No fitness impact of the knockout of the two main components of mimivirus genomic fiber and fibril layer

Authors: Jean-Marie Alempic[1]†, Hugo Bisio[1]†, Alejandro Villalta[1]†, Sébastien Santini[1], Audrey Lartigue[1], Alain Schmitt[1], Claire Bugnot[1#], Anna Notaro[2], Lucid Belmudes[3], Annie Adrait[3], Olivier Poirot[1], Denis Ptchelkine[4], Cristina De Castro[2], Yohann Couté[3], Chantal Abergel[1*]

Affiliations :

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

[2]Department of Agricultural Sciences, University of Naples Federico II, Via Università 100, 80055, Portici, Italy

[3]Univ. Grenoble Alpes, CEA, INSERM, UA13 BGE, CNRS, CEA, FR2048, 38000 Grenoble, France

[4]Aix–Marseille University, Centre National de la Recherche Scientifique, Architecture et Fonction des Macromolécules Biologiques, Unité Mixte de Recherche 7257 (IM2B), 13288 Marseille Cedex 9, France.

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

†These authors contributed equally to the work

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

One-Sentence Summary: Functional redundancy warrants mimivirus genomic fiber and fibrils formation.

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

Keywords: Mimivirus, giant virus, dsDNA, viral genome organization, genomic fiber, cryo-EM, helical reconstruction, MS-based proteomics.

Abstract

The Mimivirus 1.2Mb genome is organized into a nucleocapsid-like genomic fiber composed of a mixture of two GMC-oxidoreductase paralogs (1). Surprisingly, these proteins also constitute the glycosylated fibril layer that decorate the virion (2). The individual inactivation of each gene using homologous recombination-based knockout (KO) confirmed they were not essential and the resulting genomic fiber, as for the wild type, corresponds to a 5- and a 6-start helix composed of the remaining GMC-oxidoreductase. This result allowed to propose a model explaining the coexistence of these two helices. All mutants showed equivalent fitness in laboratory conditions and, as for the genomic fiber, there is no impact of the KO on the layer of fibrils surrounding the capsids which are still glycosylated with the same polysaccharides as the wild type. Expanding the study to additional members of the family (moumouviruses and megaviruses) revealed that, as for the glycans which are clade specific (3, 4), the main protein can also be different, a situation mimicked by the double KO of the GMC oxidoreductases in mimivirus. These results obtained on two distinct essential processes (genome packaging and virion infectivity) illustrate the exceptional functional redundancy of the *Mimiviridae*, which may be the major evolutionary force behind their giant genomes.

Significance statement

Exaptation and genetic redundancy are key factors in giant virus evolution. Exaptation is exemplified by the recruitment of metabolic enzymes, GMC oxidoreductases, for both DNA compaction and virion attachment. Despite their involvement in such essential processes these proteins can be lost during experimental evolution experiments. Here, we developed genetic tools that will allow addressing the function of thousands of genes in *Mimiviridae*. They were used to evaluate the essentially of GMC oxidoreductases for the genomic fiber and virion external fibril layer assembly. As a result, this study unveils an extensive toolkit of viral genes that can alternatively be used to secure essential functions as well as the unprecedented plasticity of mimivirus for virion assembly.

Introduction

Mimivirus is the prototype of the *Mimiviridae* family part of the *Nucleocytoviricota* phylum encompassing large and giant DNA viruses infecting eukaryotes (5). Members of the *Mimiviridae* family infecting amoeba have dsDNA genomes up to 1.5 Mb encoding over 1000 proteins, including complete glycosylation machinery (3, 4, 6–9). Mimivirus virions penetrate the cell through phagocytosis, and the acidic vacuole mediates the opening of the

stargate structure at one vertex of its icosahedral capsid (10, 11). The internal membrane unwraps and fuses with the vacuole membrane, allowing the transfer of the nucleoid compartment into the host cytoplasm, while empty capsids remain in the cell vacuole (6, 10, 12). The infectious cycle occurs in the cytoplasm where a large viral factory is developed (6, 12–14). At the late stage of the cycle, neo-synthesized virions bud at the periphery of the viral factory where they are filled by the genome. Finally, a glycosylated fibril layer, composed of two large polysaccharides synthesized by the virally encoded machinery, is added to the capsids (4, 15, 16). As a result, the 750 nm diameter virions appear as Russian dolls made of the surface layer of reticulated glycosylated fibrils decorating the icosahedral capsids. Inside, the nucleoid compartment encases the 1.2 Mb dsDNA genome organized into a 30 nm large genomic fiber with a nucleocapsid-like structure, made of a protein shell internally lined by the folded DNA, and the central channel can accommodate large proteins such as the viral RNA polymerase (1). Three independent genomic fiber structures have been determined by cryo-electron microscopy (cryo-EM): two compact 5- and 6-start, DNA containing helices, and a 5-start relaxed helix, without DNA (1). Unexpectedly, the protein shell was found to be composed of two glucose–methanol–choline (GMC) oxidoreductases sharing 69% sequence identity, and the resolution of the reconstructions (3.7 and 4.4 Å) prevented to determine whether each helix was composed of one paralog or made of a mixture of both. Interestingly, these two GMC-oxidoreductases also compose the external fibril layer of the capsids and it was hypothesized that they were the major target for glycosylation at the fibrils of mimivirus (4, 9, 17). However, despite their involvement in both the mimivirus genomic fiber structures and the virion surface fibril layer, homologs of these proteins are puzzlingly absent in the llaboratory-evolved mutant (strain M4) (17), which also lacks the glycosylation machinery responsible for synthesizing and branching the polysaccharides on the capsids. In the present study, we developed tools to knock out (KO) mimivirus genes in order to investigate the impact of the depletion of the GMC-oxidoreductases on the formation of both the genomic fiber and the fibril layer. We assessed the fitness of the different mutants compared to a wild type mimivirus (strain reunion). Cryo-EM was used to determine the structure of the genomic fiber made by the less abundant GMC-oxidoreductase in the wt genomic fiber. Nuclear magnetic resonance (NMR) and Gas Chromatography Mass Spectrometry (GC-MS) were used to complement the study by comparing the glycan compositions and structures of the two single KO viruses and the double KO virus, with the wt. We finally used Mass spectrometry (MS)-based proteomics to analyze the protein composition of the fibril layer of additional members of the family (moumouviruses and megaviruses) which glycosylate their

fibril layer with different glycans using a clade-specific glycosylation machineries (3, 9). While confirming the non-essentiality of the two GMC-oxidoreductases, our results document the unexpected resilience of mimivirus to disruptions of the main components of both genomic fiber and external fibril layer.

Results

None of the two GMC-oxidoreductases is essential

To achieve genetic manipulation of mimivirus, we adapted genetic tools previously described for melbournevirus and pandoravirus (18, 19) by combining homologous recombination with the introduction of selection cassettes. In particular, we built plasmids that introduce a selection cassette (either a nourseothricin N-acetyl transferase (NAT) or a Neomycin resistance gene (NEO)) to disrupt the gene of interest, allowing to select recombinant viruses when the gene is non-essential (Fig 1A). The resistance cassettes were expressed under the control of a megavirus promoter (Fig 1A). Consistently, the addition of nourseothricin or geneticin (the neomycin resistance gene also confers resistance to geneticin) allows to enrich the recombinant viruses, which can be cloned to obtain homogeneous populations (Fig 1B). Using these tools, we generated KO of each of the two genes encoding the GMC oxidoreductases (qu_946 and qu_143). Each mutant was easily produced and genotyped to confirm the mutation (Figure 1-figure supplement 1). Using the second selection cassette we were able to KO both genes (Figure 1-figure supplement 1) and, surprisingly, the resulting mutant was also easy to produce demonstrating that the two GMC-oxidoreductases were not essential. The absence of additional mutations in every mutant was confirmed by genome sequencing.

Mutants' fitness

To assess whether a fitness cost was associated with the mutations, we performed competition assays against wt mimivirus by measuring the abundance of each mutant over several cycles in the presence and absence of selection. Compared to the wild type, each mutant presents a common nourseothricin resistance gene, with, for the double KO (2KO), an additional geneticin resistance gene. As a result, in the presence of nourseothricin, each mutant ratio increased, with the disappearance of the wild-type virus after 5 passages (Fig 1C). These data allowed competition assays to be validated as an effective tool to assess the fitness of recombinant viruses. In the absence of selection, the relative abundance of the

mutants compared to the wild type remained around 0.5 over 5 passages, supporting the absence of a fitness cost, even for the 2KO mutant (Fig 1C).

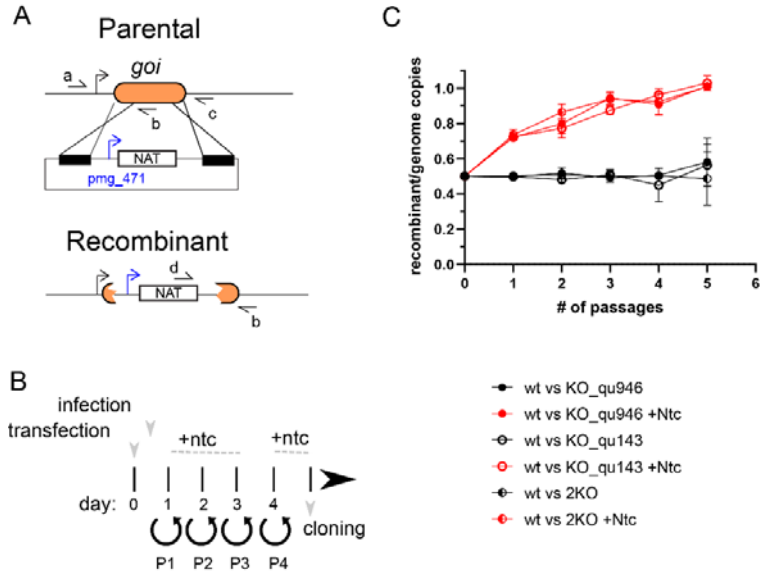

Figure 1: Mimivirus mutants' generation and phenotypic characterization. A] Schematic representation of the vector and strategy utilized for knock out of qu_143 (KO_qu143) and qu_946 (KO_qu946) in mimivirus reunion strain. Selection cassette was introduced by homologous recombination and recombinant viruses were generated as shown in Fig 1B. Primer annealing sites are also shown and the sequence of the primers is included in Figure 1-table supplement 1. B] Cartoon depicting the strategy for the selection of recombinant viruses. Viral infection was performed 1-hour post-transfection. Ntc: Nourseothricin. P= passage. C] Growth competition assays revealed no significant defects in the lytic cycle of KO strains. The competition was also performed in presence of Nourseothricin which allows the out competition of the recombinant strains due to the expression of a Nourseothricin selection cassette. Measurements were performed by qPCR of an endogenous locus (present in wt and recombinant strains) and the Nourseothricin selection cassette (only present in recombinant viruses).

Composition of the genomic fiber of single mutants

We managed to extract and purify the genomic fiber of every single mutant and assessed their protein composition using MS-based proteomics. This confirmed the most abundant protein was the remaining GMC-oxidoreductase and that proteins related to transcription and oxidative stress were still present in the genomic fiber. Negative staining microscopy highlighted surprising differences between the two structures in terms of stability (Fig 2). The two proteins qu_946 and qu_143 in mimivirus reunion strain share 69% sequence identity (81% similarity) and the genomic fiber made of qu_946 (KO_qu143), the most abundant in the wild type genomic fiber, was mostly in the unwound state (Fig 2B), while the one made of qu_143 (KO_qu946), resulted in very long and stable helices that did not

unwound (Fig 2C). The use of the qu_143 protein in the wt genomic fiber could thus contribute to its stabilization as its stability seems to be intermediate beetween each KO (Fig 2A). In the case of the double mutant, the protocol for capsid opening to extract the genomic fiber of wt and single mutants did not work to open properly the capsids. An optimized protocol allowed the extraction of a possible thinner genomic fiber, but in poor yield, preventing its purification and compositional characterization (Fig 2D).

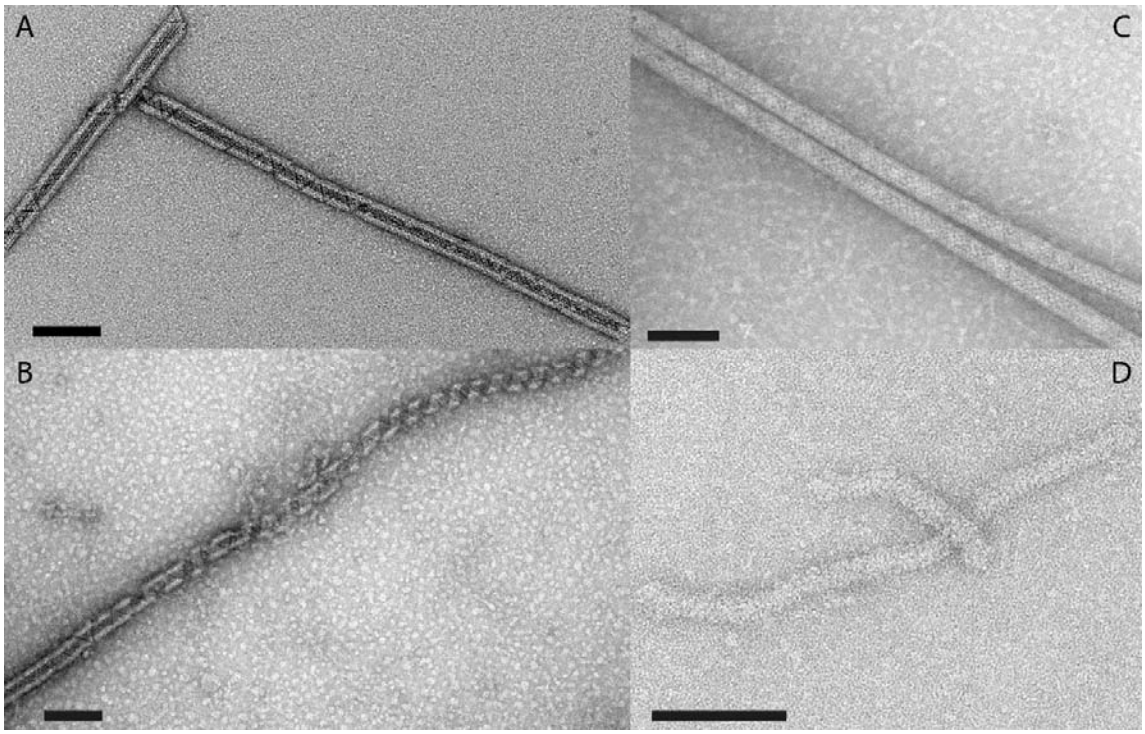

Figure 2: Micrograph of negative stained genomic fiber of A] Wild-type mimivirus, B] KO_qu143 mutant, C] KO_qu946 mutant, and D] 2KO double mutant. Scale bars 100 nm.

Cryo-EM single particle analysis of the qu_143 genomic fiber

We performed cryo-EM single-particle analysis on the KO_qu946 producing the most stable genomic fiber composed by qu_143. The 2D classification (using Relion (20, 21)) confirmed the heterogeneity of the sample, and 2D classes were sorted by applying our already described clustering protocol (1). The two main clusters corresponding in width to the compact (Cl1a) and (Cl2) structures of the wt genomic fiber (Fig 3) were respectively named Cl1 and Cl2.

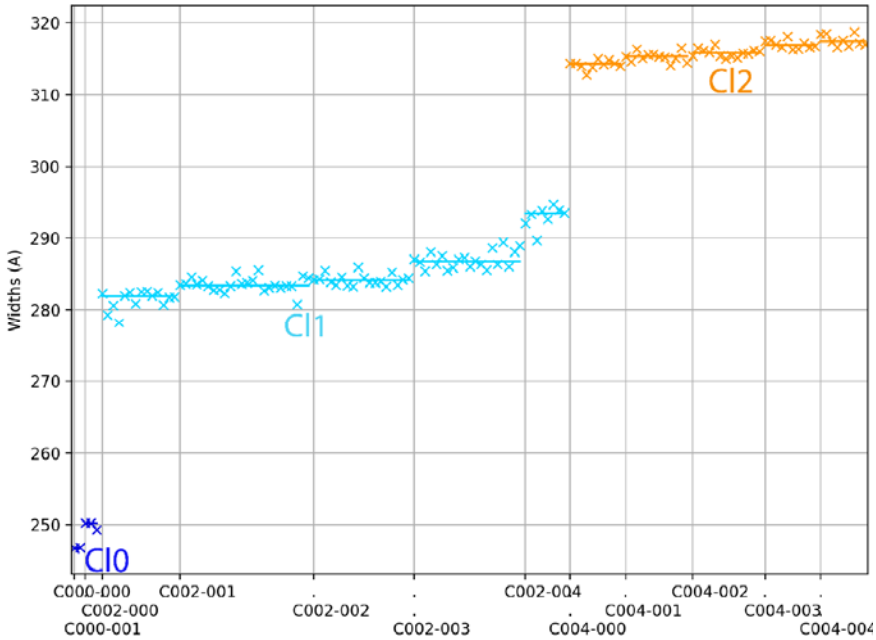

Figure 3: **Clustering analysis of the 2D classes obtained with qu_143 genomic fiber:** Automatic sorting of the 2D classes using the fiber width W1 and pairwise correlations of the 2D classes resulting in 2 main clusters (5-start Cl1 cyan; 6-start Cl2 in orange) and a smaller cluster (Cl0 dark blue).

For each, we confirmed the helical symmetry parameters were the same as the wt genomic fiber by image power spectra analyses and performed structure determination and refinement (Fig 4). The less populated thinner cluster (Cl0, ~25 nm), absent from the wt genomic fiber 2D-classes, was also identified but, due to the lower number of segments and the resulting lower resolution power spectra compared to Cl1 and Cl2, we failed to identify its helical parameters. For Cl1, after 3D-refinement, we obtained a 4.3 Å resolution helical structure (FSC threshold 0.5, masked) corresponding to the same 5-start left-handed helix as the wt, made of a ~8 nm-thick proteinaceous external shell with 5 dsDNA strands lining the interior of the shell, leaving a ~9 nm wide central channel (Fig 4). The 4.2 Å resolution Cl2 map obtained after 3D refinement (Fig 4) corresponds to the same 6-start left-handed helix as the wt, made of the ~8 nm-thick external shell, with 6 dsDNA strands lining it, leaving a ~12 nm wide inner channel (Fig 4). The Fourier shell correlation (FSC) curves for the final 3D reconstructions of the helical assemblies are presented in Fig 4-figure supplement 2.

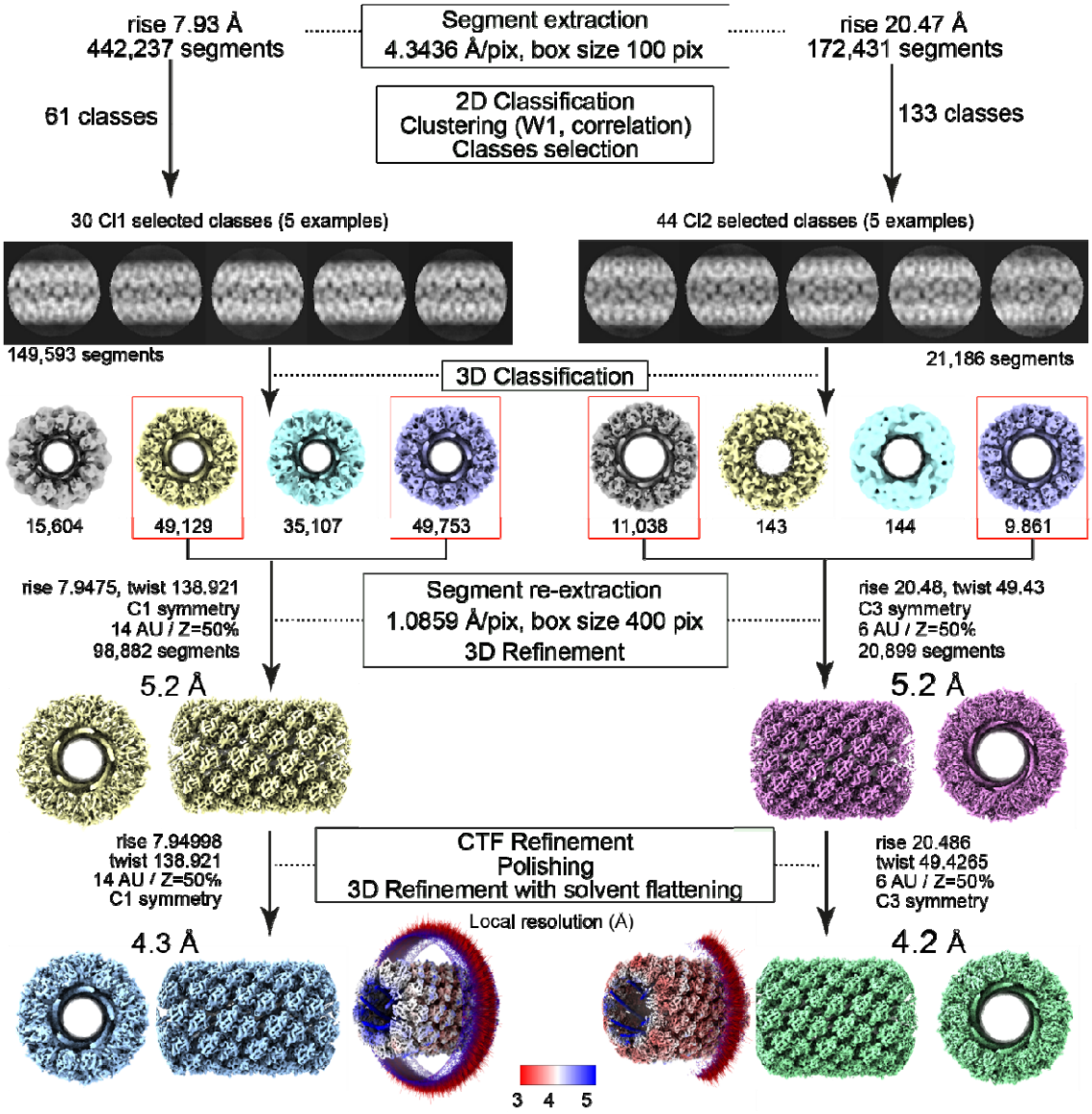

Figure 4: **Workflow of the 5- and 6-start helices reconstruction processes.** Segment extraction was performed with a box size of 400 pixels (pix) binned (box size 100 pix, 4.3436 Å/pix). The distance between consecutive boxes was equal to the axial rise calculated by indexation of the power spectrum. After clustering, 2D classes were selected for Cl1 and Cl2 and 3D classification was carried out using the selected segments, helical symmetry parameters from the power spectrum indexation, and a 300 Å or 330 Å featureless cylinder as 3D reference for Cl1 and Cl2, respectively. 3D-refinement of the 2 boxed 3D classes was achieved using one low pass filtered 3D class as reference on the unbinned segments. A first 3D-Refinement was performed with solvent flattening followed by CTF refinement and polishing of the selected segments. A last 3D refinement was achieved with solvent flattening. The EM maps colored by local resolution from 5 Å (blue) to 3 Å (red) with Euler angle distribution of the particles used for the 3D reconstruction are presented.

We used the qu_143 dimeric structure refined in the Cl1a focus refined map of the wt genomic fiber structure for refinement into the mutant Cl1 and Cl2 maps. Initial models were further refined using the real-space refinement program in PHENIX 1.18.2 (22) (Figure 4-

table supplement 3). The density that can be attributed to the FAD cofactor is present in both maps (Fig4-figure supplement 3) and the models of Cl1 and Cl2 dimers are superimposable with a core RMSD of 0.467 Å based on Cα atoms (Figure 4-table supplement 4). The relaxed Cl3 cluster in the wt was never observed with the mutant. As a result, the qu_143 helices are more stable than the wt, composed of both GMC-oxidoreductases, themselves being more stable than the qu_946-only structure of the KO_qu143, mostly found in an unwound state.

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

In the KO_qu946 genomic fiber, despite the presence of a single GMC-oxidoreductase to build the shell, there is a mixture of 5 and 6 strands of DNA in the genomic fiber as for the wt genomic fiber. We estimated the ratio of 5-start and 6-start from the clustering (Fig 3) and propose a model that reconciles the co-occurrence of the two structures. In this model, the whole genome would be folded into 6 parallel strands, 5 longer than the 6th one. The helix would then be formed initially as a 6-start helix until the sixth strand ends and, from that point becomes, a 5-start helix (Fig 5). According to this model, assuming the length of the genomic fiber is limited by the size of mimivirus nucleoid compartment, we can estimate that the maximum genome length would be ~1.4 Mb for a full 6-start helix with an 8 nm thick protein shell. The last cluster (Cl0, Fig 3) could correspond to a 4-start with an additional shorter DNA strand.

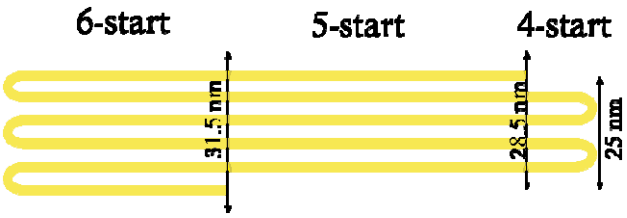

Figure 5: Model explaining the transition from a 6- to a 5-start helix. The transition from a 6- to a 5-starts involves a decrease of the helix diameter by ~3.5 nm. The small cluster around 250 nm could thus correspond to a 4-start helix.

The N-terminal cys-pro-rich domain is an addressing signal to the virion surface

The sequence of the cys-pro-rich N-ter domain of the GMC-oxidoreductases is not covered by proteomic analysis of the purified genomic fiber but is covered by peptides in the purified fibrils after extraction of the fibril layer of the capsids (1). To assess whether this cys-pro-rich N-ter domain could be an addressing signal to the fibrils, we replaced the second GMC-oxidoreductase (qu_143) in the genome of the mutant KO_qu946 by the sequence of the GFP in fusion with the sequence of the qu_143 N-terminal domain (Nqu143-GFP). We then analyzed the resulting virions by MS-based proteomics (Table 1 and Table1-table supplement 2) and fluorescence microscopy. Purified Nqu143-GFP virions showed a strong fluorescence indicating the incorporation of the chimeric protein into the viral particles (Fig 6). Moreover, Nqu143-GFP was identified in purified virions (ranked 53rd in terms of relative abundance, Table1-table supplement 2) and was found enriched in the fraction containing the fibrils (ranked 11th, Table 1 and Table1-table supplement 2), with peptides covering the N-terminal domain (Fig 6-table supplement 5). Taken together, these data indicate the N-terminal cys-pro-rich domain of the GMC oxidoreductase is necessary and sufficient to direct the proteins to the fibrils of mimivirus particles.

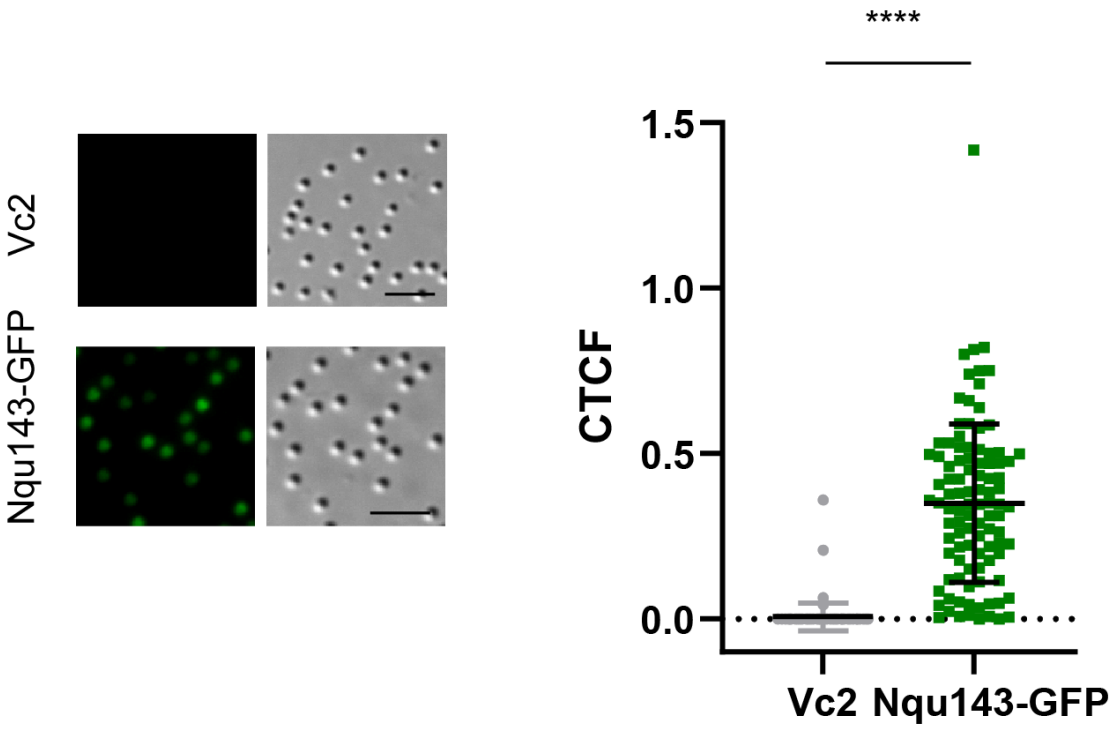

Figure 6: Quantification of fluorescent virion in infected cells overexpressing GFP (Vc2) and in cells infected by the mutant Nqu143-GFP. Scale bar 5 µm. (CTVF: corrected total virion fluorescence).

Protein composition of the fibril layer in additional members of the Mimiviridae family

The *Megavirinae* subfamily (23) of Acanthamoeba infecting *Mimiviridae* currently encompasses five clades, the mimiviruses (A clade), moumouviruses (B clade), megaviruses (C clade), tupanviruses (D clade) (8) and cotonvirus japonicus (E clade) (24). As of today, all members of the *Megavirinae* are characterized by a fibril layer that differs in thickness and length among the clades and their protein composition was unknown. While the cys-pro-rich N-ter domain of the GMC-oxidoreductases is conserved in all members of A clade, it is absent in the orthologous proteins in members of B and C clades, and the two GMC-oxidoreductases are pseudogenized in tupanviruses, suggesting another protein was composing their fibrils (Table 1-figure supplement 6). The proteomic analysis of purified fibrils of the 2KO identified a group of proteins as the most abundant in the fibrils fraction (qu_734, qu_757, qu_384 and qu_482, Table 1 and Table1-table supplement 2). These proteins were also highly ranked in the Nqu143-GFP mutant fibrils and the orthologous of qu_734 is also the most abundant in the M4 natural mutant fibrils (696-L688, Table 1 and Table1-table supplement 2). Since this mutant does not encode a glycosylation machinery, it is not glycosylated and its capsid lacks the large reticulated crown of fibrils decorating mimivirus capsid (4, 17). Interestingly the qu_734 protein also possesses a cys-pro-rich N-ter domain which is conserved in all clades (Table 1-figure supplement 6). It was not possible to determine if one of these proteins was also the building block composing the 2KO genomic fiber, given the difficulty to open its capsids and we concluded that the change in protein composition of the external fibril layer provided different properties to the capsids compared to wt and single mutants and optimization of capsid opening will be necessary to recover enough genomic fiber to be able to purify it and characterize its composition.

The proteomic analysis of fibrils purified from members of B clade, moumouvirus australiensis and maliensis, show similar composition for the most abundant proteins, but with slight differences in their relative abundances (Table 1 and Table1-table supplement 2). The moumouvirus maliensis protein mm_751, ranked 1st in its fibrils, is absent from A clade and in the top 3 in B and C clades. There are cys and pro in the N-ter domain of mm_751, but not as rich as in the GMC-oxidoreductases. For ma_195 (qu_738 in mimivirus reunion, Table 1 and Table1-table supplement 2), ranked 1st in the fibrils of moumouvirus australiensis, the N-ter domain is not cys-pro-rich. The best ranked in the fibrils of members of C clade are the same as in B clade. The first ranked in megavirus chilensis is mg749 (qu_657 in mimivirus reunion) which presents a cys-pro-rich N-ter domain, and the first ranked in megavirus vitis,

mvi_646 (qu_600 in mimivirus reunion), is lacking a cys-pro-rich N-ter domain. Yet, for each virus in each clade, there is at least one protein with a cys-pro-rich N-terminal domain among the most abundant in the fibrils.

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

Rank	Mimvirus reunion V	Mimvirus reunion F	2KO V	2KO F	2KO-GFP V	2KO-GFP F	M4 V	M4 F	Megavirus chilensis V	Megavirus chilensis F	Megavirus Vitis V	Megavirus vitis F	Moumouvirus Australiensis V	Moumouvirus ausraliensis F	Moumouvirus maliensis V	Moumouvirus maliensis fibrils
qu_143	6	1	NA	NA	NA	NA	NA	NA	182	73	177	46	91	45	225	130
qu_757	54	2	59	2	26	9	89	7	134	44	164	24	47	7	115	19
qu_734	14	3	14	1	3	5	20	1	10	7	12	14	10	14	10	16
qu_482	16	4	15	4	194	1	8	54	23	42	32	90	43	9	11	42
qu_880	29	5	33	11	64	18	NA	NA	21	6	26	18	82	18	54	24
qu_384	15	6	13	3	29	2	16	2	27	9	22	4	18	13	23	18
qu_773	7	14	9	10	75	6	8	3	5	8	3	27	3	5	3	8
qu_657	135	15	84	27	12	4	30	27	7	1	8	7	15	4	4	2
qu_600	17	16	18	7	5	9	23	4	35	10	34	1	13	3	15	4
qu_738	22	17	24	12	4	15	17	5	11	3	16	2	7	1	8	3
qu_752	41	35	16	55	86	3	2	21	NA	NA	NA	NA	NA	NA	NA	NA
qu_582	12	39	5	43	1	103	3	28	4	26	6	55	29	33	12	39
mm_751	NA	NA	NA	NA	NA	NA	NA	NA	14	2	25	3	6	2	5	1
Nqu143-GFP	NA	NA	NA	NA	53	11	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

V: Virion; F: Fibrils (grey); NA: Not applicable, the corresponding gene is absent in the corresponding genome.

Sequences of the relevant orthologues of mimivirus reunion in other viruses are provided in Table 1-figure supplement 6.

Glycan composition of the mimivirus mutants' fibrils

We recently established that the fibrils were coated with glycans (3, 9) and demonstrated that for mimivirus, the prototype of the A clade, they were made of two distinct polysaccharides (4). Since the cluster of 12 genes responsible for the biosynthesis of the polysaccharides is preserved in the A clade (3), we suspected that the fibrils of the mimivirus reunion strain had the same composition as mimivirus prototype. To assess whether the knockout of the most abundant protein composing the fibrils could also affect the branching of these polysaccharides and their composition, we analyzed the sugar content of the viral particles of the three mutants (KO_qu946, KO_qu143, and 2KO) and wt mimivirus reunion strain, as reported in (3, 4). The chemical characterization revealed for each of them the presence of sugars (Figure 7A), with rhamnose (Rha), viosamine (Vio), 2-OMe-Vio, glucose (Glc), and glucosamine (GlcN) (Fig 7A), confirming that mimivirus reunion strain and all mutants had the same glycan composition as mimivirus prototype (3). We then analyzed the fibrils of the mutants and wt by ^{1}H NMR spectroscopy (25) and compared them with mimivirus prototype. As a result, in the ^{1}H NMR spectra, the anomeric region (25) (5.5-4.4 ppm) perfectly overlapped with that of the previously studied mimivirus strain (4) (Fig 7B, Fig 7-figure supplement 7), confirming that the different sugars were assembled the same way as in the reference mimivirus to produce the same two polysaccharides (Fig 7C). Indeed, the signal at 4.80 ppm is the result of the overlap of two anomeric protons, the one of a 3-α-Rha (labeled **C**) and of a 3-β-GlcNAc (**D**) modified with a pyruvic acid at the hydroxyl functions 4 and 6. These two residues are the building blocks of the repeating unit of polysaccharide 1 (**poly_1**, **Fig 7C**) (4). The other anomeric signals, labeled with the capital letters **A, A'** (2,3-α-L-Rha), **B** (3-β-D-GlcNAc), **E** (2OMe-β-D-VioNAc) and **E'**(β-D-Vio4NAc), are part of the polysaccharide 2 (**poly_2**) that presents a backbone with a disaccharide repeating unit of 2)-α-L-Rha-(1→3)-β-D-GlcNAc-(1→, with the rhamnose residue further substituted with a viosamine methylated at position 2 (**E**) or not methylated (**E'unit**), thus taking the labels **A** and **A'**, respectively (Fig 7C) (4).

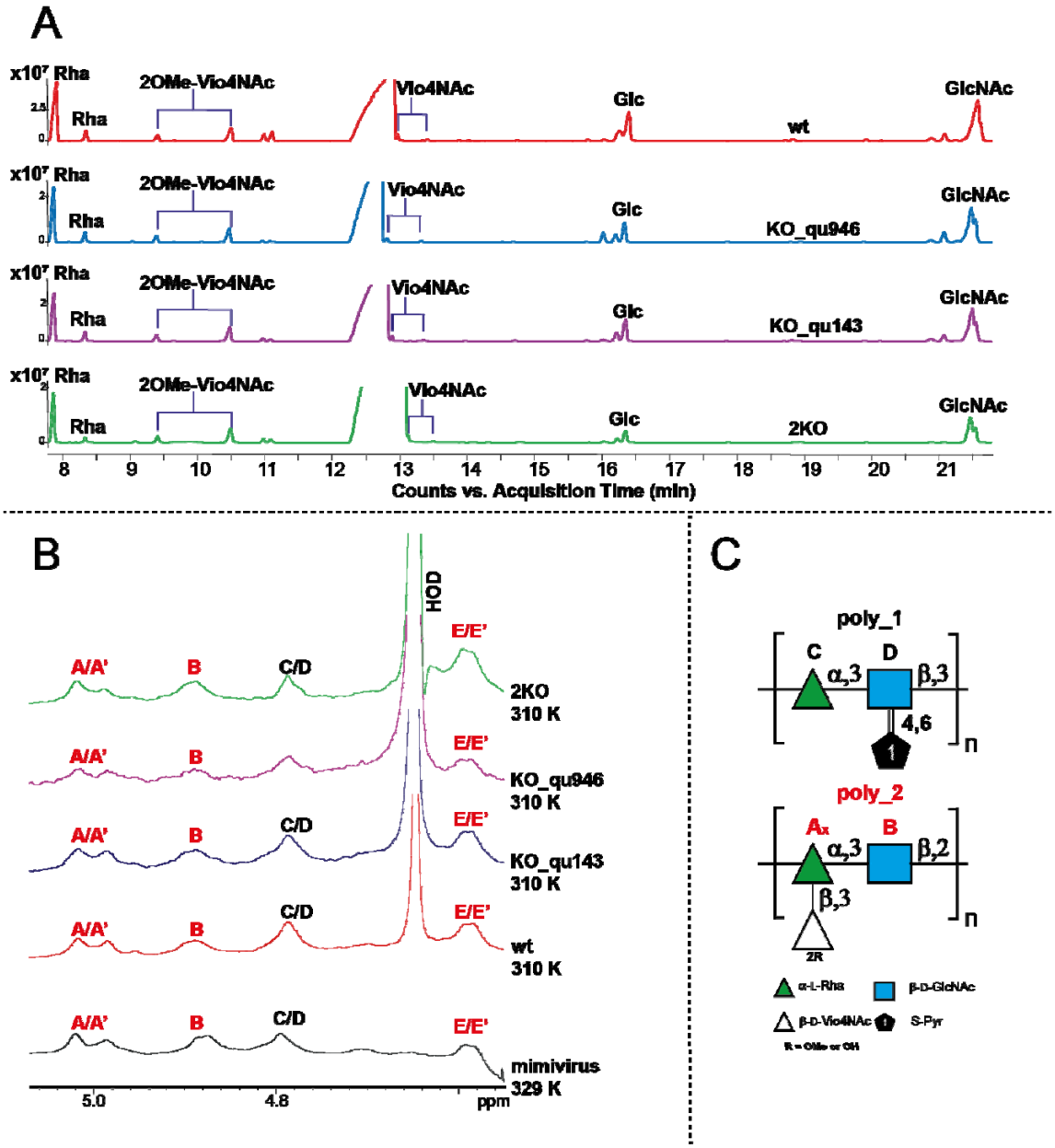

Figure 7: Compositional and NMR analysis of the fibrils of mimivirus reunion strain wt and mutants A] GC-MS chromatogram profiles of the sugars composing the fibrils of wt (a), KO_qu946 (b), KO_qu143 (c), and 2KO (d). B] Comparison of the ^{1}H NMR spectra of mimivirus reunion strain wt and related mutants with that of mimivirus prototype strain. The anomeric signals related to poly_1 (**C** and **D** units, in black), and those of poly_2 are in red. C] Structures of mimivirus polysaccharides as reported in Notaro *et al*, 2021 (4).

It was hypothesized that the mimivirus GMC oxidoreductases were the major target for glycosylation at the fibrils of mimivirus (4, 17). The results presented here indicated that the mutations, including the deletion of the two GMC-oxidoreductase genes, did not affect the glycosylation of the fibrils. These data suggest either the presence of multiple proteins being

targeted for glycosylation or that the glycosylation target is not a GMC oxidoreductase. Alternatively, it is also possible that the glycosylation machinery was able to use the qu_734 protein as support for the two polysaccharides. No glycans were identified in the M4 natural mutant fibrils.

Discussion

Genome packaging is essential for the propagation of viruses. For instance, the packaging ATPase of poxviruses (26) or the histones of marseilleviruses (19, 27) are essential for the productive infection by their virions. We have previously demonstrated that the genomic fiber of mimivirus and the external layer of glycosylated fibrils are composed of two GMC oxidoreductases (1), but in contrast to previous examples, the absence of both GMC oxidoreductases in the M4 laboratory-generated strain suggested they were not essential (17) . Implementation of the genetics in cytoplasmic giant viruses allowed us to assess the essentiality of GMC-oxidoreductases in mimivirus reunion strain. We have previously proposed an increased genome redundancy as a contributing factor to the appearance of viral giantism (18). The data presented here validate such premise and extend these predictions outside of the *Pandoraviridae*. Indeed, unsurprisingly, given the homology between the two enzymes, the KO of the most abundant one in the wt genomic fiber induced its replacement by the second one, with no apparent cost for the virus, at least in laboratory conditions. Regardless, despite the exchangeable nature of these proteins, the appearance and physical features of the genomic fiber were significantly different between recombinant viruses and the wt, indicating a certain degree of functional specialization. The KO_qu946 genomic fiber, as the wt, presents two helices, a 5- and a 6-start helixbut built with a single GMC-oxidoreductase. This led us to propose a model to explain the co-existence of both structures in which proteins, such as the chromosome condensation factor identified in the proteome of the genomic fiber (1), could fold the DNA into 6 strands before assembly in the helical structure. But, while the wt genomic fiber was highly heterogeneous, with fully relaxed helices having lost DNA and even helices that lost a strand, the KO_946 genomic fiber made of qu_143 appeared more stable than the wt with no occurrence of fully relaxed structures. Yet, despite our efforts to improve the resolution, the helices appeared less defined compared to the wild type, which may be due to internal disorders or to the cryo-sample preparation. Moreover, the deletion of qu_143, the less abundant GMC-oxidoreductase in the wt genomic fiber, seemed to increase the genomic fiber instability, suggesting a stabilizing role for

qu_143 in the wt virus. Finally, the deletion of both GMC-oxidoreductases demonstrated they were not essential, with no fitness cost associated with their loss.

Importantly, GMC oxidoreductases are not only present in the genomic fiber but also in the virion external fibril layer and have been previously predicted as targets for glycosylation (4, 17). The only difference between the genomic fiber and the external fibril layer being that the cys-pro-rich N-ter domain of the GMC-oxidoreductases was identified in the fibrils and was absent from the GMC oxidoreductase sequence from the genomic fiber. Interestingly, when fused to GFP, this cys-pro-rich N-ter domain of qu_143 is sufficient to produce fluorescent virions, suggesting it is an addressing signal to the external fibril layer. Unexpectedly, the virions of all three mutants were still decorated by long, reticulated fibrils. The MS-based proteomic analysis of the 2KO fibrils revealed they were composed by other proteins which were also enriched in the M4 natural mutant fibrils, some, including qu_734, also presenting a cys-pro-rich N-ter domain (Table 1-figure supplement 6). The glycan analysis of the GMC-oxidoreductases single and double mutants confirmed that their fibrils were still glycosylated by the same two polysaccharides as the wt virions. As previously reported, M4 was not glycosylated as it lacks a virally encoded glycosylation machinery (3, 4, 17). Moreover, the resulting change in the composition of the surface fibril layer in the 2KO virions seemed to impact the capsid sensitivity to the proteolytic treatment we originally used to recover wt and single mutants genomic fibers and we failed to recover the 2KO genomic fiber in quantities compatible with structural and compositional analysis. Yet, a thinner genomic fiber was observed (~20-25 nm, Fig 2D).

The additional analysis of the purified fibrils of members of B and C clades confirmed this versatility, with fibrils composed of different proteins inter- and even intra-clade. B and C clades virions present fibrils with a protein composition closer to each other than that of clade A. In a previous study, we reported that the glycosylation machinery was clade-specific and this is now the overall composition of the fibrils that appears to vary. The glycosylated fibril layer bear glycans echoing bacterial ones, recognized by the amoeba which feeds on bacteria (2, 3). They thus appear as key for infection, by fooling the amoeba which mistakes the virions for bacteria due to their size and “taste”. In a given environment, the composition of the fibrils could correspond to a variable, favoring the engulfment of a given virion by a given amoeba compared to others virions and even bacteria. The protein content of the fibrils also appears to have been optimized in a given clade but their complex composition suggests there

is a large set of proteins that can make the fibrils. Having a flexible toolbox for glycan synthesis and fibril protein composition would reflect the need to continuously adjust the capsids composition to outcompete other parasites and secure infection. We can thus hypothesize that in the population of virions resulting from a single infection, this composition could also be variable, providing to the virus with a toolbox to ensure the productive infection of at least one of the potential hosts present in that environment. Is it always the same protein that makes both the fibril layer and the genomic fiber when other proteins are used to make the fibril layer ? For the 2KO a thinner structure was observed (~20-25 nm). The most abundant protein composing the external fibril layer of the 2KO and M4 corresponds to the 222 amino-acids qu_734 protein, which according to the high confidence alphaFold (28) prediction could be almost twice smaller (~4.5 nm) than the GMC-oxidoreductase. Given that the central part of the genomic fiber has to correspond to a central channel of at least 9 nm to accommodate central proteins such as the RNA polymerase. There is 4 additional nm for the DNA lining the protein shell. The GMC-oxidoreductase structure is 8 nm height and makes the 16 nm helical shell, leading to a ~29 nm helix and the size of the protein making the shell is defining the final dimension of the genomic fiber. For the 2KO if this is the same protein that makes the external fibril layer and the shell of the genomic fiber, we expect a genomic fiber of 22 nm, compatible with our observations. Overall, our results reveal the resilience of mimivirus, with redundant solutions securing essential functions, such as infectiousness and genome packaging. Functional redundancy, well documented in the cellular world as a way to preserve essential function such as cell division (29) can now be expanded to the viral world.

Material and methods

Cloning of DNA constructs

Gene knock-out vectors. The plasmid for gene knock-out was generated by sequential cloning of the 3’ UTR of mg_18 (megavirus chiliensis), the promoter of mg_741 (megavirus chiliensis), and a Nourseothricin N-acetyl transferase (NAT) or a neomycin selection cassette (NEO). Each cloning step was performed using the Phusion Taq polymerase (ThermoFisher) and InFusion (Takara). Finally, 500bp homology arms were introduced at the 5’ and 3’ end of the cassette to induce homologous recombination with the viral DNA. Before transfection, plasmids were digested with EcoRI and NotI. All primers are shown in Fig1-table supplement 1.

Establishment of viral lines.

Gene knock-out. A detailed method for gene manipulation is provided in (Manuscript accepted, Nature Protocols). Gene knockout strategy was performed as previously described for pandoravirus (18). Briefly, 1.5x105 *Acanthamoeba castellanii cells* were transfected with 6 µg of linearized plasmid using Polyfect (QIAGEN) in phosphate saline buffer (PBS). One hour after transfection, PBS was replaced with PPYG, and cells were infected with 1.5x107 mimivirus reunion particles for 1 hour with sequential washes to remove extracellular virions. 24h after infection the new generation of viruses (P0) was collected and used to infect new cells. An aliquot of P0 viruses was utilized for genotyping to confirm the integration of the selection cassette. Primers used for genotyping are shown in Table S1. A new infection was allowed to proceed for 1 hour, then washed to remove extracellular virions and nourseothricin was added to the media. Viral growth was allowed to proceed for 24 hours. This procedure was repeated one more time before removing the nourseothricin selection to allow viruses to expand more rapidly. Once, the viral infection was visible, the selection procedure was repeated one more time. Viruses produced after this new round of selection were used for genotyping and cloning. Double knockout of the GMC oxidoreductases was obtaining by using a clonal population of qu_143 knockout viruses as parental strain. The locus of qu_946 was replaced by a neomycin resistance cassette. The transfection and selection of recombinant viruses' procedure was performed identical to the process to generate single knockout but replacing nourseothricin by geneticin.

Cloning and genotyping. 150,000 *A. castellanii* cells were seeded on 6 well plates with 2mL of PPYG. After adhesion, viruses were added to the well at a multiplicity of infection (MOI) of 1. One-hour post-infection, the well was washed 5 times with 1mL of PPYG, and cells were recovered by well scraping. Amoebas were then diluted until obtaining a suspension of 1 amoeba/µL. 1µL of such suspension was added in each well of a 96-well plate containing 1000 uninfected *A. castellanii* cells and 200µL of PPYG. Wells were later monitored for cell death and 100µL collected for genotyping. Genotyping was performed using Terra PCR Direct Polymerase Mix (Takara) following the manufacturer's specifications. Primers used for genotyping are shown in Fig1-table supplement 1.

Competition assay and quantitative PCR analysis.

Equal infectious particles of wild-type and recombinant mimivirus reunion were mixed and used to infect *A. castellanii* at an approximate multiplicity of infection of 0.8. Viruses were

allowed to grow overnight in the presence or absence of nourseothricin. Subsequent viral progenies were used to infect new *A. castellanii* cells in reiterative passages. A fraction of each passage was collected for genomic DNA extraction.

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

Genome sequencing and assembly of mutants' genomes

Genomic DNA was extracted from 10^{10} virus particles using the PureLink TM Genomic DNA mini kit (Invitrogen) according to the manufacturer's protocol. Clones of individual mutants and wt were sequences on an illumine platform (Novogen). For mimivirus wt, we obtained 4,819,885 150nt paired-ended reads. For KO_qu946, KO_qu946 and 2KO, 4,662,744, 4,572,873 and 5,068,030 150nt paired-ended reads, respectively. Genomes were assembled using spades v3.13, with the option "careful" and we obtained 5 contigs for the wt, 4 for KO_qu946, and 7 for KO_qu143 and 2KO. The wt sequence was consistent with the official version of mimivirus reunion (GI MW004169) and the mutants only differed in the mutated regions (in qu_946 and/or qu_143 genes). For the clone KO_qu946 the gene is interrupted from position 1151358 to 1152985 relative to wt genome (position 302 to 1741 in the corresponding qu_946 gene) and for the clone KO_qu143 the gene is interrupted from position 165934 to 167572 relative to wt genome (position 306 to 1741 in the corresponding qu_143 gene). Th region is replaced in both cases by the nourseothricin cassette construction (1637 nt). For the 2KO, the gene qu_143 is interrupted from the positions 306 to 1741 (165934 to 167572 on the wt genome) and replaced by the nourseothricin cassette construction (1637 nt), and gene qu_946 is interrupted from the positions 302 to 1206 (1151551 to 1153410 on the genome) and replaced by the geneticin cassette construction (1859 nt). To confirm the KO of each mutant, the reads were mapped on the wt genome resulting in a regular coverage along the genome, excepted for the qu_143 and qu_946 central positions thus confirming the mutations. In addition, we used the central part of the KO genes

(deleted in mutants) as blast queries against the reads of each mutant genome, which also confirmed the absent of the central portion.

Extraction and purification of the qu_946 and qu_143 mutants' genomic fiber

The genomic fiber of the mimivirus reunion single mutants of qu_946 (KO_qu946) and qu_143 (KO_qu143), were extracted as described in Villalta et al, 2022 for the wild-type virus (1). The genomic fiber was extracted from 12 mL of purified single KO mutant virions at 2 x 10^{10} particles/mL, split into 12x1 mL samples processed in parallel. Trypsin (Sigma T8003) in 40 mM Tris-HCl pH 7.5 buffer was added at a final concentration of 50 µg/mL and the virus-enzyme mix was incubated for 2h at 30°C in a heating dry block (Grant Bio PCH-1). DTT was then added at a final concentration of 10 mM and incubated at 30°C for 16h. Finally, 0.001% Triton X-100 was added to the mix and incubated for 4h at 30°C. Each tube was centrifuged at 4,000 x g for 5 min to pellet the opened capsids. The supernatant was recovered, and the fibers were concentrated by centrifugation at 15,000 x g for 4h at 4°C. Most of the supernatant was discarded leaving 12x~200 µL of concentrated fibers that were pooled and layered on top of ultracentrifuge tubes of 4 mL (polypropylene centrifuge tubes, Beckman Coulter) containing a discontinuous cesium chloride gradient (1.4, 1.3, 1.2 g/cm^3 in 40 mM Tris-HCl pH 7.5 buffer). The gradients were centrifuged at 200,000 x g for 16h at 4 °C. Since no visible band was observed, successive 0.5 mL fractions were recovered from the bottom of the tube. Each fraction was dialyzed using 20 kDa Slide-A-Lyzers (ThermoFisher) against 40 mM Tris-HCl pH 7.5 to remove the CsCl. These fractions were further concentrated by centrifugation at 15,000 x g, at 4°C for 4h, and most of the supernatant was removed, leaving ~100 µL of sample at the bottom of each tube. At each step of the extraction procedure, the sample was imaged by negative staining transmission electron microscopy (TEM) to assess the integrity of the genomic fiber. Each fraction of the gradient was finally controlled by negative staining TEM.

Negative stain TEM

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

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

Single-particle analysis by cryo-EM

Sample preparation

For the KO_qu946 mutant, 3 µL of the purified sample were applied to glow-discharged Quantifoil R 2/1 Cu grids, blotted for 2 s using a Vitrobot Mk IV (Thermo Scientific) and applying the following parameters: 4°C, 100% humidity, blotting force 0, and plunge frozen in liquid ethane/propane cooled to liquid nitrogen temperature.

Data acquisition

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

2D classification and clustering of 2D classes

All movie frames were aligned using MotionCor2 (30) and used for contrast transfer function (CTF) estimation with CTFFIND-4.1 (31). Helical segments of the purified genomic fiber manually picked with Relion 3.1.0 , were extracted with 400 pixels box sizes (decimated to 100 pixels) using a rise of 7.93 Å and a tube diameter of 300 Å. Particles were subjected to reference-free 2D classification in Relion 3.1.0 (20, 21). We then performed additional cluster analysis of the initial 2D classes provided by Relion to aim for more homogeneous clusters (1), eventually corresponding to different states (Fig 4).

Identification of candidate helical parameters

Fourier transform analysis methods have been used to confirm the helical parameters were the same as in wt (1, 32–34) for the Cl1 and Cl2.

Cryo-EM data processing and 3D reconstruction

After helical parameters determination, segments were extracted with a box size of 400 pixels (decimated to 100 pixels) using the proper rises for the Cl1 (7.93 Å, cylinder 300 Å, 442,237 segments) and the Cl2 (20.47 Å, cylinder 330 Å, 172,431 segments). A dedicated 2D classification protocol was performed independently on each extraction. For the Cl2, one round of 50 expectation-maximization (E-M) iterations was sufficient to produce 133 homogeneous 2D classes submitted to cluster analysis of which 44 2D classes were selected (21,186 segments). For the Cl1 extraction, three iterative 2D classification/selection rounds were performed (25, 50 and 100 E-M iterations) producing 61 classes from which 30 (149,593 segments) were finally selected on the basis of the cluster analysis.

Values of the helical parameters, rise and twist (Cl1: 7.9475 Å, 138.921°; Cl2: 20.48 Å, 49.43°), were then used for Relion 3D classification (20, 21), with a +/-0.5 units freedom search range, using a featureless cylinder as initial reference (diameter of 300 Å and C1 symmetry for Cl1 and 330 Å and C3 symmetry for Cl2). In real space, the helical symmetry was searched using 50% of the central part of the box for both structures. The number of asymmetrical units in each segment box was set to 1 for Cl1 and 6 for Cl2. The entire helical reconstruction and protein shell dimensions were obtained using in house developed program.

The superimposable 3D classes (same helical parameters, same helix orientation) were then selected, reducing the data set to 98,882 segments for the 5-start fiber (Cl1) and 20,899 segments for the 6-start fiber (Cl2). After re-extraction of the selected segments without scaling, a first unmasked 3D refinement was performed with a rescaled 3D classification output low pass filtered to 15 Å as reference, followed by a 3D refinement step using solvent flattened, FSCs and CTF refinement using the standard procedure described in Relion. To further improve resolution of the maps, a Bayesian polishing was applied using 10,000 segments for training and default values for polishing. A last round of 3D refinement with solvent flattening was applied to the previous refined map using the polished particles. At that point, the maps were resolved enough (Cl1: 4.3 Å, Cl2: 4.2; FSC threshold 0.5) to identify secondary structure elements (with visible cylinders corresponding to the helices) and were used to compute local resolution.

Structures refinement

The best resolution Cl2 map was used to fit the qu_143 dimeric structure (PDB 7YX3) using UCSF ChimeraX 1.5 (35). Each monomer was rigid-body fitted independently into the map for a best result. The entire protein was then inspected within coot 0.9.7 (36) to fix local

inconsistencies and was further refined against the map using the real-space refinement program in PHENIX 1.20.1 (22). The protein was submitted to 5 cycles of rigid body refinement (with each chain defined) followed by twice 10 cycles of refinement with the default PHENIX options. The resulting structure was manually corrected within coot. The resulting protein model was submitted to the same steps refinement in PHENIX. This final model was then fitted into the Cl1 map, inspected with coot and refined using 5 cycles of rigid body refinement and simulated annealing followed by twice 10 cycles of refinement with the default PHENIX options. Validations were performed into PHENIX using the comprehensive validation program (Fig4-table supplement 3). RMSD between different structures (monomers and dimers, Figure 4-table supplement 4) were computed using the align procedure in Pymol suite (Schrödinger, L., & DeLano, W. (2020). *PyMOL*).

Extraction and purification of the mutants' fibrils

Defibrillation was performed by incubating purified virions (1.5 x 10^{10}) in Eppendorf tubes in 500 µL 40 mM Tris-HCl pH 7.5 buffer, 500 mM DTT for 2h at 30°C. Tubes are then centrifuged at 14,000 g for 10 min. The supernatants containing the fibrils were recovered and concentrated on vivaspin® 3 KDa (Sartorius, VS04T91) at 3,000 g. The pellet is washed twice with 40 mM Tris-HCl pH 7.5 buffer and centrifuged at 14,000 g for 10 min and finally resuspended in the same buffer. Both pellet and fibrils were imaged by negative staining TEM.

For the 3 mutants and wt mimivirus reunion strain, 4 x 10^{11} viral particles were centrifuged at 14,000g for 10 min, the supernatant was discarded and the pellet was re-suspended in 10 ml of 50 mM DTT. Fibril extraction was performed at 100°C under stirring. After 2 h, the tube was centrifuged at 14,000 g for 15 min, at 4°C, and the fibrils recovered with the supernatant. The fibrils were then dried and purified on Biogel P10, followed by subsequent NMR analysis of each mutant and wt.

Sugar composition of viral particles of Mimivirus reunion wt and mutants

Monosaccharide composition analysis as acetylated methyl glycoside (AMG) was performed on the intact viral particles (1,25 x 10^{10}, ~250 µl) of mimivirus reunion wt and mutants, following the procedure reported by De Castro *et al* 2010 (37). The obtained AMG were analyzed via gas chromatography-mass spectrometry (GC-MS) on an Agilent instrument (GC instrument Agilent 6850 coupled to MS Agilent 5973) equipped with a SPB-5 capillary

column (Supelco, 30 m × 0.25 i.d., flow rate, 0.8 mL min–1) and He as the carrier gas. The identification of the monosaccharides derivatized as AMG, was obtained by studying the fragmentation pattern corresponding to each peak of the chromatogram and by comparison with suitable standards.

Purification and ^{1}H NMR analysis of the fibrils

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

Briefly, the glycoproteins (protein/s carrying the polysaccharides) were precipitated with cold acetone at 80%, at 4 °C, for 16 hours, for 2 times. The supernatant containing DTT and salts was discarded, while the precipitate was dissolved in water and freeze-dried. Then, the precipitate was purified by size exclusion chromatography (Biogel P10, flow: 12 ml / h) to completely remove the DTT. The eluted fractions were checked by ^{1}H NMR, revealing that the glycan-containing material was eluted at one-third of the column volume (full spectra are shown in Fig 7B).

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

genomic fiber. Each fraction of the gradient was finally controlled by negative staining TEM.

Mass spectrometry-based proteomic analyses

Proteins were solubilized with Laemmli buffer (4 volumes of sample with 1 volume of Laemmli 5X - 125 mM Tris-HCl pH 6.8, 10% SDS, 20% glycerol, 25% β-mercaptoethanol and traces of bromophenol blue) and heated for 10 min at 95 °C. The extracted proteins were stacked in the top of an SDS-PAGE gel (4-12% NuPAGE, Life Technologies), stained with Coomassie blue R-250 (Bio-Rad) before in-gel digestion using modified trypsin (Promega, sequencing grade) as previously described (38). The resulting peptides were analyzed by online nanoliquid chromatography coupled to tandem MS (UltiMate 3000 RSLCnano and Q-Exactive Plus or Q-Exactive HF, Thermo Scientific). Peptides were sampled on a 300 µm x 5 mm PepMap C18 precolumn and separated on a 75 µm x 250 mm C18 column (Reprosil-Pur 120 C18-AQ, 1.9 µm, Dr. Maisch, except for Nqu143-GFP mutant samples separated on

Aurora, 1.7 µm, IonOpticks) using a 140-min gradient (except for fibrils from Nqu143-GFP mutant for which a 60-min gradient was used). MS and MS/MS data were acquired using Xcalibur 4.0 (Thermo Scientific). Peptides and proteins were identified using Mascot (version 2.8.0, Matrix Science) through concomitant searches against homemade *A. castellanii* protein sequence database, homemade virus-specific protein sequence databases, and a homemade database containing the sequences of classical contaminant proteins found in proteomic analyses (human keratins, trypsin...). Trypsin/P was chosen as the enzyme and two missed cleavages were allowed. Precursor and fragment mass error tolerances were set at respectively at 10 and 20 ppm. Peptide modifications allowed during the search were: Carbamidomethyl (C, fixed), Acetyl (Protein N-term, variable) and Oxidation (M, variable). The Proline software version 2.2.0 (39) was used for the compilation, grouping, and filtering of the results: conservation of rank 1 peptides, peptide score ≥ 25, peptide length ≥ 6 amino acids, peptide-spectrum-match identification false discovery rate < 1% (40), and minimum of 1 specific peptide per identified protein group. Proline was then used to perform a MS1-based quantification of the identified protein groups. Intensity-based absolute quantification (iBAQ) (41) values were calculated for each protein group from the MS intensities of razor and specific peptides.

Acknowledgments

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

Data and materials availability: Genome sequences of the mutants of mimivirus reunion have been deposited to NCBI (accession numbers: KO_qu_946: OQ700912; KO_qu143: OQ700913, 2KO: OQ700914). 3D reconstruction maps and the corresponding PDB have been deposited to EMDB (Deposition number Cl1: 8ORS, EMD-17131; Cl2: 8ORH, EMD-

17125). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (42) partner repository with the dataset identifier PXD041298. AlphaFold predictions were performed using HPC/AI resources from GENCI-IDRIS (Grant 2022-AD011013526).

Competing interests: Authors declare that they have no competing interests

References

1. A. Villalta, *et al.*, The giant mimivirus 1.2 Mb genome is elegantly organized into a 30-nm diameter helical protein shield. *eLife* **11**, e77607 (2022).

2. R. A. L. Rodrigues, *et al.*, Mimivirus Fibrils Are Important for Viral Attachment to the Microbial World by a Diverse Glycoside Interaction Repertoire. *J. Virol.* **89**, 11812–11819 (2015).

3. A. Notaro, *et al.*, Giant viruses of the *Megavirinae* subfamily possess biosynthetic pathways to produce rare bacterial-like sugars in a clade-specific manner. *microLife* **3**, uqac002 (2022).

4. A. Notaro, *et al.*, Expanding the Occurrence of Polysaccharides to the Viral World: The Case of Mimivirus. *Angew. Chem. Int. Ed.* **60**, 19897–19904 (2021).

5. E. V. Koonin, *et al.*, Global Organization and Proposed Megataxonomy of the Virus World. *Microbiol. Mol. Biol. Rev.* **84**, e00061-19 (2020).

6. D. Raoult, *et al.*, The 1.2-megabase genome sequence of Mimivirus. *Science* **306**, 1344–1350 (2004).

7. C. Abergel, J. Rudinger-Thirion, R. Giegé, J.-M. Claverie, Virus-encoded aminoacyl-tRNA synthetases: structural and functional characterization of mimivirus TyrRS and MetRS. *J. Virol.* **81**, 12406–12417 (2007).

8. J. Abrahão, *et al.*, Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nat. Commun.* **9**, 749 (2018).

9. I. Speciale, *et al.*, The Astounding World of Glycans from Giant Viruses. *Chem. Rev.* **122**, 15717–15766 (2022).

10. N. Zauberman, *et al.*, Distinct DNA exit and packaging portals in the virus Acanthamoeba polyphaga mimivirus. *PLoS Biol.* **6**, e114 (2008).

11. J. R. Schrad, J. S. Abrahão, J. R. Cortines, K. N. Parent, Structural and Proteomic Characterization of the Initiation of Giant Virus Infection. *Cell* **181**, 1046-1061.e6 (2020).

12. J.-M. Claverie, C. Abergel, Mimivirus: the emerging paradox of quasi-autonomous viruses. *Trends Genet.* **26**, 431–437 (2010).

13. M. Suzan-Monti, B. L. Scola, L. Barrassi, L. Espinosa, D. Raoult, Ultrastructural Characterization of the Giant Volcano-like Virus Factory of Acanthamoeba polyphaga Mimivirus. *PLoS ONE* **2**, e328 (2007).

703 14. Y. Mutsafi, N. Zauberman, I. Sabanay, A. Minsky, Vaccinia-like cytoplasmic replication of the
704 giant Mimivirus. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 5978–5982 (2010).

705 15. Y. G. Kuznetsov, T. Klose, M. Rossmann, A. McPherson, Morphogenesis of mimivirus and its
706 viral factories: an atomic force microscopy study of infected cells. *J. Virol.* **87**, 11200–11213
707 (2013).

708 16. I. L. M. de Aquino, *et al.*, Diversity of Surface Fibril Patterns in Mimivirus Isolates. *J. Virol.*,
709 e01824-22 (2023).

710 17. M. Boyer, *et al.*, Mimivirus shows dramatic genome reduction after intraamoebal culture. *Proc.*
711 *Natl. Acad. Sci. U. S. A.* **108**, 10296–10301 (2011).

712 18. H. Bisio, *et al.*, Evolution of giant pandoravirus revealed by CRISPR/Cas9. *Nat. Commun.* **14**, 428
713 (2023).

714 19. Y. Liu, *et al.*, Virus-encoded histone doublets are essential and form nucleosome-like structures.
715 *Cell* **184**, 4237-4250.e19 (2021).

716 20. S. He, S. H. W. Scheres, Helical reconstruction in RELION. *J. Struct. Biol.* **198**, 163–176 (2017).

717 21. S. H. W. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure
718 determination. *J. Struct. Biol.* **180**, 519–530 (2012).

719 22. D. Liebschner, *et al.*, Macromolecular structure determination using X-rays, neutrons and
720 electrons: recent developments in *Phenix. Acta Crystallogr. Sect. Struct. Biol.* **75**, 861–877
721 (2019).

722 23. L. Gallot-Lavallée, G. Blanc, J.-M. Claverie, Comparative Genomics of Chrysochromulina Ericina
723 Virus and Other Microalga-Infecting Large DNA Viruses Highlights Their Intricate Evolutionary
724 Relationship with the Established Mimiviridae Family. *J. Virol.* **91**, pii: e00230-17 (2017).

725 24. H. Takahashi, S. Fukaya, C. Song, K. Murata, M. Takemura, Morphological and taxonomic
726 properties of the newly isolated *Cotonvirus japonicus* , a new lineage of the subfamily
727 *Megavirinae*. *J. Virol.* (2021) https:/doi.org/10.1128/JVI.00919-21 (July 26, 2021).

728 25. I. Speciale, *et al.*, Liquid-state NMR spectroscopy for complex carbohydrate structural analysis:
729 A hitchhiker's guide. *Carbohydr. Polym.* **277**, 118885 (2022).

730 26. M. C. Cassetti, M. Merchlinsky, E. J. Wolffe, A. S. Weisberg, B. Moss, DNA Packaging Mutant:
731 Repression of the Vaccinia Virus A32 Gene Results in Noninfectious, DNA-Deficient, Spherical,
732 Enveloped Particles. *J. Virol.* **72**, 5769–5780 (1998).

733 27. M. I. Valencia-Sánchez, *et al.*, The structure of a virus-encoded nucleosome. *Nat. Struct. Mol.*
734 *Biol.* **28**, 413–417 (2021).

735 28. J. Jumper, *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–
736 589 (2021).

737 29. H. P. Erickson, M. Osawa, Cell division without FtsZ--a variety of redundant mechanisms. *Mol.*
738 *Microbiol.* **78**, 267–270 (2010).

30. S. Q. Zheng, *et al.*, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

31. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

32. N. Coudray, *et al.*, Deducing the symmetry of helical assemblies: Applications to membrane proteins. *J. Struct. Biol.* **195**, 167–178 (2016).

33. R. Diaz, W. J. Rice, D. L. Stokes, Fourier-Bessel reconstruction of helical assemblies. *Methods Enzymol.* **482**, 131–165 (2010).

34. C. Sachse, Single-particle based helical reconstruction—how to make the most of real and Fourier space. *AIMS Biophys.* **2**, 219–244 (2015).

35. E. F. Pettersen, *et al.*, UCSF Chimera-A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

36. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of *Coot*. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

37. C. De Castro, M. Parrilli, O. Holst, A. Molinaro, "Microbe-Associated Molecular Patterns in Innate Immunity" in *Methods in Enzymology*, (Elsevier, 2010), pp. 89–115.

38. M. G. Casabona, Y. Vandenbrouck, I. Attree, Y. Couté, Proteomic characterization of Pseudomonas aeruginosa PAO1 inner membrane. *Proteomics* **13**, 2419–2423 (2013).

39. D. Bouyssié, *et al.*, Proline: an efficient and user-friendly software suite for large-scale proteomics. *Bioinforma. Oxf. Engl.* **36**, 3148–3155 (2020).

40. Y. Couté, C. Bruley, T. Burger, Beyond Target-Decoy Competition: Stable Validation of Peptide and Protein Identifications in Mass Spectrometry-Based Discovery Proteomics. *Anal. Chem.* **92**, 14898–14906 (2020).

41. B. Schwanhäusser, *et al.*, Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).

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

Funding:

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

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

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

CDC gratefully acknowledges STARPLUS 2020 (project no. 21□UNINA□EPIG□042) from the University of Napoli for financial support.