteomics to analyze for each of the three mutants

97 the protein composition of their layer of fibrils and extended the study to members belonging  
98 to B and C clades (moumouvirus and megaviruses), known to glycosylate their layer of  
99 fibrils with different glycans using a clade-specific glycosylation machineries (3, 9). While  
100 confirming the non-essentiality of the two GMC-oxidoreductases, our results document the  
101 unexpected resilience of mimivirus to the deletion of these two genes through the use of  
102 alternative proteins to compensate their disruptions.

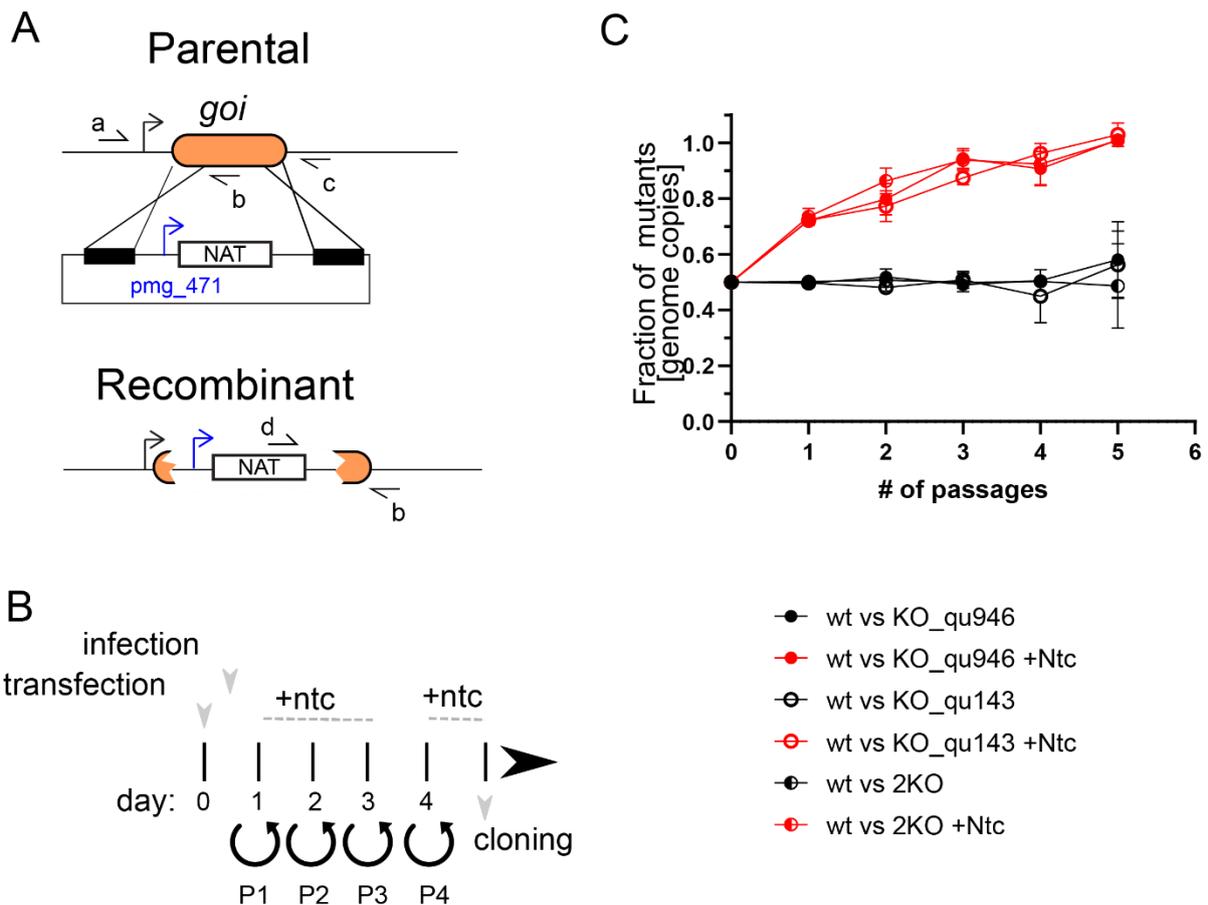
## 103 **Results**

### 104 **None of the two GMC-oxidoreductases is essential**

105 We used our recently developed protocol (19) combining homologous recombination  
106 with the introduction of a nourseothricin N-acetyl transferase (NAT) selection cassette to  
107 delete each of the two genes encoding the GMC oxidoreductases (qu\_946 and qu\_143). We  
108 selected recombinant viruses that were cloned to obtain homogeneous populations (Fig 1B)  
109 (19). Each mutant was easily produced and genotyped to confirm the mutation (Fig. S1).  
110 Using a second Neomycin resistance gene (NEO) selection cassette we were able to delete  
111 both genes (Fig. S1) demonstrating that the two GMC-oxidoreductases were not essential.  
112 The absence of additional mutations in every mutant was confirmed by genome sequencing.

### 113 *Mutants' fitness*

114 To assess whether a fitness cost was associated with the mutations, we performed  
115 competition assays against wt mimivirus reunion strain by measuring the abundance of each  
116 mutant over several cycles in the presence and absence of selection. In contrast to the wt, each  
117 mutant presents a common nourseothricin resistance gene. The double deletion mutant (2KO)  
118 encodes for an additional geneticin resistance gene. As a result, in the presence of  
119 nourseothricin, each mutant ratio increased, with the disappearance of the wt virus after 5  
120 passages (Fig 1C). These data allowed competition assays to be validated as an effective tool  
121 to assess the fitness of recombinant viruses. In the absence of selection, the relative  
122 abundance of the mutants compared to wt remained around 0.5 over 5 passages, supporting  
123 the absence of a fitness cost, even for the double deletion mutant (Fig 1C).



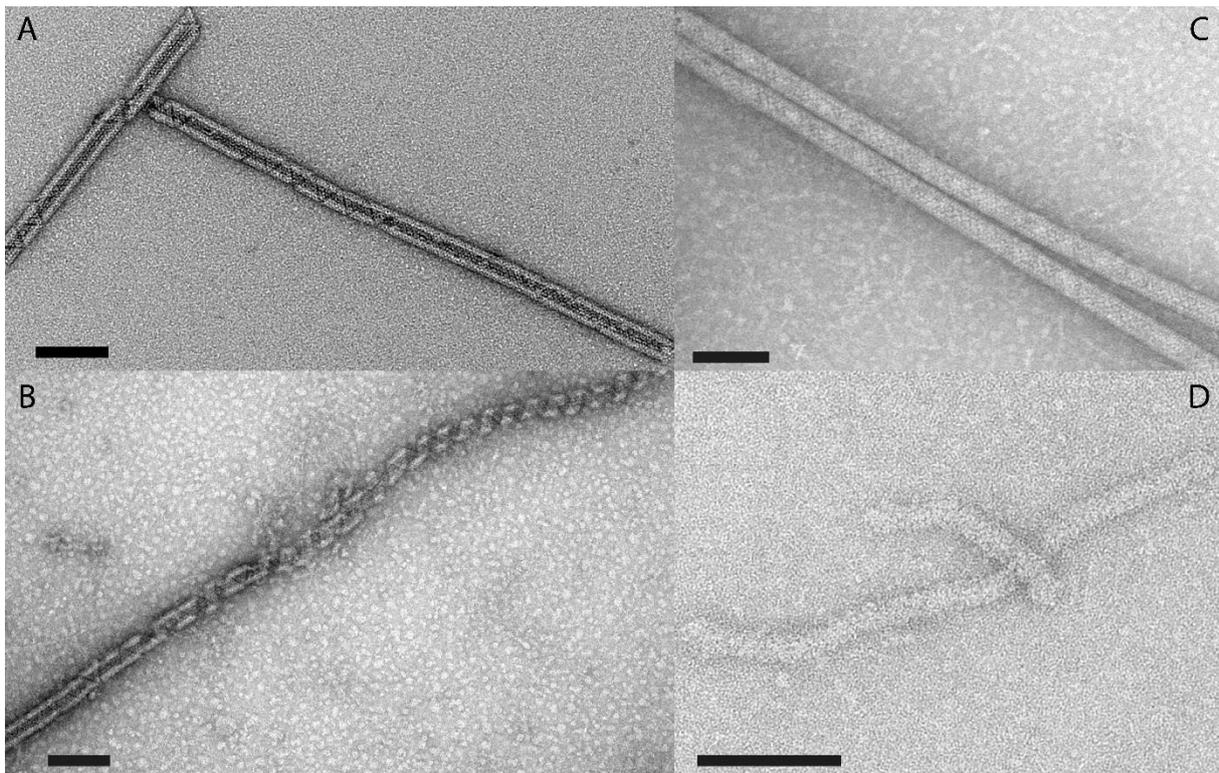
124

125 **Figure 1: Mimivirus mutants' generation and phenotypic characterization.** A) Schematic  
 126 representation of the vector and strategy utilized for deletion of qu\_143 (KO\_qu143) and  
 127 qu\_946 (KO\_qu946) in mimivirus reunion strain. *goi*: gene of interest. Selection cassette was  
 128 introduced by homologous recombination and recombinant viruses were generated as shown  
 129 in Fig 1B. Primer annealing sites are also shown and the sequence of the primers is included  
 130 in Table S1. B) Cartoon depicting the strategy for the selection of recombinant viruses. Viral  
 131 infection was performed 1-hour post-transfection. Ntc: Nourseothricin. P= passage. C)  
 132 Growth competition assays revealed no significant defects in the lytic cycle of deletion  
 133 strains. The competition was also performed in presence of Nourseothricin which allows the  
 134 outcompetition of the recombinant strains due to the expression of a Nourseothricin selection  
 135 cassette. Measurements were performed by qPCR of an endogenous locus (present in wt and  
 136 recombinant strains) and the Nourseothricin selection cassette (only present in recombinant  
 137 viruses).

138 *Composition of the genomic fiber of single mutants*

139 To determine the composition of the genomic fiber of each single mutant we extracted  
 140 and purified their genomic fiber. MS-based proteomics confirmed that the most abundant  
 141 protein was the remaining GMC-oxidoreductase. Negative staining transmission electron  
 142 microscopy (NS-TEM) highlighted surprising differences between the two structures (Fig 2),  
 143 despite the fact that the two proteins (qu\_946 and qu\_143) share 69% sequence identity (81%  
 144 similarity). Specifically, the genomic fiber extracted from the KO\_qu143 mutant and made of

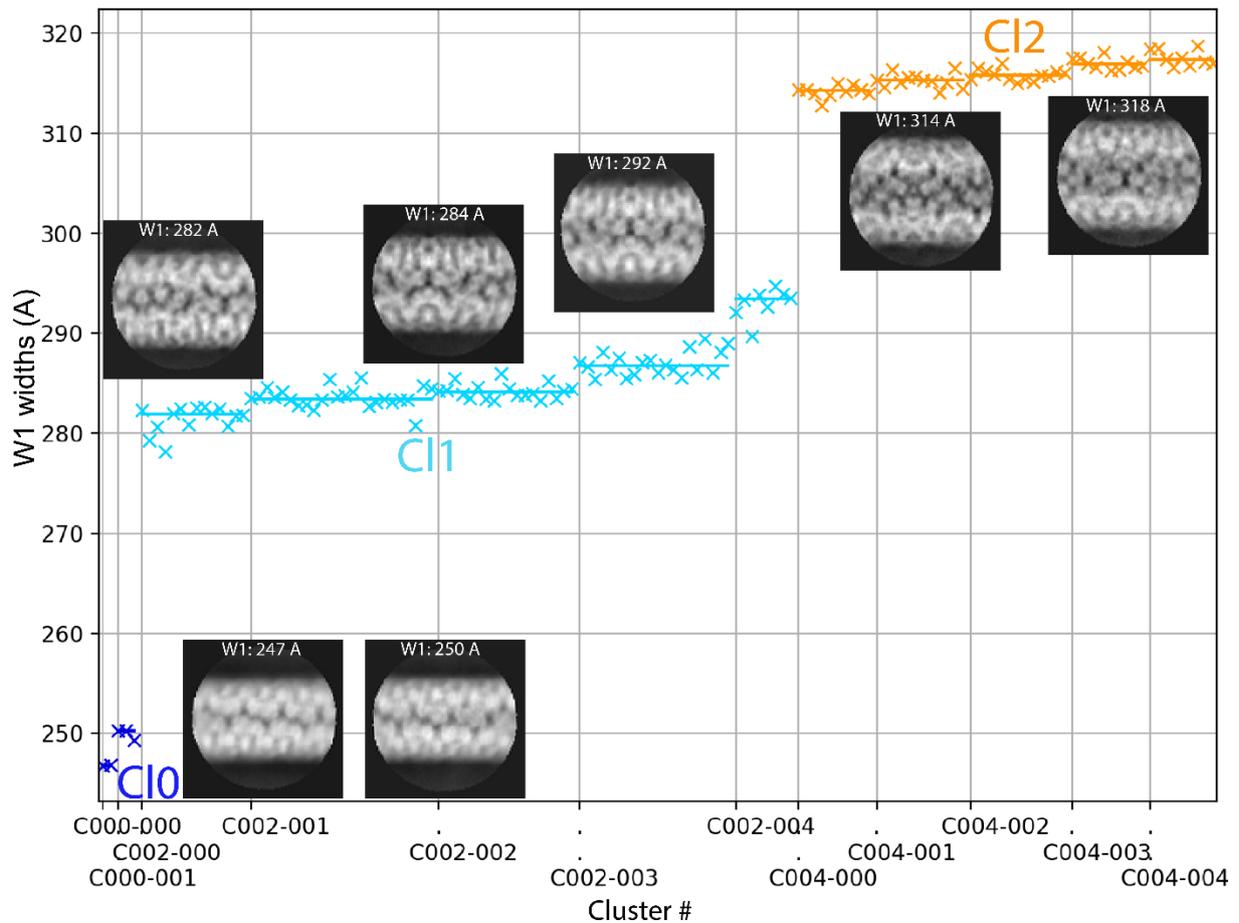
145 qu\_946 (the most abundant in the wild type genomic fiber) was mostly in the unwound state  
146 (Fig 2B), while the one made of qu\_143 (KO\_qu946), resulted in very long and stable helices  
147 that did not unwind (Fig 2C). Thus, the use of both GMC-oxidoreductases in the wt genomic  
148 fiber could contribute to a fine tuning of its biophysical properties, with an intermediate state  
149 between the wt and each deletion mutant (Fig 2A). In the case of the double mutant, the  
150 protocols for capsid opening to extract the genomic fiber of wt or single mutants did not work  
151 properly. An optimized protocol allowed the extraction of a possible thinner genomic fiber,  
152 but in poor yield, preventing its purification and compositional characterization (Fig 2D).



153  
154 **Figure 2: Micrograph of negative stained genomic fiber of A] Wild-type mimivirus, B]**  
155 **KO\_qu143 mutant, C] KO\_qu946 mutant, and D] 2KO double mutant. Scale bars 100 nm.**

156 **Cryo-EM single particle analysis of the qu\_143 genomic fiber**

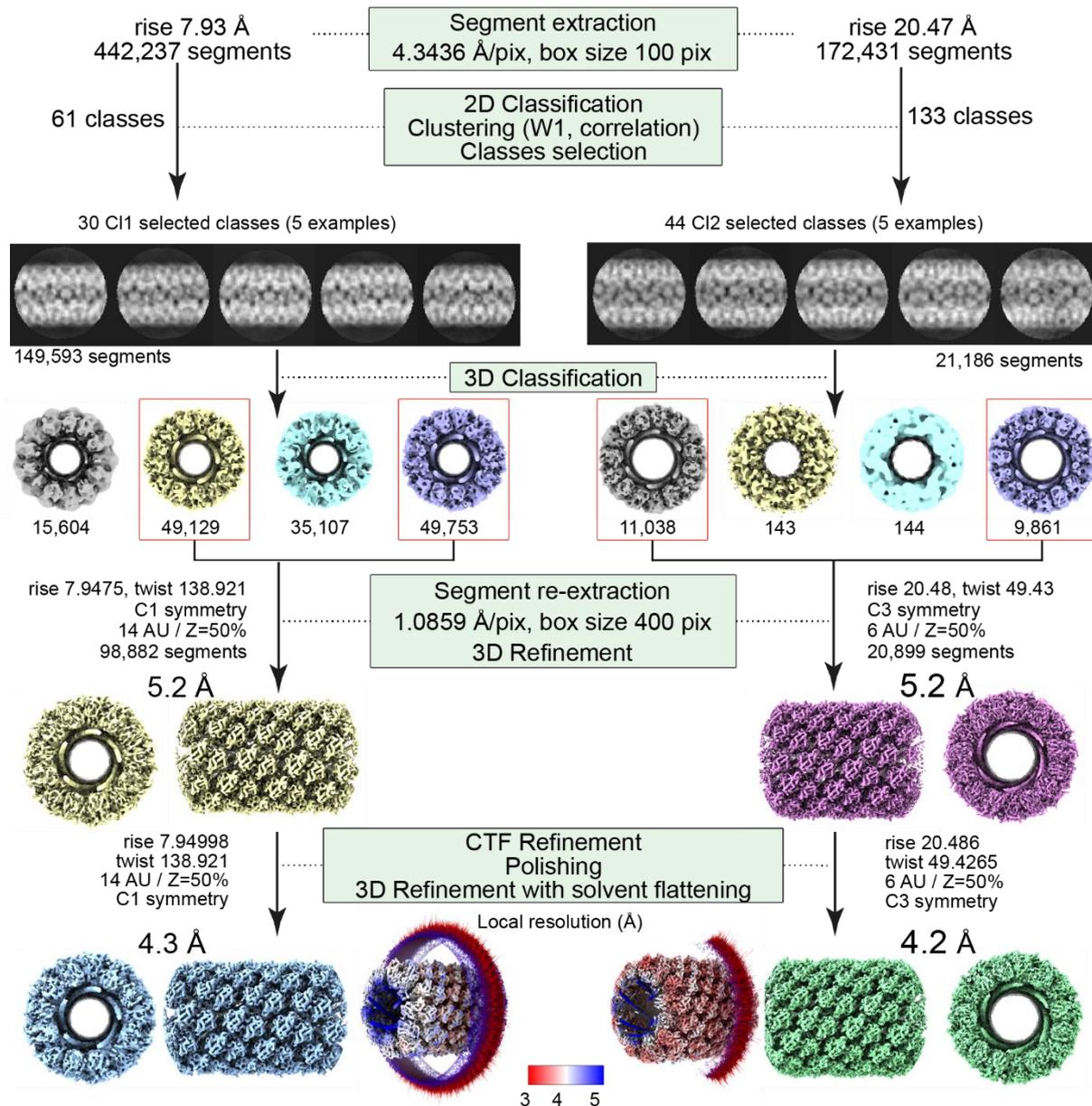
157 To determine the contribution of each GMC-oxidoreductase to the wt genomic fiber  
158 we performed cryo-EM single-particle analysis on the most stable genomic fiber composed by  
159 qu\_143 (mutant KO\_qu946). As for the wt, the 2D-classification revealed an heterogeneity of  
160 the sample, and the 2D classes were sorted by applying our already described clustering  
161 protocol (1). The two main clusters corresponding in width to the compact (C11a) and (C12)  
162 structures of the wt genomic fiber were respectively named C11 and C12 (Fig 3).



163

164 **Figure 3: Clustering of the 2D classes obtained with qu\_143 genomic fiber:** Automatic  
 165 sorting of the 2D classes using the fiber width W1 and pairwise correlations of the 2D classes  
 166 resulted in 2 main clusters (5-start CI1 in cyan; 6-start CI2 in orange) and a smaller cluster  
 167 (CI0 in dark blue). Each cross corresponds to a 2D class and its associated W1.  
 168 Representative 2D classes are displayed with their respective W1.

169 For each main cluster, we confirmed that the helical symmetry parameters were the  
 170 same as the wt genomic fiber and proceeded to structure determination and refinement (Fig  
 171 4). For the less populated smaller cluster (CI0, ~25 nm), absent from the wt genomic fiber 2D-  
 172 classes, we failed to identify its helical parameters due to the lower number of segments and  
 173 the resulting lower resolution. After 3D-refinement, we obtained a 4.3 Å resolution helical  
 174 ~29 nm diameter structure (FSC threshold 0.5, masked) for CI1. This structure corresponds to  
 175 the same 5-start left-handed helix as the wt (CI1a), made of a ~8 nm-thick proteinaceous  
 176 external shell with 5 dsDNA strands lining the interior of the shell and a ~9 nm wide central  
 177 channel (Fig 4, Fig. S2). The 4.2 Å resolution CI2 map obtained after 3D refinement (Fig 4,  
 178 Fig. S 2) corresponds to the same ~32 nm diameter 6-start left-handed helix as the wt, with 6  
 179 dsDNA strands lining the external shell and a ~12 nm wide inner channel (Fig 4).



180

181 **Figure 4: Workflow of the 5- and 6-start helices reconstruction processes.** Segment  
 182 extraction was performed with a box size of 400 pixels (pix) binned (box size 100 pix, 4.3436  
 183 Å/pix). The distance between consecutive boxes was equal to the axial rise calculated by  
 184 indexation of the power spectrum. After clustering, 2D classes were selected for C11 and C12  
 185 and 3D classification was carried out using the selected segments, helical symmetry  
 186 parameters from the power spectrum indexation, and a 300 Å or 330 Å featureless cylinder as  
 187 3D reference for C11 and C12, respectively. 3D-refinement of the 2 boxed 3D classes was  
 188 achieved using one low pass filtered 3D class as reference on the unbinned segments. A first  
 189 3D-Refinement was performed with solvent flattening followed by CTF refinement and  
 190 polishing of the selected segments. A last 3D refinement was achieved with solvent flattening.  
 191 The EM maps colored by local resolution from 5 Å (blue) to 3 Å (red) with Euler angle  
 192 distribution of the particles used for the 3D reconstruction are presented.

193

194

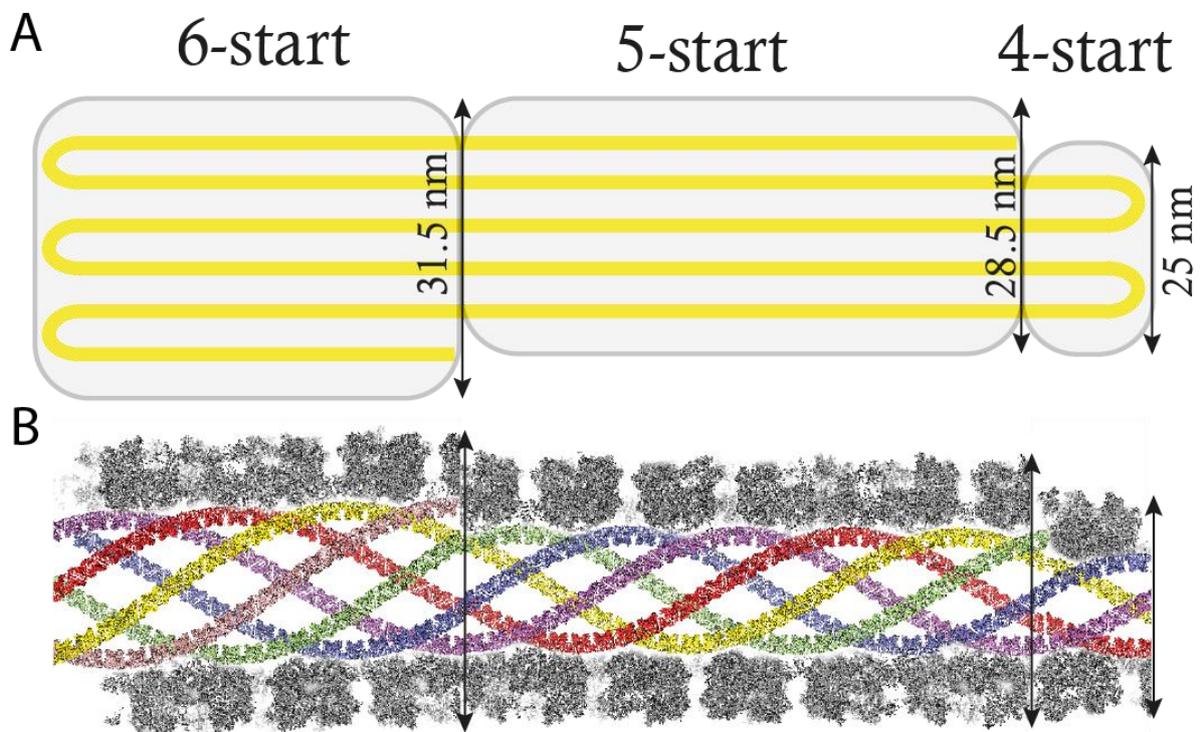
195

Since the helical parameters between the wt C11a and the mutant C11 are the same, we used  
 the qu\_143 dimeric structure refined in the C11a focus refined map for refinement into the  
 mutant maps (Material and Methods section and Table S3). The density that can be attributed

196 to the FAD cofactor was present in both maps (Fig. S3) and the models of C11 and C12 dimers  
197 are superimposable with a core RMSD of 0.467 Å based on C $\alpha$  atoms (Table S4). As the  
198 qu\_143 helices are more stable than the wt (composed of both GMC-oxidoreductases), the  
199 relaxed C13 cluster was never observed with this mutant.

## 200 **Model explaining the co-existence of 5- and 6-start helices**

201 As for the wt, the genomic fiber of KO\_qu946 is composed of a mixture of 5 and 6  
202 strands of DNA, despite the presence of a single GMC-oxidoreductase into the shell. We  
203 estimated the ratio of 5-start and 6-start from the clustering (Fig 3) and can now propose a  
204 model that reconciles the co-occurrence of the two structures. In this model, the whole  
205 genome would be folded into 6 parallel strands, 5 longer than the 6<sup>th</sup> one. The helix would  
206 then be formed initially as a 6-start helix until the sixth strand ends and, from that point,  
207 becomes a 5-start helix (Fig 5). According to this model, assuming the length of the genomic  
208 fiber is limited by the size of mimivirus nucleoid compartment, we can estimate that the  
209 maximum genome length would be ~1.4 Mb for a full 6-start helix with an 8 nm thick protein  
210 shell. We hypothesize that the last cluster (C10, Fig 3) could correspond to a 4-start with an  
211 additional shorter DNA strand.



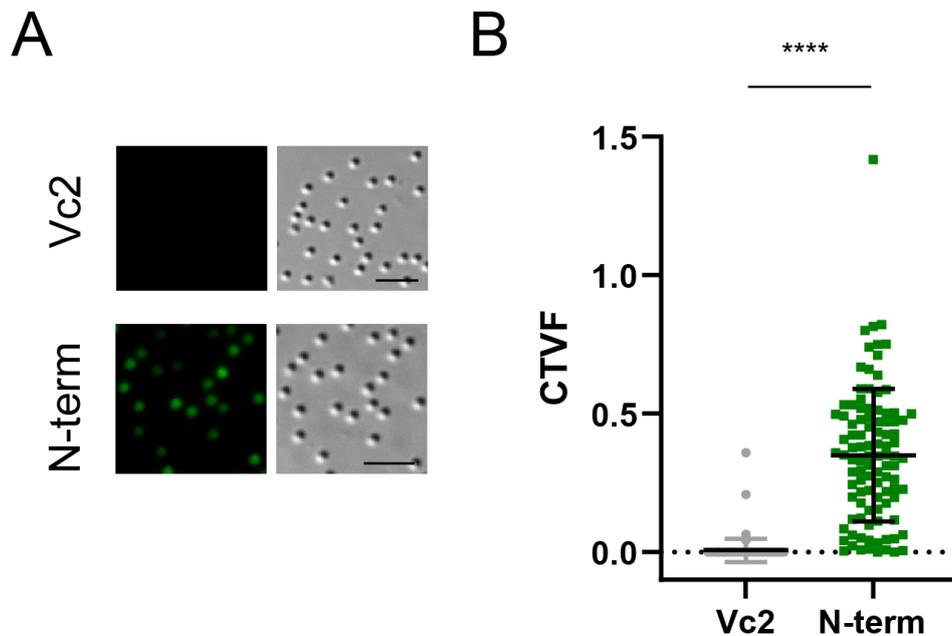
212

213 **Figure 5: Model explaining the transition from a 6- to a 5-start helix.** A] Flat models of  
214 the transition from a 6- to a 5-starts involving a decrease of the helix diameter by ~3.5 nm.  
215 The small cluster could thus correspond to a 25 nm diameter 4-start helix. B] Cartoon model  
216 of the different helices. A longitudinal section of the GMC-oxidoreductase shell is

217 represented around the central channel and each helix was positioned to produce a certain  
218 continuity of the DNA strands.

### 219 **The N-terminal cys-pro-rich domain is an addressing domain to the virion surface**

220 The sequence of the cys-pro-rich N-ter domain of the GMC-oxidoreductases is not  
221 covered by proteomic analysis of the purified genomic fiber but is covered by peptides in the  
222 purified fibrils forming the external layer at the surface of the capsids (1). To assess whether  
223 this cys-pro-rich N-ter domain could be a structural signature used to address proteins to the  
224 layer of fibrils, we replaced the second GMC-oxidoreductase (qu\_143) in the genome of the  
225 mutant KO\_qu946 by the sequence of the GFP in fusion with the sequence of the qu\_143 N-  
226 terminal domain (Nqu143-GFP). We then analyzed the resulting virions by MS-based  
227 proteomics (Table 1 and Table S2) and fluorescence microscopy. Purified Nqu143-GFP  
228 virions showed a strong fluorescence supporting the incorporation of the chimeric protein into  
229 the viral particles (Fig 6). Moreover, after defibrillation of the virions the GFP fluorescence  
230 was lost. In addition, Nqu143-GFP was identified in purified virions (ranked 97<sup>th</sup> in terms of  
231 relative abundance, Table S2) and was found enriched 12-times in the fraction containing the  
232 external fibrils (ranked 20<sup>th</sup>, Table 1 and Table S2) with peptides covering the N-terminal  
233 domain (Fig. S5). Taken together, these data indicate that the N-terminal cys-pro-rich domain  
234 of the GMC oxidoreductase is sufficient to direct the proteins to the layer of fibrils at the  
235 surface of mimivirus particles.



236

237 **Figure 6: Fluorescent virion (A) images and (B) fluorescence virion quantification** in infected cells overexpressing GFP (Vc2) and in cells infected by the mutant Nqu143-GFP. Scale bar 5  
238  $\mu\text{m}$ . (CTVF: corrected total virion fluorescence).  
239

## 240 **Composition of the purified external fibrils**

### 241 *Protein composition of the layer of fibrils in additional members of the family Mimiviridae*

242 The members of the *Megavirinae* subfamily (20) currently encompasses five clades,  
243 the mimiviruses (A clade), moomouviruses (B clade), megaviruses (C clade), tupanviruses (D  
244 clade) (8) and cottonvirus japonicus (E clade) (21). As of today, all members of the  
245 *Megavirinae* are characterized by a layer of fibrils that differs in thickness and lengths among  
246 the clades and, excepted for clade A (1, 18), their protein composition was unknown. While  
247 the cys-pro-rich N-ter domain of the GMC-oxidoreductases is conserved in all members of  
248 the A clade, it is absent in the orthologous proteins in members of the B and C clades.  
249 Moreover, homologs of the two GMC-oxidoreductases are pseudogenized in tupanviruses,  
250 suggesting that different proteins should compose their fibrils (Fig. S6). We thus conducted a  
251 systematic proteomic analysis of the purified fibrils of different members of the 3 clades, in  
252 addition to the mutants. For the double mutant (2KO) we identified a group of proteins as the  
253 most abundant in the fibrils fraction (qu\_734, qu\_757, qu\_384 and qu\_482, Table 1 and Table  
254 S2). These proteins were also highly ranked in the Nqu143-GFP double mutant fibrils and the  
255 orthologous of qu\_734 was identified as the most abundant protein in the laboratory-evolved  
256 M4 mutant fibrils (696-L688, Table 1 and Table S2). This mutant does not encode a  
257 glycosylation machinery and is not glycosylated (4, 18), thus its capsid lacks the large  
258 reticulated layer of fibrils decorating mimivirus capsid (Fig. S8F). Interestingly the qu\_734  
259 protein also possesses a cys-pro-rich N-ter domain which is conserved in all clades (Fig. S6).  
260 It was not possible to determine if one of these proteins was also the building block  
261 composing the 2KO genomic fiber, given the difficulty to open the 2KO capsids. Thus, we  
262 concluded that the change in protein composition of the fibrils led to capsids with different  
263 stability properties. While, as for the wt, the most abundant protein in KO\_qu946 remains  
264 qu\_143, the lack of qu143 in KO\_qu143 does not lead to its replacement by the second GMC-  
265 oxidoreductase. Instead, it is replaced by a group of proteins, with qu\_465 (predicted as a  
266 thioredoxin domain containing protein) as the most abundant, followed by qu\_757, the second  
267 most abundant in the wt fibrils (Table I and Table S2).

268 The proteomic analysis of fibrils purified from two isolates of the B clade,  
269 moomouvirus australiensis and maliensis, showed similar protein compositions, but with  
270 slight differences in their relative abundances (Table 1 and Table S2). The moomouvirus  
271 maliensis protein mm\_751, ranked 1<sup>st</sup> in its fibrils, is absent from A clade fibrils and is in the  
272 top 3 in the fibrils of members of the B and C clades. Cystein and proline amino acids are

273 present in the N-ter domain of mm\_751, but less abundant than in the GMC-oxidoreductases.  
274 For ma\_195 (qu\_738 in mimivirus reunion, Table 1 and Table S2), ranked 1<sup>st</sup> in the fibrils of  
275 mousmouvirus australiensis, the N-ter domain is not cys-pro-rich. The best ranked in the fibrils  
276 of members of the C clade are the same as for members of the B clade. The first ranked in  
277 megavirus chilensis is mg749 (qu\_657 in mimivirus reunion) which presents a cys-pro-rich  
278 N-ter domain, while the first ranked in megavirus vitis, mvi\_646 (qu\_600 in mimivirus  
279 reunion), lacks a cys-pro-rich N-ter domain. Yet, for each virus in each clade, there is at least  
280 one protein with a cys-pro-rich N-terminal domain among the most abundant in the fibrils.

Rank	Mimivirus reunion	KO_946	KO_143	2KO	2KO-GFP	M4	Megavirus chilensis	Megavirus vitis	Moumouvirus ausraliensis	Moumouvirus maliensis
qu_143	1	1	NA	NA	NA	NA	73	46	45	130
qu_757	2	3	2	2	11	7	44	24	7	19
qu_734	3	7	6	1	8	1	7	14	14	16
qu_482	4	42	45	4	1	54	42	90	9	42
qu_880	5	13	13	11	22	NA	6	18	18	24
qu_384	6	4	8	3	2	2	9	4	13	18
qu_773	14	8	4	10	9	3	8	27	5	8
qu_657	15	5	3	27	6	27	1	7	4	2
qu_600	16	6	5	7	3	4	10	1	3	4
qu_738	17	15	23	12	5	5	3	2	1	3
qu_752	35	17	22	55	4	21	NA	NA	NA	NA
qu_582	39	11	7	43	45	28	26	55	33	39
qu_465	10	2	1	6	18	6	17	42	32	29
mm_751	NA	NA	NA	NA	NA	NA	2	3	2	1
Nqu143-GFP	NA	NA	NA	NA	20	NA	NA	NA	NA	NA

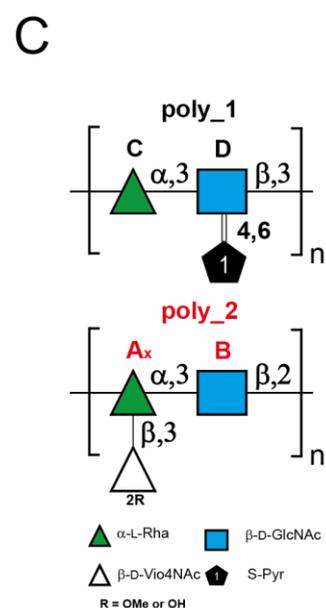
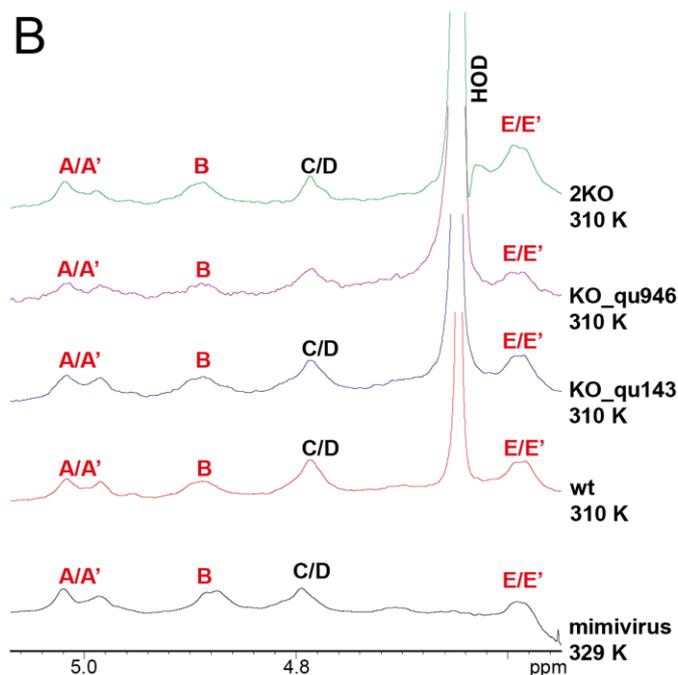
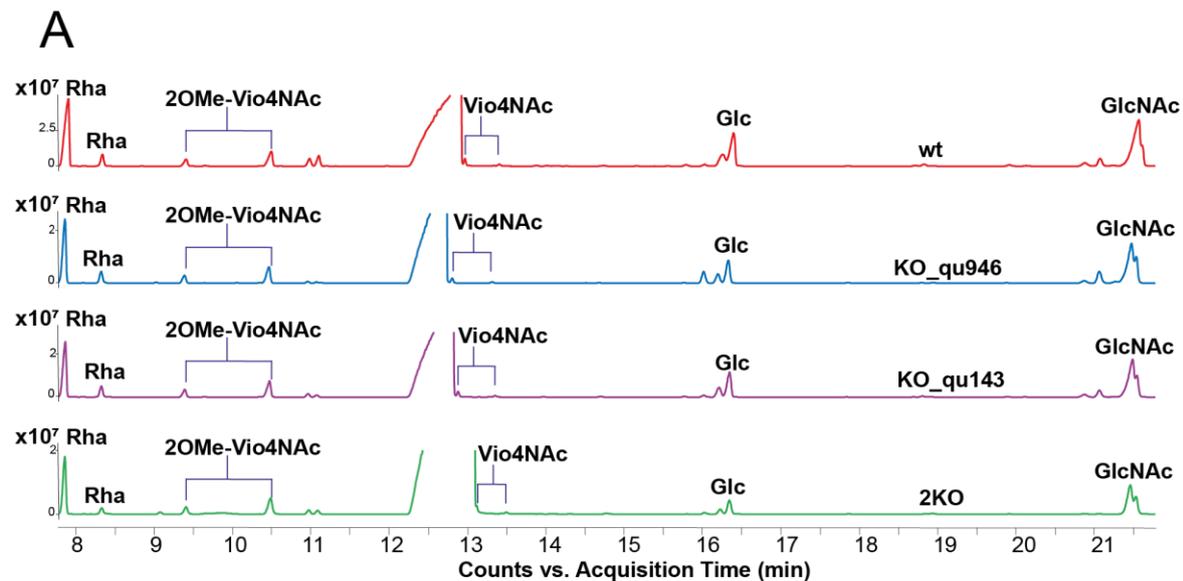
281 **Table 1: Rank of the most abundant proteins in the purified fibrils of members of clade A, B and C**

282 NA: Not applicable, the corresponding gene is absent in the corresponding genome. Sequences of the relevant orthologues of mimivirus reunion  
283 in other viruses are provided in Fig. S6.

284

285 *Glycan composition of the fibrils in mimivirus mutants*

286 We recently established that the fibrils were coated with glycans (3, 9) and demonstrated  
287 that for mimivirus, the prototype of the A clade, they were made of two distinct large molecular  
288 weight polysaccharides (4). Since the cluster of 12 genes responsible for the biosynthesis of the  
289 polysaccharides is conserved in the A clade (3), we hypothesized that the fibrils of the  
290 mimivirus reunion strain would have the same composition as mimivirus. To assess whether  
291 the knockout of the most abundant protein composing the fibrils could also affect the branching  
292 of these polysaccharides and their composition, we analyzed the sugar content of the viral  
293 particles of all three mimivirus reunion mutants (KO\_qu946, KO\_qu143, and 2KO) together  
294 with the wt strain, as reported in (3, 4). The chemical characterization revealed for each of them  
295 the presence of sugars (Figure 7A), with rhamnose (Rha), viosamine (Vio), 2-OMe-Vio,  
296 glucose (Glc), and glucosamine (GlcN) (Fig 7A), confirming that mimivirus reunion strain and  
297 all mutants had the same glycan composition as mimivirus prototype (3). We then analyzed the  
298 fibrils of the mutants and wt by <sup>1</sup>H NMR spectroscopy (22) and compared them with the  
299 mimivirus prototype, which confirmed that the different sugars were assembled the same way  
300 as in the reference mimivirus to produce the same two polysaccharides (Fig 7, Fig. S7) (4). NS-  
301 TEM images of the virions of the different mutants were obtained after methylcellulose staining  
302 to assess whether the mutations changed the layer of fibrils appearance, which would suggest  
303 that the change in protein impacted their level of glycosylation (23). All mutants' virions  
304 showed the same cross-linked outer layer, suggesting that despite their differences in protein  
305 composition, their glycosylation was not affected (Fig. S8).



306  
 307 **Figure 7: Compositional and NMR analysis of the fibrils** of mimivirus reunion strain wt and  
 308 mutants A] GC-MS chromatogram profiles of the sugars composing the fibrils of wt (a),  
 309 KO\_qu946 (b), KO\_qu143 (c), and 2KO (d). B] Comparison of the <sup>1</sup>H NMR spectra of  
 310 mimivirus reunion strain wt and related mutants with that of mimivirus prototype strain. The  
 311 anomeric signals related to poly\_1 (C and D units, in black), and those of poly\_2 are in red. C]  
 312 Structures of mimivirus polysaccharides as reported in Notaro *et al*, 2021 (4).

313 It was hypothesized that mimivirus R135 GMC oxidoreductases (qu\_143 in mimivirus  
 314 reunion) was the major target for glycosylation in the surface fibrils of mimivirus (4, 18). The  
 315 results presented here indicate that the mutations, including the deletion of the two GMC-  
 316 oxidoreductase genes, did not affect the surface glycosylation. These data suggest that the GMC

317 oxidoreductases are not the main target for glycosylation. Alternatively, it is also possible that  
318 the glycosylation machinery was able to use the qu\_734 protein as support for the two  
319 polysaccharides. As expected, no glycans were identified in the laboratory-evolved M4 mutant  
320 layer of fibrils.

## 321 **Discussion**

322 Genome packaging is essential for the propagation of viruses. For instance, the  
323 packaging ATPase of poxviruses (24) or the histones of marseilleviruses (25, 26) are essential  
324 for the productive infection by their virions. In addition, proteins at the surface of viral  
325 capsids usually play a central role in interacting with the host cell surface and the initial steps  
326 of infection (27). The two processes, genome encapsidation and viral entry, are thus essential.  
327 As the GMC oxidoreductase are composing both the genomic fiber and the surface fibrils (1)  
328 we could have expected them to be essential. However, these two enzymes are absent in the  
329 laboratory-evolved mimivirus M4 strain that also lost the glycosylation machinery (18). Our  
330 recent implementation of genetics tools for cytoplasmic giant viruses provided the first  
331 opportunity to directly assess the impact of the GMC-oxidoreductases deletions in mimivirus  
332 reunion strain (19). Given the homology between the two enzymes, the deletion of the most  
333 abundant one in the wt genomic fiber induced its replacement by the second one, with no  
334 apparent cost for the virus, in laboratory conditions. Yet, the single mutant genomic fibers  
335 were significantly different compared to the wt (Fig. 2), indicating a certain degree of  
336 functional specialization for each GMC-oxidoreductase. While the wt genomic fiber was  
337 highly heterogeneous, with fully relaxed helices having lost DNA (1), the KO\_qu946  
338 genomic fiber appeared more stable, with no occurrence of fully relaxed structures. In  
339 contrast, the deletion of qu\_143 seemed to increase the genomic fiber instability, suggesting a  
340 stabilizing role for this second GMC-oxidoreductase (Fig. 2). The cryo-EM analysis of the  
341 KO\_qu946 genomic fiber structure, as for the wt, produced the same two 5- and 6-start  
342 helices, but composed of a single GMC-oxidoreductase. This led us to propose a model to  
343 explain the co-existence of both structures. In this model, the genome is folded into six  
344 strands before assembly into a 6-start helix. The first and the last strand can be shorter than  
345 the other strands, leading to a 5- start helix with 5 DNA strands and a 4-start helix with 4  
346 DNA strands. Finally, the deletion of both GMC-oxidoreductases with no significant loss of  
347 viral fitness demonstrated they were not essential.

348 The qu\_143 GMC oxidoreductase is 5 times less abundant than qu\_946 in the  
349 genomic fiber but it is ranked 1<sup>st</sup> in the layer of fibrils and is 3 times more abundant than the  
350 next one, qu\_757. It was previously predicted as the main target for glycosylation (4, 17, 18)  
351 while qu\_946 is only ranked 13<sup>th</sup> in the fibrils (Table S2). MS-based proteomics revealed that  
352 the cys-pro-rich N-ter domain of the GMC-oxidoreductase is present in the fibrils and cleaved  
353 in the genomic fiber (1), but the protease and precise cleavage site remains to be identified.  
354 Interestingly, an additional mutant in which this cys-pro-rich N-ter domain was fused to GFP,  
355 produced virions in which the GFP was identified at the surface of the virions and is enriched  
356 in the surface layer of fibrils. Further studies will be needed to elucidate the mechanism  
357 behind the recognition of this domain and its addressing to the fibrils. In addition, the virions  
358 of all mutants were decorated by long, reticulated fibrils as evidenced by TEM (Fig. S8).  
359 Except for the KO\_qu946 mutant, MS-based proteomic analysis of the fibrils revealed they  
360 were composed by other proteins some of which also presenting a cys-pro-rich N-ter domain  
361 (Fig. S6). Finally, the glycan analysis of the GMC-oxidoreductases single and double mutants  
362 confirmed that their fibrils were still glycosylated by the same two polysaccharides as the wt  
363 virions.

364 In a previous study, it was reported that the glycosylation machinery was clade-  
365 specific and produced different glycans (3). In the present study, the analysis of the purified  
366 fibrils of members of the B and C clades revealed they were also composed by different  
367 proteins inter- and even intra-clade and that B and C clades virions' fibrils present a protein  
368 composition closer to each other than that of A clade.

369 The glycosylated layer of fibrils bear glycans echoing bacterial ones, recognized by  
370 the amoeba which feeds on bacteria (2, 3). They thus appear as key for productive infection.  
371 In a given environment, the variable composition of the fibrils could be used to favor the  
372 engulfment of a given virion by a given amoeba, compared to others virions and even  
373 bacteria. The protein content of the layer of fibrils appears to have been optimized in a given  
374 clade, but their complex composition suggests they can be made from a large set of diverse  
375 proteins. Having a flexible toolbox for building the external layer of fibrils would reflect the  
376 need to constantly adjust the capsid composition to outcompete other parasites and secure  
377 infection. We can thus hypothesize that in the population of virions resulting from a single  
378 infection, this composition might also be variable, helping out the virus to ensure the

379 productive infection of at least one of the possible hosts present in that environment by at  
380 least one virion.

381 The question whether it is always the same protein that makes both the layer of fibril  
382 and the genomic fiber, even when other proteins than the GMC-oxidoreductases are used,  
383 remains unanswered. Yet, the most abundant protein composing the external layer of fibrils of  
384 the 2KO and M4 corresponds to the 222 amino-acids qu\_734 protein, which according to the  
385 high confidence alphaFold (28) prediction could be almost twice smaller (~4.5 nm) than the  
386 GMC-oxidoreductase. From the previous study, the central part of the genomic fiber has to  
387 correspond to a central channel of at least 9 nm to accommodate proteins such as the RNA  
388 polymerase (1). There is 4 additional nm for the DNA lining the protein shell and the GMC-  
389 oxidoreductase is 8 nm height and makes the 16 nm helical shell, leading to a ~29 nm helix.  
390 The protein making the shell is thus defining the final dimension of the genomic fiber. For the  
391 2KO, if this is the same protein that makes the external layer of fibrils and the shell of the  
392 genomic fiber, we expect a genomic fiber of 22 nm, compatible with the thinner structure  
393 observed in Fig 1D (~20-25 nm).

394 We have previously proposed an increased genome redundancy as a contributing  
395 factor to the appearance of viral gigantism (29). The data presented here validate such premise  
396 and extend these predictions outside of the *Pandoraviridae*. Overall, our results reveal the  
397 resilience of mimivirus, with redundant solutions securing essential functions such as  
398 infectiousness and genome packaging. Functional redundancy, well documented in the  
399 cellular world as a way to preserve essential function such as cell division (30), and until now  
400 a hallmark of the optimized microorganisms, may thus also be at work in the viral world.

## 401 **Material and methods**

### 402 *Cloning of DNA constructs*

403 A detailed protocol for gene manipulation of giant viruses and their host is provided in (19).

404 *Gene knock-out vectors.* The plasmid for gene knock-out was generated by sequential cloning  
405 of the 3' UTR of mg\_18 (megavirus chiliensis), the promoter of mg\_741 (megavirus chiliensis),  
406 and a Nourseothricin N-acetyl transferase (NAT) or a neomycin selection cassette (NEO). Each  
407 cloning step was performed using the Phusion Taq polymerase (ThermoFisher) and InFusion  
408 (Takara). Finally, 500bp homology arms were introduced at the 5' and 3' end of the cassette to

409 induce homologous recombination with the viral DNA (19). Before transfection, plasmids were  
410 digested with EcoRI and NotI. All primers are shown in FigS1.

#### 411 ***Establishment of viral lines.***

412 *Gene knock-out.* Gene knockout strategy was performed as previously described for  
413 pandoravirus (29). Briefly,  $1.5 \times 10^5$  *Acanthamoeba castellanii* cells were transfected with 6  $\mu$ g  
414 of linearized plasmid using Polyfect (QIAGEN) in phosphate saline buffer (PBS). One hour  
415 after transfection, PBS was replaced with PPYG, and cells were infected with  $1.5 \times 10^7$   
416 mimivirus reunion particles for 1 hour with sequential washes to remove extracellular virions.  
417 24h after infection the new generation of viruses (P0) was collected and used to infect new  
418 cells. An aliquot of P0 viruses was utilized for genotyping to confirm the integration of the  
419 selection cassette. Primers used for genotyping are shown in Table S1. A new infection was  
420 allowed to proceed for 1 hour, then washed to remove extracellular virions and nourseothricin  
421 was added to the media. Viral growth was allowed to proceed for 24 hours. This procedure was  
422 repeated one more time before removing the nourseothricin selection to allow viruses to expand  
423 more rapidly. Once, the viral infection was visible, the selection procedure was repeated one  
424 more time. Viruses produced after this new round of selection were used for genotyping and  
425 cloning (19). Double knockout of the GMC oxidoreductases was obtained by using a clonal  
426 population of qu\_143 knockout viruses as parental strain. The locus of qu\_946 was replaced by  
427 a neomycin resistance cassette. The transfection and selection of recombinant viruses'  
428 procedure was performed identical to the process to generate single knockout but replacing  
429 nourseothricin by geneticin.

430 *Cloning and genotyping.* 150,000 *A. castellanii* cells were seeded on 6 well plates with 2 mL  
431 of PPYG. After adhesion, viruses were added to the well at a multiplicity of infection (MOI) of  
432 1. One-hour post-infection, the well was washed 5 times with 1mL of PPYG, and cells were  
433 recovered by well scraping. Amoebas were then diluted until obtaining a suspension of 1  
434 amoeba/ $\mu$ L. 1 $\mu$ L of such suspension was added in each well of a 96-well plate containing 1000  
435 uninfected *A. castellanii* cells and 200  $\mu$ L of PPYG. Wells were later monitored for cell death  
436 and 100  $\mu$ L collected for genotyping (19). Genotyping was performed using Terra PCR Direct  
437 Polymerase Mix (Takara) following the manufacturer's specifications. Primers used for  
438 genotyping are shown in Table S1.

#### 439 ***Competition assay and quantitative PCR analysis.***

440 Equal infectious particles of wild-type and recombinant mimivirus reunion were mixed and  
441 used to infect *A. castellanii* at an approximate multiplicity of infection of 0.8. Viruses were  
442 allowed to grow overnight in the presence or absence of nourseothricin. Subsequent viral  
443 progenies were used to infect new *A. castellanii* cells in reiterative passages. A fraction of each  
444 passage was collected for genomic DNA extraction.

445 Viral genomes were purified using Wizard genomic DNA purification kit (PROMEGA). To  
446 determine the amplification kinetic, the fluorescence of the EvaGreen dye incorporated into the  
447 PCR product was measured at the end of each cycle using SoFast EvaGreen Supermix 2× kit  
448 (Bio-Rad, France). A standard curve using the gDNA of purified viruses was performed in  
449 parallel with each experiment. For each point, a technical triplicate was performed. Quantitative  
450 real-time PCR (qRT-PCR) analyses were performed on a CFX96 Real-Time System (Bio-Rad).

#### 451 ***Genome sequencing and assembly of mutants' genomes***

452 Genomic DNA was extracted from 10<sup>10</sup> virus particles using the PureLink TM Genomic DNA  
453 mini kit (Invitrogen) according to the manufacturer's protocol. Clones of individual mutants  
454 and wt were sequenced on an illumina platform (Novogen). For mimivirus wt, we obtained  
455 4,819,885 150nt paired-end reads. For KO\_qu946, KO\_qu946 and 2KO, 4,662,744, 4,572,873  
456 and 5,068,030 150nt paired-end reads, respectively. Genomes were assembled using spades  
457 v3.13, with the option "careful" and we obtained 5 contigs for the wt, 4 for KO\_qu946, and 7  
458 for KO\_qu143 and 2KO. The wt sequence was consistent with the original mimivirus reunion  
459 genome sequence (GI MW004169) and for the clone KO\_qu946 the gene was interrupted from  
460 position 1,151,358 to 1,152,985 relative to wt genome (position 302 to 1741 in the  
461 corresponding qu\_946 gene) and for the clone KO\_qu143 the gene was interrupted from  
462 position 165,934 to 167,572 relative to wt genome (position 306 to 1741 in the corresponding  
463 qu\_143 gene). The region is replaced in both cases by the nourseothricin cassette sequence  
464 (1637 nt). For the 2KO, the gene qu\_143 is interrupted from positions 306 to 1741 (165,934 to  
465 167,572 on the wt genome) and replaced by the nourseothricin cassette sequence (1637 nt), and  
466 gene qu\_946 is interrupted from positions 302 to 1206 (1,151,551 to 1,153,410 on the genome)  
467 and replaced by the geneticin cassette sequence (1859 nt). To confirm the deletion of each  
468 mutant, the reads were mapped on the wt genome resulting in homogeneous coverage along the  
469 genome, except for the qu\_143 and qu\_946 central positions which are covered. In addition,  
470 we used the central part of the GMC-oxidoreductase genes (deleted in mutants) as blast queries

471 against the reads of each mutant genome, which also confirmed the absence of the central  
472 region.

### 473 ***Extraction and purification of the qu\_946 and qu\_143 mutants' genomic fiber***

474 The genomic fiber of the mimivirus reunion single mutants of qu\_946 (KO\_qu946) and qu\_143  
475 (KO\_qu143), were extracted as described in Villalta et al, 2022 for the wt virus (1). The  
476 genomic fiber was extracted from 12 mL of purified single deletion mutant virions at  $2 \times 10^{10}$   
477 particles/mL, split into 12 x 1 mL samples processed in parallel. Trypsin (Sigma T8003) in 40  
478 mM Tris-HCl pH 7.5 buffer was added at a final concentration of 50  $\mu$ g/mL and the virus-  
479 enzyme mix was incubated for 2h at 30°C in a heating dry block (Grant Bio PCH-1). DTT was  
480 then added at a final concentration of 10 mM and incubated at 30°C for 16h. Finally, 0.001%  
481 Triton X-100 was added to the mix and incubated for 4h at 30°C. Each tube was centrifuged at  
482 4,000 x g for 5 min to pellet the opened capsids. The supernatant was recovered and  
483 concentrated by centrifugation at 15,000 x g for 4h at 4°C. Most of the supernatant was  
484 discarded leaving 12x~200  $\mu$ L of concentrated broken pieces of genomic fiber that were pooled  
485 and layered on top of ultracentrifuge tubes of 4 mL (polypropylene centrifuge tubes, Beckman  
486 Coulter) containing a discontinuous cesium chloride gradient (1.4, 1.3, 1.2 g/cm<sup>3</sup> in 40 mM  
487 Tris-HCl pH 7.5 buffer). The gradients were centrifuged at 200,000 x g for 16h at 4 °C. Since  
488 no visible band was observed, successive 0.5 mL fractions were recovered from the bottom of  
489 the tube. Each fraction was dialyzed using 20 kDa Slide-A-Lyzers (ThermoFisher) against 40  
490 mM Tris-HCl pH 7.5 to remove the CsCl. These fractions were further concentrated by  
491 centrifugation at 15,000 x g, at 4°C for 4h, and most of the supernatant was removed, leaving  
492 ~100  $\mu$ L of sample at the bottom of each tube. At each step of the extraction procedure, the  
493 sample was imaged by negative staining transmission electron microscopy (NS-TEM) to assess  
494 the integrity of the genomic fiber. Each fraction of the gradient was finally controlled by NS-  
495 TEM.

### 496 ***Negative staining TEM***

497 300 mesh ultra-thin carbon-coated copper grids (Electron Microscopy Sciences, EMS) were  
498 prepared for negative staining by adsorbing 4-7  $\mu$ L of the sample for 3 min., followed by  
499 blotting excess liquid and staining for 2 min in 2% uranyl acetate to image the genomic fiber.  
500 For fibrils and mutant virions, staining was performed with a drop of 1% uranyl followed by  
501 blotting after 10-15 s, and a drop of uranyl acetate coupled with methylcellulose (2% and 0.2%,  
502 respectively) was added twice and left for 5-10 s before blotting.

503 The grids were imaged either on an FEI Tecnai G2 microscope operated at 200 keV and  
504 equipped with an Olympus Veleta 2k camera (IBDM microscopy platform, Marseille, France);  
505 an FEI Tecnai G2 microscope operated at 200 keV and equipped with a Gatan OneView camera  
506 (IMM, microscopy platform, France).

### 507 *Single-particle analysis by cryo-EM*

#### 508 *Sample preparation*

509 For the KO\_qu946 mutant, 3  $\mu\text{L}$  of the purified sample were applied to glow-discharged  
510 Quantifoil R 2/1 Cu grids, blotted for 2 s using a Vitrobot Mk IV (Thermo Scientific) and  
511 applying the following parameters: 4°C, 100% humidity, blotting force 0, and plunge frozen in  
512 liquid ethane/propane cooled to liquid nitrogen temperature.

#### 513 *Data acquisition*

514 Grids were imaged using a Titan Krios (Thermo Scientific) microscope operated at 300 keV  
515 and equipped with a K3 direct electron detector and a GIF BioQuantum energy filter (Gatan).  
516 2,224 movie frames were collected using the EPU software (Thermo Scientific) at a nominal  
517 magnification of 81,000x with a pixel size of 1.0859 Å and a defocus range of -0.6 to -2.8  
518  $\mu\text{m}$ . Micrographs were acquired using EPU (Thermo Scientific) with 2.3 s exposure time,  
519 fractionated into 40 frames, and 18.25  $\text{e}^-/\text{pixel}/\text{s}$  (total fluence of 35.597  $\text{e}^-/\text{Å}^2$ ).

#### 520 *2D classification and clustering of 2D classes*

521 All movie frames were aligned using MotionCor2 (31) and used for contrast transfer function  
522 (CTF) estimation with CTFFIND-4.1 (32). Helical segments of the purified genomic fiber  
523 manually picked with Relion 3.1.0, were extracted with 400 pixels box sizes (decimated to 100  
524 pixels) using a rise of 7.93 Å and a tube diameter of 300 Å. Particles were subjected to  
525 reference-free 2D classification in Relion 3.1.0 (33, 34). We then performed additional cluster  
526 analysis of the initial 2D classes provided by Relion to aim for more homogeneous clusters (1),  
527 eventually corresponding to different states (Fig. 3 & 4).

#### 528 *Identification of candidate helical parameters*

529 Fourier transform analysis methods have been used to confirm the helical parameters were the  
530 same as in wt (1, 35–37) for the Cl1 (Cl1a in wt) and Cl2.

#### 531 *Cryo-EM data processing and 3D reconstruction*

532 After helical parameters determination, segments were extracted with a box size of 400 pixels  
533 (decimated to 100 pixels) using the proper rises for the C11 (7.93 Å, cylinder 300 Å, 442,237  
534 segments) and the C12 (20.47 Å, cylinder 330 Å, 172,431 segments). A dedicated 2D  
535 classification protocol was performed independently on each extraction. For the C12, one round  
536 of 50 expectation-maximization (E-M) iterations was sufficient to produce 133 homogeneous  
537 2D classes submitted to cluster analysis and 44 2D classes were selected (21,186 segments).  
538 For the C11 extraction, three iterative 2D classification/selection rounds were performed (25,  
539 50 and 100 E-M iterations) producing 61 classes from which 30 (149,593 segments) were  
540 finally selected based on the cluster analysis.

541 Values of the helical parameters, rise and twist (C11: 7.9475 Å, 138.921°; C12: 20.48 Å,  
542 49.43°), were then used for Relion 3D classification (33, 34), with a +/-0.5 units freedom search  
543 range, using a featureless cylinder as initial reference (diameter of 300 Å and C1 symmetry for  
544 C11 and 330 Å and C3 symmetry for C12). In real space, the helical symmetry was searched  
545 using 50% of the central part of the box for both structures. The number of asymmetrical units  
546 in each segment box was set to 1 for C11 and 6 for C12. The entire helical reconstruction and  
547 protein shell dimensions were obtained using an in-house developed program.

548 The superimposable 3D classes (same helical parameters, same helix orientation) were then  
549 selected, reducing the data set to 98,882 segments for the 5-start helix (C11) and 20,899  
550 segments for the 6-start helix (C12). After re-extraction of the selected segments without  
551 scaling, a first unmasked 3D refinement was performed with a rescaled 3D classification output  
552 low pass filtered to 15 Å as reference, followed by a 3D refinement step using solvent flattened,  
553 FSCs and CTF refinement using the standard procedure described in Relion. To further improve  
554 the resolution of the maps, Bayesian polishing was applied using 10,000 segments for training  
555 and default values for polishing. A last round of 3D refinement with solvent flattening was  
556 applied to the previous refined map using the polished particles. At that point, the maps were  
557 resolved enough (C11: 4.3 Å, C12: 4.2; FSC threshold 0.5) to identify secondary structure  
558 elements (with visible cylinders corresponding to the helices) and were used to compute local  
559 resolution.

#### 560 *Structures refinement*

561 The best resolution C12 map was used to fit the qu\_143 dimeric structure (PDB 7YX3) using  
562 UCSF ChimeraX 1.5 (38). Each monomer was then rigid-body fitted independently into the  
563 map. The entire protein was then inspected within coot 0.9.7 (39) to fix local inconsistencies

564 and was further refined against the map using the real-space refinement program in PHENIX  
565 1.20.1 (40). The protein was submitted to 5 cycles of rigid body refinement (with each chain  
566 defined) followed by twice 10 cycles of refinement with the default PHENIX options. The  
567 resulting structure was manually corrected within coot. The resulting protein model was  
568 submitted to the same steps refinement in PHENIX. This final model was then fitted into the  
569 C11 map, inspected with coot and refined using 5 cycles of rigid body refinement and simulated  
570 annealing followed by twice 10 cycles of refinement with the default PHENIX options.  
571 Validations were performed with PHENIX using the comprehensive validation program (Table  
572 S3). RMSD between different structures (monomers and dimers, Table S4) were computed  
573 using the align procedure in Pymol suite (Schrodinger, L., & DeLano, W. (2020). *PyMOL*).

#### 574 ***Extraction and purification of the mutants' fibrils***

575 To analyze the glycan composition and polysaccharides structures of the 3 mutants and wt  
576 mimivirus reunion strain we applied an already described protocol (4). Briefly,  $4 \times 10^{11}$  viral  
577 particles were centrifuged at 14,000g for 10 min, the supernatant was discarded and the pellet  
578 was re-suspended in 10 ml of 50 mM DTT. Fibril extraction was performed at 100°C under  
579 stirring. After 2 h, the tube was centrifuged at 14,000 g for 15 min, at 4°C, and the fibrils  
580 recovered with the supernatant. The fibrils were then dried and purified on Biogel P10, followed  
581 by subsequent NMR analysis of each mutant and wt.

582 We also developed a softer defibrillation protocol to recover the fibrils without contaminating  
583 them with proteins released by damaged virions in order to analyze the fibrils protein  
584 composition by MS-based proteomics. Purified virions ( $1.5 \times 10^{10}$ ) were incubated in Eppendorf  
585 tubes in 500  $\mu$ L 40 mM Tris-HCl pH 7.5 buffer, 500 mM DTT for 2h at 30°C. Tubes were then  
586 centrifuged at 14,000 g for 10 min. The supernatants containing the fibrils were recovered and  
587 concentrated on vivaspin® 3 KDa (Sartorius, VS04T91) at 3,000 g. The pellet was washed  
588 twice with 40 mM Tris-HCl pH 7.5 buffer and centrifuged at 14,000 g for 10 min and finally  
589 resuspended in the same buffer. Intact virions, pellets and fibrils were imaged by NS-TEM to  
590 assess the integrity of the defibrillated virions in the pellet and the presence of fibrils in the  
591 supernatant. For the Nqu143-GFP mutants, defibrillated virions were also observed by  
592 fluorescence microscopy which confirmed the absence of fluorescence due to the removal of  
593 the GFP together with the layer of fibrils.

#### 594 ***Sugar composition of viral particles of Mimivirus reunion wt and mutants***

595 Monosaccharide composition analyses as acetylated methyl glycoside (AMG) were performed  
596 on the intact viral particles ( $1,25 \times 10^{10}$ ,  $\sim 250 \mu\text{l}$ ) of mimivirus reunion wt and mutants,  
597 following the procedure reported by De Castro *et al* 2010 (41). The obtained AMG were  
598 analyzed via gas chromatography-mass spectrometry (GC-MS) on an Agilent instrument (GC  
599 instrument Agilent 6850 coupled to MS Agilent 5973) equipped with a SPB-5 capillary column  
600 (Supelco,  $30 \text{ m} \times 0.25 \text{ i.d.}$ , flow rate,  $0.8 \text{ mL min}^{-1}$ ) and He as the carrier gas. The identification  
601 of the monosaccharides derivatized as AMG, was obtained by studying the fragmentation  
602 pattern corresponding to each peak of the chromatogram and by comparison with suitable  
603 standards.

#### 604 *Purification and $^1\text{H}$ NMR analysis of the fibrils*

605 The fibrils of mimivirus reunion wt and mutants, extracted as reported above, were purified to  
606 remove the DTT used for the extraction procedure.

607 Briefly, the glycoproteins (protein/s carrying the polysaccharides) were precipitated with cold  
608 acetone at 80%, at  $4 \text{ }^\circ\text{C}$ , for 16 hours, twice. The supernatant containing DTT and salts was  
609 discarded, while the precipitate was dissolved in water and freeze-dried. Then, the precipitate  
610 was purified by size exclusion chromatography (Biogel P10, flow:  $12 \text{ ml / h}$ ) to completely  
611 remove the DTT. The eluted fractions were checked by  $^1\text{H}$  NMR, revealing that the glycan-  
612 containing material was eluted at one-third of the column volume (full spectra are shown in Fig  
613 7B).

614 The  $^1\text{H}$  NMR measurements were carried out on 600 MHz Bruker instrument, equipped with a  
615 CryoProbe™ at 310 K. The intensity of the solvent signal was reduced by measuring a mono-  
616 dimensional DOSY spectrum, setting  $\delta$  and  $\Delta$  to 2.4 ms and 100 ms, respectively, and the  
617 variable gradient to 50% of its maximum power. Spectra were processed and analyzed using  
618 Bruker TopSpin 4.0.9 program.

619 In the  $^1\text{H}$  NMR spectra, the anomeric region (22) (5.5-4.4 ppm) perfectly overlapped with that  
620 of the previously studied mimivirus strain (4) (Fig 7B, Fig 7-figure supplement 7), confirming  
621 that the different sugars were assembled the same way as in the reference mimivirus to produce  
622 the same two polysaccharides (Fig 7C). The signal at 4.80 ppm is the result of the overlap of  
623 two anomeric protons, the one of a 3- $\alpha$ -Rha (labeled **C**) and of a 3- $\beta$ -GlcNAc (**D**) modified  
624 with a pyruvic acid at the hydroxyl functions 4 and 6. These two residues are the building blocks  
625 of the repeating unit of polysaccharide 1 (**poly\_1**, **Fig 7C**) (4). The other anomeric signals,

626 labeled with the capital letters **A**, **A'** (2,3- $\alpha$ - L-Rha), **B** (3- $\beta$ -D-GlcNAc), **E** (2OMe- $\beta$ -D-  
627 VioNAc) and **E'**( $\beta$ -D-Vio4NAc), are part of the polysaccharide 2 (**poly\_2**) that presents a  
628 backbone with a disaccharide repeating unit of 2)- $\alpha$ -L-Rha-(1 $\rightarrow$ 3)- $\beta$ -D-GlcNAc-(1 $\rightarrow$ , with the  
629 rhamnose residue further substituted with a viosamine methylated at position 2 (**E**) or not  
630 methylated (**E'unit**), thus taking the labels **A** and **A'**, respectively (Fig 7C) (4).

### 631 **Mass spectrometry-based proteomic analyses**

632 Proteins were solubilized with Laemmli buffer (4 volumes of sample with 1 volume of Laemmli  
633 5X - 125 mM Tris-HCl pH 6.8, 10% SDS, 20% glycerol, 25%  $\beta$ -mercaptoethanol and traces of  
634 bromophenol blue) and heated for 10 min at 95 °C. The extracted proteins were stacked in the  
635 top of an SDS-PAGE gel (4-12% NuPAGE, Life Technologies), stained with Coomassie blue  
636 R-250 (Bio-Rad) before in-gel digestion using modified trypsin (Promega, sequencing grade)  
637 as previously described (42). The resulting peptides were analyzed by online nanoliquid  
638 chromatography coupled to tandem MS (UltiMate 3000 RSLCnano and Q-Exactive Plus or Q-  
639 Exactive HF, Thermo Scientific). Peptides were sampled on a 300  $\mu$ m x 5 mm PepMap C18  
640 precolumn and separated on a 75  $\mu$ m x 250 mm C18 column (Reprosil-Pur 120 C18-AQ, 1.9  
641  $\mu$ m, Dr. Maisch, except for KO\_qu143, KO\_qu946 and Nqu143-GFP mutant samples separated  
642 on Aurora, 1.7  $\mu$ m, IonOpticks) using a 140-min gradient (except for fibrils from Nqu143-GFP  
643 mutant for which a 60-min gradient was used). MS and MS/MS data were acquired using  
644 Xcalibur 4.0 (Thermo Scientific). Peptides and proteins were identified using Mascot (version  
645 2.8.0, Matrix Science) through concomitant searches against homemade *A. castellanii* protein  
646 sequence database, homemade virus-specific protein sequence databases, and a homemade  
647 database containing the sequences of classical contaminant proteins found in proteomic  
648 analyses (human keratins, trypsin...). Trypsin/P was chosen as the enzyme and two missed  
649 cleavages were allowed. Precursor and fragment mass error tolerances were set at respectively  
650 at 10 and 20 ppm. Peptide modifications allowed during the search were: Carbamidomethyl (C,  
651 fixed), Acetyl (Protein N-term, variable) and Oxidation (M, variable). The Proline software  
652 version 2.2.0 (43) was used for the compilation, grouping, and filtering of the results:  
653 conservation of rank 1 peptides, peptide length  $\geq$  6 amino acids, peptide-spectrum-match  
654 identification false discovery rate  $<$  1% (44), and minimum of 1 specific peptide per identified  
655 protein group. Proline was then used to perform a MS1-based quantification of the identified  
656 protein groups. Intensity-based absolute quantification (iBAQ) (45) values were calculated for  
657 each protein group from the MS intensities of razor and specific peptides (Table S2). The

658 relative abundance of individual proteins in virions and fibrils was calculated as the ratio of the  
659 individual protein iBAQ values to the sum of the iBAQ values of all proteins in each sample.  
660 The relative enrichment of individual proteins between virions and fibrils was calculated as the  
661 ratio of their relative abundances in each fraction (1).

662 **Acknowledgments:** We thank Jean-Michel Claverie for his comments on the manuscript and  
663 discussions all along the project and Elsa Garcin for editing the revised manuscript. The data  
664 were collected at the Cryo-EM Swedish National Facility funded by the Knut and Alice  
665 Wallenberg, Family Erling Persson and Kempe Foundations, SciLifeLab, Stockholm  
666 University, and Umeå University under the expert assistance of Dr. Matthieu Coinçon. The  
667 preliminary electron microscopy experiments were performed on the PiCSL-FBI core facility  
668 (Aïcha Aouane, IBDM, AMU-Marseille), a member of the France-BioImaging national  
669 research infrastructure, sample freezing conditions and cryo-EM preliminary acquisitions  
670 were performed at the AFMB microscopy platform and on the IMM imaging platform (Dr.  
671 Artemis Kosta). We thank the PACA Bioinfo platform for computing support.

672 **Funding:** This project has received funding from the European Research Council (ERC)  
673 under the European Union's Horizon 2020 research and innovation program (grant agreement  
674 No 832601). Proteomic experiments were partly supported by ProFI (ANR-10-INBS-08-01)  
675 and GRAL, a program from the Chemistry Biology Health (CBH) Graduate School of  
676 University Grenoble Alpes (ANR-17-EURE-0003). France-BioImaging national research  
677 infrastructure (ANR-10-INBS-04). CDC gratefully acknowledges STARPLUS 2020 (project  
678 no. 21-UNINA-EPIG-042) from the University of Napoli for financial support.

679 **Data and materials availability:** Genome sequences of the mutants of mimivirus reunion  
680 have been deposited to NCBI (accession numbers: KO\_qu\_946: OQ700912; KO\_qu143:  
681 OQ700913, 2KO: OQ700914). 3D reconstruction maps and the corresponding PDB have  
682 been deposited to EMDB (Deposition number C11: 8ORS, EMD-17131; C12: 8ORH, EMD-  
683 17125). The mass spectrometry proteomics data have been deposited to the ProteomeXchange  
684 Consortium via the PRIDE (46) partner repository with the dataset identifier PXD041298.  
685 AlphaFold predictions were performed using HPC/AI resources from GENCI-IDRIS (Grant  
686 2022-AD011013526).

687 **Authors contributions:** C.A. conceived and designed the research; J-M.A., H.B., A.V., S.S.,  
688 A.L., A.S., C.B. A.N., L.B., A.A., O.P. C.D.C., Y.C., and C.A. performed research; D.P.  
689 granted access to microscopy platform and helped optimizing sample freezing with A.V.; J-

690 M. A and A.L. developed methodologies for genomic fiber and fibrils extraction and  
691 purification; H.B. developed the genetic tools and analyzed mutants; C.B. and A.L. built the  
692 Nqu143-GFP mutant and analyzed the data; A.V., collected CryoEM data; A.V., S.S., A.S.,  
693 and C.A. analyzed CryoEM data Y.C., L.B., A.A., performed MS-based proteomics and  
694 analyzed the data; A.N. and C.D.C., performed GC-MS and NMR analyses and analyzed the  
695 data, H.B., A.V., A.S., S.S. A.N., C.D.C., Y.C. and C.A. wrote the manuscript.

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

## 697 **References**

- 698 1. A. Villalta, *et al.*, The giant mimivirus 1.2 Mb genome is elegantly organized into a 30-nm  
699 diameter helical protein shield. *eLife* **11**, e77607 (2022).
- 700 2. R. A. L. Rodrigues, *et al.*, Mimivirus Fibrils Are Important for Viral Attachment to the Microbial  
701 World by a Diverse Glycoside Interaction Repertoire. *J. Virol.* **89**, 11812–11819 (2015).
- 702 3. A. Notaro, *et al.*, Giant viruses of the *Megavirinae* subfamily possess biosynthetic pathways to  
703 produce rare bacterial-like sugars in a clade-specific manner. *microLife* **3**, uqac002 (2022).
- 704 4. A. Notaro, *et al.*, Expanding the Occurrence of Polysaccharides to the Viral World: The Case of  
705 Mimivirus. *Angew. Chem. Int. Ed.* **60**, 19897–19904 (2021).
- 706 5. E. V. Koonin, *et al.*, Global Organization and Proposed Megataxonomy of the Virus World.  
707 *Microbiol. Mol. Biol. Rev.* **84**, e00061-19 (2020).
- 708 6. D. Raoult, *et al.*, The 1.2-megabase genome sequence of Mimivirus. *Science* **306**, 1344–1350  
709 (2004).
- 710 7. C. Abergel, J. Rudinger-Thirion, R. Giegé, J.-M. Claverie, Virus-encoded aminoacyl-tRNA  
711 synthetases: structural and functional characterization of mimivirus TyrRS and MetRS. *J. Virol.*  
712 **81**, 12406–12417 (2007).
- 713 8. J. Abrahão, *et al.*, Tailed giant Tupanvirus possesses the most complete translational apparatus  
714 of the known virosphere. *Nat. Commun.* **9**, 749 (2018).
- 715 9. I. Speciale, *et al.*, The Astounding World of Glycans from Giant Viruses. *Chem. Rev.* **122**, 15717–  
716 15766 (2022).
- 717 10. N. Zauberman, *et al.*, Distinct DNA exit and packaging portals in the virus *Acanthamoeba*  
718 *polyphaga* mimivirus. *PLoS Biol.* **6**, e114 (2008).
- 719 11. J. R. Schrad, J. S. Abrahão, J. R. Cortines, K. N. Parent, Structural and Proteomic Characterization  
720 of the Initiation of Giant Virus Infection. *Cell* **181**, 1046-1061.e6 (2020).
- 721 12. J.-M. Claverie, C. Abergel, Mimivirus: the emerging paradox of quasi-autonomous viruses.  
722 *Trends Genet.* **26**, 431–437 (2010).

- 723 13. M. Suzan-Monti, B. L. Scola, L. Barrassi, L. Espinosa, D. Raoult, Ultrastructural Characterization  
724 of the Giant Volcano-like Virus Factory of *Acanthamoeba polyphaga* Mimivirus. *PLoS ONE* **2**,  
725 e328 (2007).
- 726 14. Y. Mutsafi, N. Zauberman, I. Sabanay, A. Minsky, Vaccinia-like cytoplasmic replication of the  
727 giant Mimivirus. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 5978–5982 (2010).
- 728 15. Y. G. Kuznetsov, T. Klose, M. Rossmann, A. McPherson, Morphogenesis of mimivirus and its  
729 viral factories: an atomic force microscopy study of infected cells. *J. Virol.* **87**, 11200–11213  
730 (2013).
- 731 16. I. L. M. de Aquino, *et al.*, Diversity of Surface Fibril Patterns in Mimivirus Isolates. *J. Virol.*,  
732 e01824-22 (2023).
- 733 17. T. Klose, *et al.*, A Mimivirus Enzyme that Participates in Viral Entry. *Struct. Lond. Engl.* **1993** **23**,  
734 1058–1065 (2015).
- 735 18. M. Boyer, *et al.*, Mimivirus shows dramatic genome reduction after intraamoebal culture. *Proc.*  
736 *Natl. Acad. Sci. U. S. A.* **108**, 10296–10301 (2011).
- 737 19. N. Philippe, A. Shukla, C. Abergel, H. Bisio, Genetic manipulation of giant viruses and their host,  
738 *Acanthamoeba castellanii*. *Nat. Protoc.* (2023) <https://doi.org/10.1038/s41596-023-00910-y>.
- 739 20. L. Gallot-Lavallée, G. Blanc, J.-M. Claverie, Comparative Genomics of Chrysochromulina Ericina  
740 Virus and Other Microalga-Infecting Large DNA Viruses Highlights Their Intricate Evolutionary  
741 Relationship with the Established Mimiviridae Family. *J. Virol.* **91**, pii: e00230-17 (2017).
- 742 21. H. Takahashi, S. Fukaya, C. Song, K. Murata, M. Takemura, Morphological and taxonomic  
743 properties of the newly isolated *Cotonvirus japonicus*, a new lineage of the subfamily  
744 *Megavirinae*. *J. Virol.* (2021) <https://doi.org/10.1128/JVI.00919-21> (July 26, 2021).
- 745 22. I. Speciale, *et al.*, Liquid-state NMR spectroscopy for complex carbohydrate structural analysis:  
746 A hitchhiker’s guide. *Carbohydr. Polym.* **277**, 118885 (2022).
- 747 23. C. Hacker, *et al.*, Nanoparticle suspensions enclosed in methylcellulose: a new approach for  
748 quantifying nanoparticles in transmission electron microscopy. *Sci. Rep.* **6**, 25275 (2016).
- 749 24. M. C. Cassetti, M. Merchlinsky, E. J. Wolffe, A. S. Weisberg, B. Moss, DNA Packaging Mutant:  
750 Repression of the Vaccinia Virus A32 Gene Results in Noninfectious, DNA-Deficient, Spherical,  
751 Enveloped Particles. *J. Virol.* **72**, 5769–5780 (1998).
- 752 25. Y. Liu, *et al.*, Virus-encoded histone doublets are essential and form nucleosome-like structures.  
753 *Cell* **184**, 4237-4250.e19 (2021).
- 754 26. M. I. Valencia-Sánchez, *et al.*, The structure of a virus-encoded nucleosome. *Nat. Struct. Mol.*  
755 *Biol.* **28**, 413–417 (2021).
- 756 27. H. Sobhy, A comparative review of viral entry and attachment during large and giant dsDNA  
757 virus infections. *Arch. Virol.* **162**, 3567–3585 (2017).
- 758 28. J. Jumper, *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–  
759 589 (2021).

- 760 29. H. Bisio, *et al.*, Evolution of giant pandoravirus revealed by CRISPR/Cas9. *Nat. Commun.* **14**, 428  
761 (2023).
- 762 30. H. P. Erickson, M. Osawa, Cell division without FtsZ—a variety of redundant mechanisms. *Mol.*  
763 *Microbiol.* **78**, 267–270 (2010).
- 764 31. S. Q. Zheng, *et al.*, MotionCor2: anisotropic correction of beam-induced motion for improved  
765 cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
- 766 32. A. Rohou, N. Grigorieff, CTFIND4: Fast and accurate defocus estimation from electron  
767 micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
- 768 33. S. He, S. H. W. Scheres, Helical reconstruction in RELION. *J. Struct. Biol.* **198**, 163–176 (2017).
- 769 34. S. H. W. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure  
770 determination. *J. Struct. Biol.* **180**, 519–530 (2012).
- 771 35. N. Coudray, *et al.*, Deducing the symmetry of helical assemblies: Applications to membrane  
772 proteins. *J. Struct. Biol.* **195**, 167–178 (2016).
- 773 36. R. Diaz, W. J. Rice, D. L. Stokes, Fourier-Bessel reconstruction of helical assemblies. *Methods*  
774 *Enzymol.* **482**, 131–165 (2010).
- 775 37. C. Sachse, Single-particle based helical reconstruction—how to make the most of real and  
776 Fourier space. *AIMS Biophys.* **2**, 219–244 (2015).
- 777 38. E. F. Pettersen, *et al.*, UCSF Chimera-A visualization system for exploratory research and  
778 analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- 779 39. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of *Coot*. *Acta*  
780 *Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
- 781 40. D. Liebschner, *et al.*, Macromolecular structure determination using X-rays, neutrons and  
782 electrons: recent developments in *Phenix*. *Acta Crystallogr. Sect. Struct. Biol.* **75**, 861–877  
783 (2019).
- 784 41. C. De Castro, M. Parrilli, O. Holst, A. Molinaro, “Microbe-Associated Molecular Patterns in  
785 Innate Immunity” in *Methods in Enzymology*, (Elsevier, 2010), pp. 89–115.
- 786 42. M. G. Casabona, Y. Vandenbrouck, I. Attree, Y. Couté, Proteomic characterization of  
787 *Pseudomonas aeruginosa* PAO1 inner membrane. *Proteomics* **13**, 2419–2423 (2013).
- 788 43. D. Bouyssié, *et al.*, Proline: an efficient and user-friendly software suite for large-scale  
789 proteomics. *Bioinforma. Oxf. Engl.* **36**, 3148–3155 (2020).
- 790 44. Y. Couté, C. Bruley, T. Burger, Beyond Target-Decoy Competition: Stable Validation of Peptide  
791 and Protein Identifications in Mass Spectrometry-Based Discovery Proteomics. *Anal. Chem.* **92**,  
792 14898–14906 (2020).
- 793 45. B. Schwanhäusser, *et al.*, Global quantification of mammalian gene expression control. *Nature*  
794 **473**, 337–342 (2011).

795 46. Y. Perez-Riverol, *et al.*, The PRIDE database resources in 2022: a hub for mass spectrometry-  
796 based proteomics evidences. *Nucleic Acids Res.* **50**, D543–D552 (2022).

797