Running Title: 4.4 Å structure of *Marseilleviridae*

The 4.4 Å structure of the giant Melbournevirus virion belonging to the *Marseilleviridae* family

Raymond N. Burton-Smith[1], Hemanth Kumar Narayana Reddy[2], Martin Svenda[3], Chantal Abergel[4], Kenta Okamoto[2], Kazuyoshi Murata[1,5,6*]

[1] Exploratory Research Center on Life and Living Systems (ExCELLS), National Institutes of Natural Sciences, Okazaki, Aichi, Japan,

[2] Program in Molecular Biophysics, Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden,

[3] Biomedical and X-Ray Physics, Department of Applied Physics, AlbaNova University Center, KTH Royal Institute of Technology, 106 91, Stockholm, Sweden

[4] Structural and Genomic Information Laboratory, UMR 7256 (IMM FR 3479, IM2B), Centre National de la Recherche Scientifique & Aix-Marseille University, Marseille 13288, France

[5] National Institute for Physiological Sciences, National Institutes of Natural Sciences, Okazaki, Aichi, Japan

[6] Department of Physiological Sciences, School of Life Science, The Graduate University for Advanced Studies (SOKENDAI), Okazaki, Aichi, Japan

* Correspondence and Lead author: Kazuyoshi Murata (kazum@nips.ac.jp)

Summary

Members of *Marseilleviridae*, one family of icosahedral giant viruses classified in 2012 have been identified worldwide in all types of environments. The virion shows a characteristic internal membrane extrusion at the five-fold vertices of the capsid, but its structural details need to be elucidated. We now report the 4.4 Å cryo-electron microscopy structure of the Melbournevirus capsid. An atomic model of the major capsid protein (MCP) shows a unique cup structure on the trimer that accommodates additional proteins. A polyalanine model of the penton base protein shows internally extended N- and C-terminals, which indirectly connect to the internal membrane extrusion. The Marseilleviruses share the same orientational organisation of the MCPs as PBCV-1 and CroV, but do not appear to possess a protein akin to the "tape measure" of these viruses. Minor capsid proteins named PC-β, zipper, and scaffold are proposed to control the dimensions of the capsid during assembly.

Keywords: cryo-electron microscopy; giant virus; NCLDV; single particle analysis; major capsid protein

42 **Introduction**

43

44 Members of the *Marseilleviridae* family belonging to the nucleo-cytoplasmic large DNA viruses
45 (NCLDVs) (Brandes and Linial, 2019) infect Acanthamoeba (Colson et al., 2013), and present a
46 highly complex ~360 kbp genome and icosahedral capsid with a diameter of ~250 nm. The
47 superfamily of NCLDVs was initially proposed by Iyer et al. (Iyer et al., 2001) and currently
48 consist of nine families and several subfamilies of viruses (Subramaniam et al., 2020). According
49 to ICTY classification, the members of the *Nucleocytoviricota* realm (PMID: 31187277) are
50 restricted to seven families of giant viruses possessing double-stranded DNA and targeting a
51 wide range of eukaryotic hosts. Since the first discovery of Marseillevirus in 2009 (Boyer et al.,
52 2009), the *Marseilleviridae* has expanded to more than 10 members and classified into five
53 lineages, A to E (Aoki et al., 2019; Fabre et al., 2017). Members of *Marseilleviridae* have been
54 found in some human patients, but their link with a particular disease is yet to be identified
55 (Aherfi et al., 2016; Kutikhin et al., 2014).

56

57 Melbournevirus is a member of the *Marseilleviridae* isolated from a freshwater pond in
58 Melbourne, Australia in 2014 (Doutre et al., 2014). It shares common genomic and structural
59 features with the other members of the family and classifies into lineage A. The first cryo-EM
60 structure was reported at 26 Å resolution by using a 200 kV microscope (Okamoto et al., 2018),
61 revealing the major capsid protein (MCP) array and a triangulation number of the icosahedral
62 capsid as T=309. In addition, characteristic morphological features of Melbournevirus capsids
63 were elucidated such as an extrusion of the internal membrane at the five-fold vertices of the

64 capsid and a nucleoid containing a large density body, the nature of which has not yet been
65 elucidated.
66
67 The sub-nanometer resolution structure of tokyovirus, one member of the *Marseilleviridae*, has
68 been reported using high voltage cryo-electron microscopy (cryo-HVEM) (Chihara et al., 2021).
69 The map did not reach a high enough resolution to be able to build *de novo* the structures of the
70 various minor capsid proteins (mCPs) and only allowed a semi-fitted model for the MCP and a
71 hypothetical fit of the PBCV-1 penton protein (Chihara et al., 2021). However, manual data
72 collection with the HVEM microscope restricted the number of images collected, and the current
73 resolution is limited to ~8 Å after using a capsid-focussed mask. The tokyovirus cryo-EM map
74 elucidated the novel capsid protein network of the giant virus including the scaffold proteins
75 which could assist in the formation of the characteristic internal membrane extrusion at the five-
76 fold vertices of the capsid and function in a similar manner to a tape measure protein (Xian et al.,
77 2020). Furthermore, it suggested a glycosylated cap protein was sitting on the MCP trimer
78 (Chihara et al., 2021).
79
80 High resolution cryo-EM maps of the giant viruses have been reported in three structures
81 exceeding 5 Å resolution: 3.5 Å for PBCV-1 (Fang et al., 2019), 4.6 Å and 4.8 Å for African
82 swine fever virus (ASFV) (Liu et al., 2019; Wang et al., 2019). These structures utilised 300 kV
83 microscopes and were reconstructed with a technique termed "block based" reconstruction (Zhu
84 et al., 2018) to improve obtained resolutions. Leveraging the high symmetry of icosahedral
85 viruses, this technique can decrease the size of boxes necessary for 3D reconstruction and allow
86 localized defocus refinement across the particle. For all objects >150 nm, this defocus gradient is

87 considerable and has a negative impact on the attainable resolution. Consequently, they
88 successfully constructed the atomic structures of the MCP and several mCPs in PBCV-1 and
89 ASFV. In PBCV-1, the structures of 13 different types of mCP were identified in addition to the
90 MCP (Fang et al. 2019), while in ASFV, the structures of four mCPs were determined in
91 addition to the MCP (Liu et al., 2019; Wang et al., 2019).
92
93 Here, we report the structure of Melbournevirus to 4.9 Å for the whole particle, with only 3,124
94 particle images, and a maximum resolution of 4.42 Å obtained on the acquired dataset for the
95 five-fold, three-fold and two-fold "blocks", using symmetry expansion, recentring and defocus
96 refinement. This is functionally the same as block-based reconstruction (Zhu et al., 2018) which
97 has been applied to the study of other giant viruses but operates entirely within the RELION
98 (Fernandez-Leiro and Scheres, 2017; Scheres, 2012; Zivanov et al., 2018; Zivanov et al., 2020)
99 processing suite. The capsid, particularly in the block-based reconstructions, permits clearer
100 views of the complex lattice of capsid proteins. While the hard limit of resolution in this data
101 makes residue identification difficult, poly-alanine chain models could be created in some cases.
102 An atomic model of the MCP, and a polyalanine model of the penton base protein which has no
103 identified polypeptide sequence were built and compared with other giant viruses reported
104 previously. The Marseilleviruses share the same orientational organisation of the major capsid
105 proteins as PBCV-1 and CroV, but do not appear to possess a protein akin to the "tape measure"
106 of these viruses, and the internal scaffold array does not extend to the pentasymmetron. This
107 means some other mechanism to control the dimensions of the capsid during assembly, possibly
108 utilising the PC-β, zippers, and scaffold proteins in concert, is required. These differences
109 demonstrate the variety and biological flexibility of the giant viruses, while simultaneously

110 utilising similar strategies for some critical features. The cryo-EM map of Melbournevirion
111 provides clearer views of the complex lattice of the capsid, and the molecular interactions
112 between capsid proteins, revealing the functional viral capsid network.
113
114
115 **Results**
116
117 **The structure of Melbournevirus**
118 The image processing flowchart and the dataset collection details of the whole-virion 3D
119 reconstruction is shown in Fig. S1 and Table 1. Briefly, 3,367 micrograph movies of ice
120 embedded Melbournevirions, in two independently collected datasets, consisting of 20 and 40
121 frames per micrograph respectively, were collected with a 300 kV Titan Krios (Thermo Fisher
122 Scientific) microscope using automated image acquisition software. The image data were
123 processed in the RELION 3.1 suite (Zivanov et al., 2018; Zivanov et al., 2020) throughout, as
124 described in Methods. Using Ewald sphere correction, a 4.9Å global resolution map of
125 Melbournevirus was reconstructed by assuming icosahedral symmetry (Fig. 1, Fig. S1F).
126 Externally, Melbournevirus shares a high structural similarity to that of tokyovirus (Chihara et
127 al., 2021), another member of the *Marseilleviridae*, sharing the same triangulation number
128 (T=309), and the same MCP pattern on the capsid surface. Like tokyovirus, it shows a uniform
129 MCP array each with a lower resolution “cap” density (Chihara et al., 2021). Internally, the inner
130 membrane shows the characteristic extrusion at the five-fold axis, and the scaffold network
131 between the capsid layer and inner membrane is comparable to that of tokyovirus (red circles in
132 Fig. 1B). A low resolution "plug" of density fills a space between the pentasymmetron and the

133 internal membrane extrusion (arrow in Fig. 1B), although it does not directly contact the internal
134 membrane. These components may be flexible in the structure because they are not well resolved
135 even in the non-symmetrized map of the block-based reconstruction (Fig. 2).
136
137 With this Melbournevirus dataset, we approached the resolutions at which *de novo* model
138 building can be attempted. Using the "block based" reconstruction strategy (Zhu et al., 2018)
139 permitted further improvements in resolution (Fig. S2). Blocks were reconstructed with no
140 symmetry applied (C1), with a soft mask focussed on the capsid. Carrying out the block
141 reconstructions without defocus refinement barely improved resolution above that of the Ewald-
142 sphere corrected whole virus reconstruction (the global FSC for the fivefold block estimated
143 4.8Å) but a single pass of defocus refinement resulted in the reconstructions reaching the
144 maximum resolution of 4.42 Å (Fig. S2-S5, Table 2).
145
146 Figure 2 shows the capsid viewed from the inside, filtered by local resolution, and focussed on
147 the asymmetric unit encompassing a two, three and fivefold symmetry axes, describing the
148 details of the protein network formed with mCPs. Each component is labelled according to the
149 nomenclature proposed in the segmentation of tokyovirus by Chihara *et al.* (Chihara et al., 2021)
150 and closely follows the motifs previously described. These are individually detailed as follows.
151
152 **The molecular structure of the individual proteins**
153 While the genome of Melbournevirus has been sequenced, most of the capsid proteins are
154 unannotated with a few hypothetical assignments (Okamoto et al., 2018). As a result, we built a

155 model of the MCP, and carried out chain tracing and fitting of poly-alanine chains for some other
156 regions of density which could be segmented and followed without significant ambiguity.
157
158 ***The major capsid protein***
159 The amino acid sequence of the Melbournevirus MCP was aligned against those of tokyovirus,
160 PBCV-1 and other NCLDVs, where it scores highly against tokyovirus, and within acceptable
161 parameters against Iridovirus and PBCV-1 (Fig. S6). The MCP was predicted to possess the
162 same "double jelly roll" motif as other NCLDV MCPs (Fig. S7). This is confirmed by the
163 striated density on the MCP trimer extracted from the centre of the threefold axis (Fig. 3A),
164 which show a β-sheet structure. A homology model of the Melbournevirus MCP was built by
165 using the PBCV-1 MCP (PDB ID: 5TIQ) as a structural template, and flexible-fitted to the
166 density, clearly showing the molecular model of the MCP trimer (Fig. 3B). As previously
167 reported for PBCV-1 (Fang et al. 2019) and ASFV (Wang et al. 2019), the MCP trimer is formed
168 with interactions between the FG1 loops which extends inside the trimer with the two short α-
169 helices (dotted red circle in Fig. 3D), and stabilized with the N-terminal domain (NTD) extended
170 from the adjacent monomer which functions like an anchor (arrow in Fig. 3D). Furthermore, in
171 the Melbournevirus MCP trimer, the elongated HI1 loop interacts with the adjacent MCP
172 monomer (arrow heads in Fig. 3D-E, Fig. S7A), and forms a unique cup structure with the
173 relatively long loops of DE1 and FG1 on the top of the MCP trimer (Fig. 3E). The cup structure
174 serves to accommodate a similarly uniform and symmetric "cap" region to tokyovirus (Chihara
175 et al., 2021) which cannot be attributed to the MCP polypeptide (Fig. 3B, C). Previously,
176 Okamoto et al. reported that an uncharacterised protein, MEL_236, appeared to have the same
177 abundance as the MCP protein (Okamoto et al., 2018). At 16kDa, MEL_236 is approximately

178 the correct size (given a margin of error for SDS-PAGE size determination) to correspond with
179 the PAS stain-sensitive 14 kDa protein of tokyovirus previously reported (Chihara et al., 2021).
180 Genomic studies of Noumeavirus reported the orthologue of this, NMV_189, as the most
181 abundant protein (Fabre et al., 2017). Unfortunately, in these maps the cap region is not clear
182 enough to build a model from this sequence (arrowheads in Fig. S3D and S4D).
183
184 ***The penton base protein***
185 As reported in other giant virus capsids (Chihara et al. 2021, Fang et al. 2019, Wang et al. 2019,
186 Born et al., 2018), the five-fold axis of the Melbournevirus capsid is filled with a penton base
187 protein pentamer. We performed a rigid body fit of the penton base protein from PBCV-1 (Fang
188 et al., 2019) to this density, but the quality of the fit was marginal. Therefore, the penton base
189 pentamer was extracted from the fivefold block reconstruction (Fig. S3) using UCSF Chimera
190 (Pettersen et al., 2004) and SEGGER (Pintilie and Chiu, 2012), and DeepTracer (Pfab et al.,
191 2021) used with this extracted density together with a short polyalanine sequence to trace the
192 backbone (Fig. 4). The penton protein exhibited a single jelly roll (SJR) motif similar to those of
193 other giant viruses (Fig. S8). The pentamer is formed by two interactions, external and internal;
194 the external is the loop extended inward of the pentamer in the outer capsid region (Fig. 4C), and
195 the internal is the interactions between the N-terminals (black asterisks in Fig. 4D) and between
196 the N-terminal and C-terminal (red asterisks in Fig. 4D). The entire structure of the penton base
197 protein is similar to that of PBCV-1, which does not have the large insertion domain found in
198 CroV-dependent mavirus (Fig. S8). In the cryo-EM map, some inner capsid densities also exist
199 under the penton base protein models (arrow in Fig. 4). Another component may exist in this
200 region and connect the penton to the inner membrane extrusion.

201

202 The five-fold axis of Melbournevirus capsid with the characteristic extrusion of the inner
203 membrane shows a low resolution "plug" of density which fills the area underneath the penton,
204 completely enveloping the pentasymmetron component proteins (Fig. 1B), but this plug does not
205 directly contact the inner membrane extrusion.

206

207 ***The pentasymmetron components***

208 As the minor capsid proteins underneath the pentasymmetron, three pentasymmetron
209 components (PC-α, PC-β, PC-γ) have been identified in tokyovirus (Chihara et al. 2021). Similar
210 components were also identified in Melbournevirus (Fig. 2). With the improved resolution of the
211 Melbournevirus block-based reconstruction, we can clearly see one of the pentasymmetron
212 component proteins (named as PC-α (Chihara et al., 2021)) revealing two quadruple bundles
213 (each reminiscent of a ferritin monomer (Kayama et al., 2021)) of α-helices, stacked one on top
214 of the other (Fig. 2, Fig. S9). A polyalanine model was built *de novo* and fitted into the density.
215 However, it is less clear whether this is a single protein or two independent ones, as the
216 extremity of this region is relatively unsupported and as a result is ~6Å (Fig. S3). Our current
217 hypothesis is that they are two independent bundles, where the polypeptide chain continues into
218 the capsid framework and are located surrounding the inner membrane extrusion (Fig. S8, Fig.
219 1B).

220

221 ***Other minor capsid proteins***

222 Like tokyovirus, the mCPs of Melbournevirus form an intricate lattice network (Fig. 2). The two
223 viruses, extremely close in genetic terms, should demonstrate high structural similarity. Thus,

224 Melbournevirus shows the same lattice array with the trapezoidal lattice consisting of the lattice
225 protein (salmon in Fig. 2), which are linked by cement proteins (pale blue in Fig. 2) and with a
226 glue/zipper protein arrangement (pink and orange, respectively, in Fig. 2) along the trisymmetron
227 interface. These are internally supported by other proteins; one is the scaffold protein, and the
228 other is the support protein (Fig. 2). The scaffold proteins (yellow in Fig. 2) extend from the
229 "horseshoe" shaped terminus (bracket in Fig. 2) of one scaffold array near the pentasymmetron
230 to the scaffold array along the trisymmetron interface. The support proteins (dark blue in Fig. 2)
231 showed a large density and three additional smaller densities, running parallel to the
232 trisymmetron interface and the cement protein array. The three additional small densities were
233 not observed in tokyovirus capsid at 7.7 Å resolution (Chihara et al. 2021). As the scaffold and
234 support proteins (yellow and dark blue, respectively, in Fig. 2) are underneath the lattice layer of
235 the capsid, they are more flexible and as such resolution suffers when the map is filtered by local
236 resolution, however without local filtering they are exceedingly difficult to see (Fig. 2). The glue
237 proteins are located on the edge of the trisymmetron and serve to connect the adjacent
238 trisymmetron. The bisymmetric structure looks like a dimer of the protein, but it is unclear at this
239 resolution. The zipper protein consists of ten components (orange in Fig. 2), eight of which are
240 located along the trisymmetron interface and bind to each of the two-fold symmetric glue
241 proteins (pink in Fig. 2). The remaining terminal two zipper proteins are located between the
242 glue proteins and the pentasymmetron components (arrowhead in Fig. 2), and their orientation is
243 significantly different from the other zipper proteins (arrow in Fig. 2).
244
245 **Orientation of MCP trimers on the capsid surface**

246 The pentasymmetron assembly of Melbournevirus shares the MCP orientation ("golf club")
247 motif displayed in CroV (Xiao et al. 2017) and PBCV-1 (Xian et al. 2020), in addition to the
248 trisymmetron (Fig. 5). In the motif, the corner, P_{nd} (n=1,2,3,4,5), of the MCP trimer in the
249 asymmetric unit shows a 60° rotation to match that of the adjacent asymmetric unit. This is
250 curious as Xian *et al.* hypothesise this orientational variance in the pentasymmetron asymmetric
251 unit to be caused by the tape measure protein (Xian et al. 2020), which Melbournevirus does not
252 exhibit, but instead appears to use the scaffold array (Fig. 2) and some other mCPs which were
253 originally reported by Chihara *et al.* (2021). Examining underneath this "golf club" motif
254 positions PC-α and PC-β at the interfaces of the individual motifs, with PC-γ at the interface of
255 the motif and the neighbouring trisymmetron (white circles in Fig. 5). Looking at the minor
256 capsid protein network in the Melbournevirus capsid, the function of the tape measure proteins is
257 possibly replaced by the coordinated use of the PC-β, zippers, and scaffold proteins (dotted curve
258 in Fig. 2).
259
260
261 **Discussion**
262
263 Achieving close to the Nyquist limit of this dataset when reconstructing the complete viral
264 particle with Ewald sphere correction demonstrates how vital proper analysis and correction of
265 the various physical (electro-optical) effects can be for the study of giant viruses. The Ewald
266 sphere reconstruction and solvent-corrected FSC curve does not fall to zero (black FSC curve,
267 Fig. S1J). Therefore, we also show the unmasked (non-solvent corrected) Ewald sphere-
268 corrected curve (grey curve, Fig. S1J), and the respective FSC curves without Ewald sphere

269 correction (maroon and orange, respectively, Fig. S1J) all of which reach zero, indicating that the
270 Ewald sphere-corrected FSC curve not reaching zero is a result of how close it is to maximum
271 resolution. In our study, a processing method analogous to block-based reconstruction (Zhu et
272 al., 2018) (Fig. S2-5) was used to improve the clarity of the map further. When we used 1MV
273 cryo-HVEM for single particle analysis (SPA) of tokyovirus (Chihara et al., 2021), Ewald sphere
274 correction did not improve the resolution. This suggests that the electron accelerated 300 kV
275 microscope has a significant defocus gradient limitation for giant viruses above 200 nm,
276 although symmetry expansion and block-based reconstruction helps to counter this. Further,
277 block-based reconstruction permits subtle variation in particle orientation, which is not possible
278 with symmetry imposition. This may also play a role in improving resolution in the block
279 reconstructions.
280
281 This Melbournevirus cryo-EM 3D reconstruction has a comparatively low particle count of
282 3,124 particles when compared to the previously reported structures of PBCV-1 (~13,000 for a
283 4.4 Å reconstruction (Fang et al., 2019)) and ASFV (16,266 for a 14.1 Å reconstruction (Liu et
284 al., 2019), 63,348 for an 8.8 Å reconstruction (Wang et al., 2019)). We attribute this to both
285 biological and computational reasons. ASFV possesses an outer membrane which will negatively
286 impact the resolutions reported for a whole virus reconstruction due to its disordered nature,
287 which neither PBCV-1 nor Melbournevirus possess. PBCV-1, on the other hand, achieves a
288 higher whole virus resolution, which can be considered consistent with the increased particle
289 count. It is also ~25% smaller than Melbournevirus, at 190 nm, so the intra-particle defocus
290 gradient should not be as large. Computationally, improvements in image processing software,
291 specifically, magnification anisotropy estimation and Ewald sphere correction played a

292 significant role in improving our Melbournevirus reconstruction (see Methods). As these
293 computational methods advance further, we look forward to seeing further improvements for the
294 study of giant viruses.
295
296 The Melbournevirus MCP is structurally analogous to the PBCV-1 (Fang et al., 2019),
297 Faustovirus (Klose et al., 2016) and ASFV (Liu et al., 2019; Wang et al., 2019) MCPs, with the
298 sequence similarity including a double jelly roll motif each composed of eight β-strands (Benson
299 et al., 2004). The greatest level of similarity is with tokyovirus (Fig. S6), where only a few
300 residues differ, which would suggest the MCP structure is highly preserved in *Marseilleviridae*.
301 With the improved resolution of the Melbournevirus block reconstructions, we could better
302 visualise the HI1 loop of Melbournevirus MCP, contacting the HI2 loop of the adjacent
303 monomer and forming a tight trimeric complex of MCP (Fig. 3D-E, Fig. S7A). This differs from
304 ASFV, where it is the FG1 loop which is extended (Liu et al., 2019), and from PBCV-1, where a
305 short FG1 loop is in close proximity to the HI2 loop of the neighbouring MCP (De Castro et al.,
306 2018). Furthermore, the HI1 loop and the other external loops of DE1 and HI2 extend externally
307 (Fig. 3B, Fig. S7) and form a cup structure on the top of the MCP trimer to accommodate
308 additional cap densities (orange in Fig. 3C). PBCV-1 MCP has a highly glycosylated "cap"
309 region, while ASFV and faustovirus have a large additional β-sheet rich domain (Fig. S7B-D).
310 Melbournevirus, however, has an ordered region, which is still lower resolution than the main
311 body of the MCP trimer (Figs. S3D and S4D), which does not correspond to any part of the MCP
312 model (Fig. 3B,C). This ordered structural motif would imply the presence of protein, rather than
313 sugars, which tend to be highly mobile. However, the reduced clarity of the density compared to
314 that of the MCP may be caused by several factors. Firstly, the cap may be in multiple

315 conformations across the capsid surface, resulting in loss of clarity due to icosahedral averaging.
316 Secondly, the cap may present partial occupancy of each MCP trimer across the capsid surface,
317 also resulting in loss of clarity due to icosahedral averaging. To conclude this, further
318 investigation is necessary to identify the cap protein and determine the stoichiometry with the
319 MCP. As Fabre *at al.* (Fabre et al., 2017) and Okamoto *et al.* (Okamoto et al., 2018) report
320 MEL_236 at equal or greater stoichiometry to the MCP protein, we can strongly assume that
321 MEL_236 corresponds to the 14kDa glycosylated cap protein we previously reported (Chihara et
322 al., 2021). Aoki et al. reported that in hokutovirus and kashiwazakivirus, which belong to lineage
323 B of *Marseilleviridae*, the newly replicated viruses are aligned on the surface of the host cells,
324 exhibiting a "bunch" formation consisting of normal and infected cells, similar to tupanvirus
325 (Aoki et al. 2019). The flexible characteristics of this glycosylated cap protein on the MCP
326 trimer may be different within the members of *Marseilleviridae*, playing a different functional
327 role in the interface with the viral host.
328
329 With the improvement in resolution of this Melbournevirus reconstruction, we have been able to
330 build a polyalanine model of the penton base protein. The DeepTracer (Pfab et al., 2021) chain
331 tracing predicted a SJR fold. When comparing the DeepTracer predicted structure to that of the
332 PBCV-1 penton base protein (Fang et al., 2019), the SJR fold aligns, but the less structured loops
333 are of varying length. The polyalanine model fits the density well, however the loops show some
334 dislocation from that of PBCV-1 (Fig. S8), which may explain why we have been unable to
335 identify candidate penton protein primary sequence *via* BLAST (Altschul et al., 1990; Johnson et
336 al., 2008) sequence homology. The CroV-dependent mavirus penton (Born et al., 2018), which
337 we previously tested against tokyovirus, again aligns at the SJR fold but is otherwise a poor fit,

338 even when the external insertion domain (curly bracket in Fig. S10) which is absent in PBCV-1
339 and Marseilleviruses, is ignored. The penton shows more density underneath the fitted
340 polyalanine model (arrow in Fig. 4B). Weng et al. reports that the penton base pentamer hung
341 lantern-like densities underneath the protein, connecting the penton and inner membrane (Wang
342 et al. 2019). In the Melbournevirus penton, we do not find the lantern-like density under the
343 penton. The low-resolution "plug" of density following the penton may have a similar function to
344 connect the penton with the internal membrane extrusion.
345
346 There are two orientation combinations between MCP trimers, a non-rotating combination and a
347 60°-rotating combination (Wang et al. 2019). In the non-rotating combination, a jelly roll motif 1
348 (JR1) in the MCP trimer interacts with a jelly roll motif 2 (JR2) in the adjacent MCP trimer,
349 which creates a relatively flat MCP array, like the trisymmetron surface. In contrast, in the 60°-
350 rotating combination, a JR1 in the MCP trimer interacts with a JR1 in the adjacent MCP trimer,
351 while a JR2 in the MCP trimer interacts with a JR2 in the adjacent MCP trimer. These
352 combinations create a relatively angled edge, like the trisymmetron interface. In the
353 pentasymmetron, the orientation of the MCP trimers is more complicated. Five asymmetric units
354 of the pentasymmetron, ideally each consisting of six non-rotating combination MCP trimers,
355 surround the central penton protein, and the interface of each asymmetric unit interacts with the
356 60°-rotating MCP trimer of the neighboring asymmetric unit. Finally, these MCP trimers creates
357 a penton vertex (Fig. S11). However, Xiao et al. reported in PBCV-1 and CroV that one of six
358 MCP trimers of the asymmetric unit, which located at the corner of the pentasymmetron, was
359 exceptionally rotated 60° (Xiao et al. 2017, Xian et al. 2020). They called this MCP trimer array
360 of the asymmetric unit the "golf club motif". Furthermore, the "golf club motif" was

361 hypothesized to be caused by specific localization of the tape measure proteins (Xian et al.
362 2020). In our observations, Melbournevirus showed a "golf club motif" of the MCP trimer array
363 in the same orientation as PBCV-1 and CroV (Fig. 5), but Melbournevirus lacks a tape measure
364 protein. However, the rotation of the single MCP trimer by 60° (P_{1d}, P_{2d}, P_{3d} in Fig. 5) orients it
365 to match that of the adjacent trisymmetron and penton asymmetric unit. If we picture all the
366 MCP trimers in a pentasymmetron asymmetric unit in the same orientation (Fig. S11), we can
367 better visualise why the rotation of a single MCP trimer may occur. The single MCP trimer
368 would cause a mismatch in the MCP orientational alignment of the trisymmetron interface
369 (arrows in Fig. S11). The trisymmetron interface demonstrates a greater angle across it than the
370 trisymmetron, which is of a gentler curvature. As such, it is likely that this orientational
371 arrangement is a result of improving the flexibility of the capsid across the trisymmetron
372 interface.
373
374 Without a tape measure protein, we propose that PC-β, which is present underneath the interface
375 between the 60°-rotated P_d MCP trimer and the P_b and P_e MCP trimers, may play a role in
376 determining trimer orientation. In a similar manner, PC-α may support the interface between the
377 penton and two asymmetric units, while PC-γ supports the interface between the
378 pentasymmetron and the adjacent trisymmetron. However, these components do not directly
379 contact with the 60°-rotated P_d MCP trimer. While we have previously proposed that the scaffold
380 protein array acts as an equivalent to the tape measure protein, the horseshoe terminus of the
381 scaffold does not appear to be able to directly interact with any of the PC proteins, as the
382 glue/zipper protein array is present in between the horseshoe terminus and either PC-β or PC-γ
383 (Fig. 2). PC-β may interact with the terminal two zipper proteins, which have a rotated

384 orientation when compared to the remaining eight zipper proteins running down the edges of the
385 glue proteins along the trisymmetron interface, while PC-γ interacts with the glue proteins
386 themselves along two trisymmetron interfaces. As the scaffold proteins run parallel to the
387 glue/zipper array, some interaction may also occur there, with the three acting in concert in a
388 similar manner to the tape measure protein in controlling capsid construction. Therefore, the
389 combination of PC-β, zipper proteins, and scaffold proteins (dotted curve in Fig. 2) is most likely
390 to function as the tape measure protein in Melbournevirus.
391
392 The lattice proteins show strong similarity to the motifs displayed in the lower resolution
393 tokyovirus reconstruction. By using local resolution filtering in RELION 3.1 (Zivanov et al.,
394 2020), we were able to avoid losing clarity in the more rigid lattice proteins, while
395 simultaneously visualising the more flexible scaffold array (Fig. 2). Interestingly, this permitted
396 visualisation of three additional weak densities extending from the previously reported support
397 protein (dark blue in Fig. 2), which run parallel to the zipper/glue proteins and cement proteins
398 on the trapezoidal lattice, but which themselves were not previously observed in tokyovirus.
399 Missing them in tokyovirus was likely due to resolution limitations and caused by decreased
400 signal to noise from fewer particles and lack of dose weighting for HVEM data (Chihara et al.,
401 2021). These may act to further support the lattice array as the array extends from the
402 pentasymmetron construct along the trisymmetron interface, in addition to supporting the
403 characteristic internal membrane structure.
404
405 Unfortunately, while at 4.42 Å resolution we can see some secondary structure, most if not all
406 residue sidechains are unclear. Furthermore, at many points it is extremely difficult to determine

407 which direction of two apparently divergent paths the density should follow. As the primary
408 sequence is unknown, and large or interacting sidechains can create a density bridge which is
409 difficult to discriminate from backbone density, we erred on the side of caution for the time
410 being. One of the mCPs, a repeating segment of the lattice protein (salmon in Fig. 2), has a
411 section exposed on the internal face of the capsid (following the chain into the mCP layer created
412 too much ambiguity in direction of the polypeptide) which fits a short section consisting of three
413 linked α-helices (Fig. S10).
414
415 Here we have increased the resolution of a cryo-EM 3D reconstruction of a Marseillevirus to a
416 resolution approaching that necessary to build *de novo* structural models. This has permitted us
417 to improve our previous Marseillevirus MCP model and begin the process of building a whole-
418 capsid model. This is currently limited by both a lack of comprehensive annotation of capsid-
419 related proteins and reaching the maximum resolution possible with this dataset. To fully clarify
420 the structure of Melbournevirus, and build a complete model of the *Marseilleviridae* capsid, we
421 must fully utilize our understanding of the genetics of the *Marseilleviridae*, using copy numbers
422 to help to identify the structural components and selected knockout genes to retrieve a
423 phenotype, whilst simultaneously achieving resolutions that allow *ab initio* structure
424 determination.
425
426
427 **Methods**
428
429 ***Viral propagation and purification***

430 Melbournevirus particles were propagated and purified as described previously (Doutre et al.,
431 2014; Okamoto et al., 2018). It is described briefly as follows. Seed Melbournevirus particles
432 were propagated in *Acanthamoeba castellanii* cells that were cultured in PPYG medium and
433 confluent in ten 75 cm^3 culture flasks. The infected culture fluid was collected three days post
434 infection and centrifuged for 10 min at 500 *g*, 4ºC. The supernatant was centrifuged for 35 min at
435 6,500 *g*, 4ºC, and the pellet was suspended in 1 mL of PBS (-). The sample was ultracentrifuged
436 in 10-60% (w/v) continuous sucrose gradient for 90 min at 6,500 *g*, 4ºC. The virus particles'
437 band was collected from the sucrose gradient, and centrifuged for 35 min at 6,500 *g*, 4ºC. The
438 pellet was rinsed five times in 50 mL of PBS (-) by centrifugation to remove excess sucrose.
439 Finally, the virus particles were suspended in 100 µL PBS (-) for further cryo-EM analysis.
440
441 ***Cryo-EM sample preparation and acquisition***
442 Three µL of purified virus particles were loaded on a glow-charged lacey carbon copper grid and
443 flash frozen in liquid ethane after 2 s blotting in 100% humidity at 4ºC using a Vitrobot Mark IV
444 (Thermo Fisher Scientific). The particles images were collected using a 300 kV Titan Krios G2
445 microscope (Thermo Fisher Scientific) equipped with a K2 Summit direct electron detector using
446 counting mode and a GIF energy filter of 20 eV slit width (Gatan Inc). Two datasets were
447 collected, dataset 1 with 20 frames per micrograph movie, and dataset 2 with 40 frames per
448 micrograph movie. These datasets were taken at a nominal magnification of 64,000× (pixel size
449 on specimen is 2.21 Å/pixel), at 0.8 – 3.0 µm under focus, and at 26.4 $e^-/Å^2$ total dose.
450
451 ***Image processing***

452 The image processing workflow of the whole virus particle in summarised graphically in Fig. S1.
453 3,367 micrograph movies were imported into RELION 3.1 (Zivanov et al., 2018; Zivanov et al.,
454 2020) and motion correction performed using MotionCor2 (1.4.0) (Zheng et al., 2017) using
455 whole frame motion correction. Contrast transfer function (CTF) parameters were calculated on
456 the resulting micrographs using CTFFIND 4.1.14 (Rohou and Grigorieff, 2015), using the
457 RELION 3.1 default parameters except for defocus step size, which was set to 100 Å. A total of
458 229 particles were manually picked, extracted using 4× downsampling (to 8.84 Å/pixel in 360-
459 pixel boxes), and 2D classified into 10 classes with 0.5° angular sampling and "ignore CTFs
460 until first peak" enabled. Three classes were used to generate three initial models using I3
461 symmetry with a mask diameter of 2550 Å and an initial angular sampling of 3.7°, of which the
462 best was chosen. These three clear classes, comprising 182 particles, were used as references for
463 RELION template autopicking of the full dataset, resulting in 7,000 particles picked. These were
464 extracted using 4× downsampling and 2D classified into 40 classes with a mask diameter of 2650
465 Å, angular sampling of 1° and "ignore CTFs until first peak" enabled. 19 classes, totalling 5,047
466 particles were selected in clear classes and subjected to a further round of 2D classification, this
467 time with 0.5° angular sampling and full CTF correction. This resulted in 3 clear classes,
468 containing 3,852 particles. 3D classification into five classes was carried out with an angular
469 sampling of 1.8° and "ignore CTFs until first peak" enabled. Three classes, totalling 3,569
470 particles, were selected from the 3D classification and a 3D refinement carried out. The
471 maximum resolution possible with the downsampled particles was achieved (17.68 Å). Particles
472 were re-extracted using 2× downsampling (4.42 Å/pixel in 720-pixel boxes) and a further 3D
473 refinement and single round of per-particle defocus refinement again achieved the maximum
474 resolution possible with downsampled particles (8.84 Å). Particles were re-extracted without

475 downsampling (2.21 Å/pixel) which permitted a reconstruction to achieve 7 Å. Unfortunately,
476 CTF refinement with 1,440-pixel boxes would fail. Particles were re-extracted with 1.5×
477 downsampling (3.315 Å/pixel in 960-pixel boxes) and anisotropic magnification distortion
478 estimated, followed by further defocus refinement. This resulted in a 6.65 Å reconstruction. 3D
479 classification with alignment disabled was carried out into five classes, with the two best classes,
480 comprising 3,124 particles, selected, resulting in a 6.45 Å reconstruction. Ewald sphere
481 correction was used with *relion_reconstruct* to generate independent half-maps using the
482 parameters from this final refinement, which results in a final gold standard FSC (Chen et al.,
483 2013) global estimated resolution of 4.9 Å with a soft capsid only mask.
484
485 At this point, we moved toward a strategy akin to that of block-based reconstruction (Zhu et al.,
486 2018), but only using the functions contained within RELION 3.1 (Zivanov et al., 2020). The
487 image processing workflow of the block-based reconstruction is summarised in Fig. S2. First,
488 symmetry expansion was carried out. Then three masks, individually focussed on a five-, three-
489 and two-fold axis were generated in UCSF Chimera (Pettersen et al., 2004), before particle
490 subtraction was used to calculate the shifts to centre the focussed region in the box. Particle
491 subtraction was cancelled after shifts were calculated. Using these RELION-calculated shifts, the
492 five-, three- and two-fold axes were extracted in 440-, 500- and 700-pixel boxes respectively,
493 each containing a total of 187,440 boxes. All block reconstructions were carried out with C1 (no)
494 symmetry imposed. The *relion_reconstruct* module was used for each extracted particle set to
495 generate a reference for refinement and check positioning was correct. 3D refinements were
496 carried out for each segment, achieving 4.8-5 Å with post-processing. Defocus refinement was
497 carried out once for each, and *relion_reconstruct* used to reconstruct half-maps using the

498 defocus-refined particle parameters. The five-fold and three-fold reconstructions achieved 4.42
499 Å, the maximum resolution possible with these data. The two-fold reconstruction achieved 4.43
500 Å, likely caused by the larger box size used. Further CTF parameter optimisation and Bayesian
501 polishing were not carried out.
502

503 ***Model building***

504 The amino acid sequence of the MCP of Melbournevirus is annotated, along with four other
505 candidates which may be mCPs previously reported by Okamoto *et al.* (Okamoto et al., 2018). A
506 homology model of the MCP was generated using SWISS-MODEL (Waterhouse et al., 2018)
507 using a structural template of that of PBCV-1 (PDBID: 5TIQ). This model was rigid body fit into
508 a SEGGER-extracted (Pintilie and Chiu, 2012) MCP from the centre of the threefold axis and
509 fitted into the map using ISOLDE (Croll, 2018) with restraints on predicted secondary structure.
510 Real-space refinement was carried out in PHENIX (Adams et al., 2010) with intermediate rounds
511 of geometry minimisation.
512

513 The PDB model of the PBCV-1 penton protein was rigid-body fit to the Melbournevirus penton
514 density. It was deemed a poor fit, so the map blocks were segmented using SEGGER (Pintilie
515 and Chiu, 2012) and extracted segments were processed with DeepTracer (Pfab et al., 2021)
516 using a poly-alanine sequence for chain tracing. The best of these poly-alanine chain models was
517 rigid-body fit to fill the penton with five identical chains. ISOLDE (Croll, 2018) was used to
518 better fit some sections of the chain. Real-space refinement and validation were carried out in
519 PHENIX (Adams et al., 2010). Other parts of the mCPs (PC-α and Lattice protein) were also
520 process with the same way and the polyalanine models were built.

521

522 ***Data availability***

523 The density maps have been deposited in the EMDB with the accession codes WXYZ (whole
524 virus), WXYZ (five-fold block), WXYZ (three-fold block) and WXYZ (two-fold block).

525

526

527 **Acknowledgements**

528 We thank Marta Carroni at SciLifeLab cryo-EM facility in Stockholm for data collection. This
529 study was funded by the following grants: MEXT KAKENHI (JP19H04845 to K.M.), the Joint
530 Research of ExCELLS (20-004 to K.M.), Vetenskapsrådet (VR)/The Swedish Research Council
531 (to K.O., grant no. 2018-03387), the Swedish Foundation for International Cooperation in
532 Research and Higher Education (STINT) (to Janos Hajdu and K.O., grant no. JA2014-5721),
533 FORMAS research grant from the Swedish Research Council for Environment, Agricultural
534 Sciences, and Spatial Planning (to K.O., grant no. 2018-00421), the Royal Swedish Academy of
535 Sciences (to K.O., grant no. BS2018-0053).

536

537 **Author contributions**

538 K.O. and K.M. conceived the project. C.A. and M.S. provided Melbournevirus sample. H.K.N.R.
539 and K.O. prepared cryo-EM sample. H.K.N.R. and K.O. collected cryo-EM data. R.N.B.S.
540 processed cryo-EM data. R.N.B.S., and K.M. prepared figures and wrote the main manuscript
541 text. All authors reviewed the manuscript.

542

543 **Competing interests**

544 The authors declare there to be no competing financial interests.

545

546 **Tables**

547 **Table 1.** Dataset collection details

Melbournevirus data collection	Dataset 1	Dataset 2
Microscope	FEI Titan Krios G2	
Accelerating voltage (kV)	300	
Spherical aberration (mm)	2.7	
Detector	Gatan K2 Summit	
Total dose (e-/A^2)	26.4	
Micrographs	1,240	2,127
Frames per micrograph	20	40
Nominal magnification	64,000×	
Pixel spacing on the specimen (Å/pixel)	2.21	
Target defocus range (μm)	-0.8 – -3.0	
Initial particles picked	7,000	
Final particles used	3,124	

548

549

550 **Table 2.** Reconstruction details

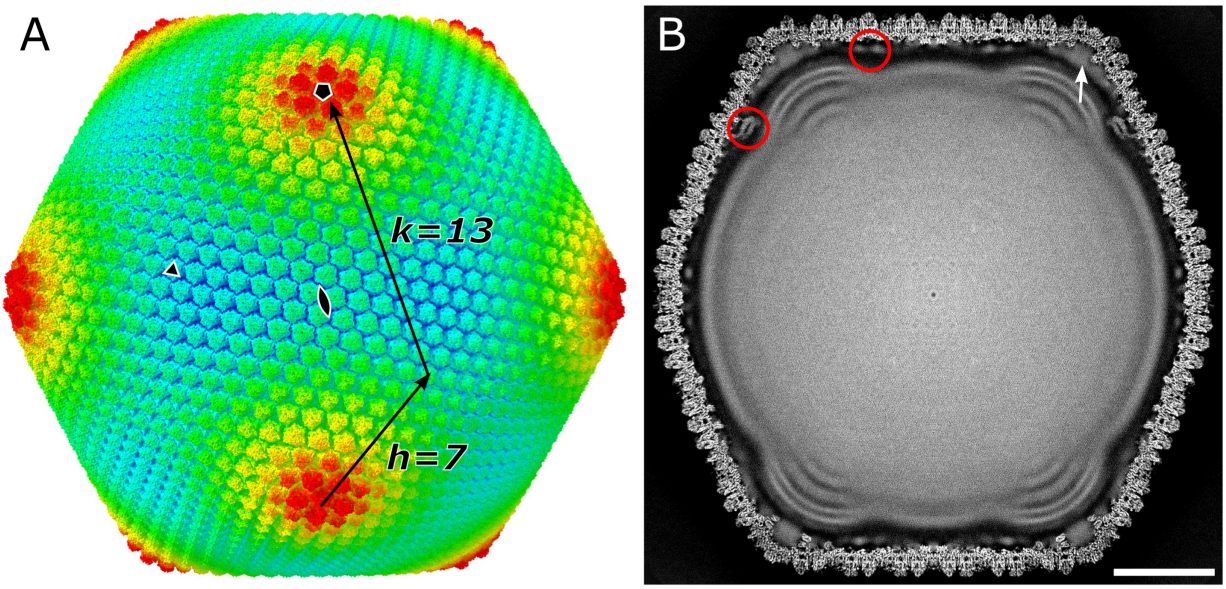
Reconstruction	Whole virus	Five-fold block	Three-fold block	Two-fold block
Particles used	3,124	187,440	187,440	187,440
Symmetry imposed	I3	C1	C1	C1
Resolution	4.9	4.42	4.42	4.43
EMDB accession code				

551

552

553 **Figures**

554



555

556

557 **Figure 1.** SPA 3D reconstruction of Melbournevirus at 4.9 Å. A) External view of

558 Melbournevirus, coloured by radius with the following parameters: blue, 1,050 Å, turquoise,

559 1,100 Å; green, 1,150 Å; yellow, 1,200 Å; red, 1250 Å, and symmetry axes are marked by a

560 pentagon (five-fold), triangle (three-fold) and double teardrop (two-fold). Index of h and k for

561 triangulation number is indicated. B) Central slice of Melbournevirus. Red circles indicate the

562 scaffold proteins between the capsid layer and inner membrane. A white arrow indicates a low

563 resolution "plug" of density which fills a space between the pentasymmetron and the inner

564 membrane extrusion but does not directly contact the inner membrane. Scale bar equals 50 nm.

565

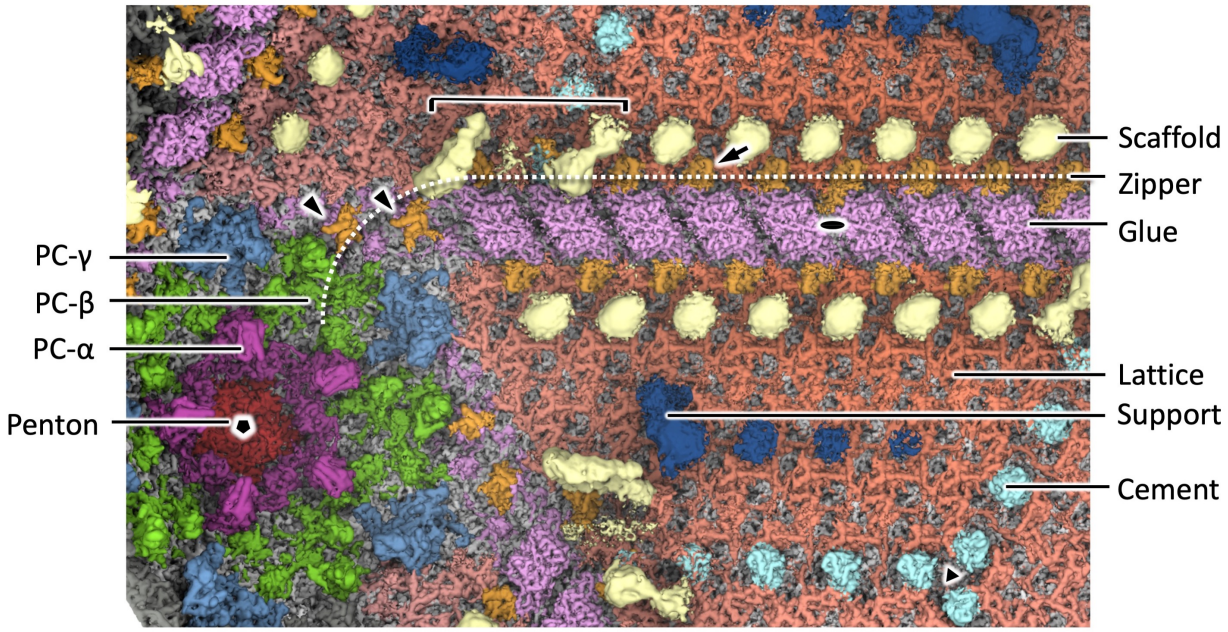
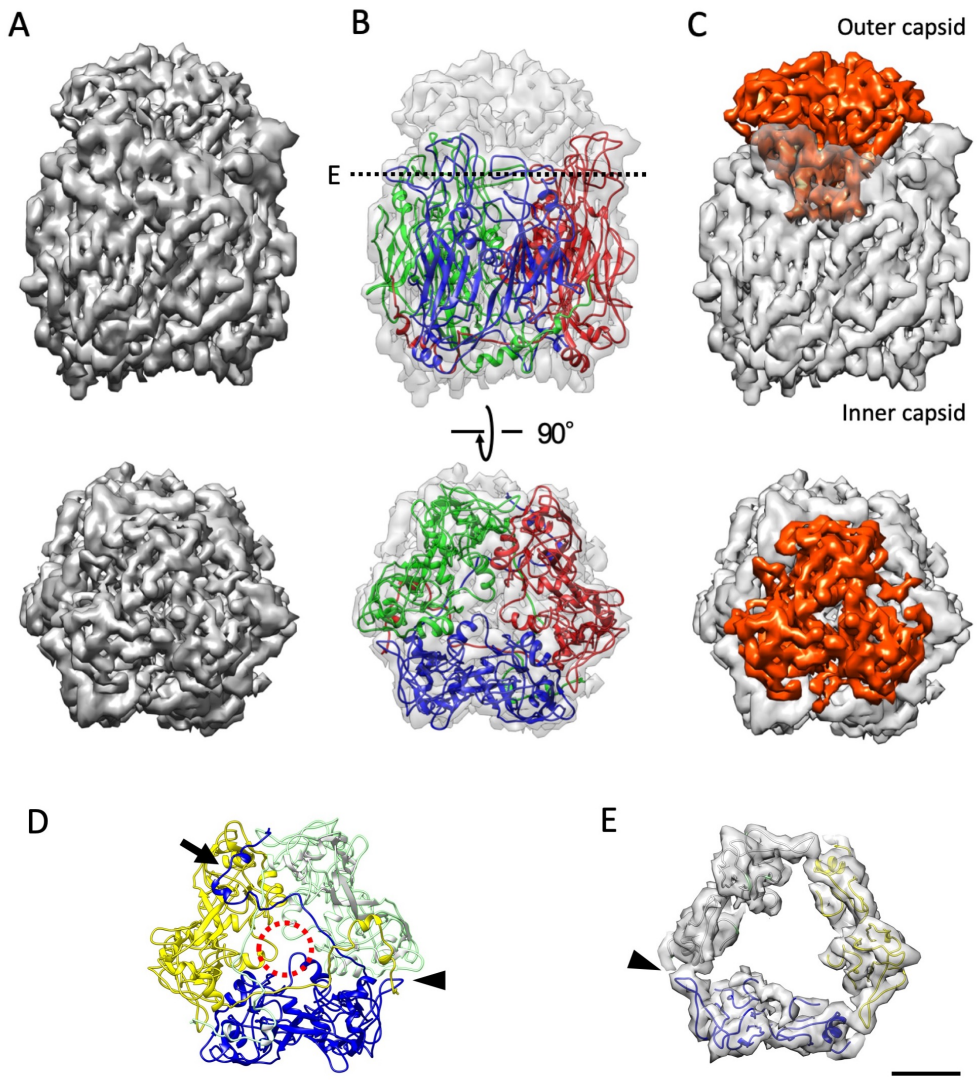
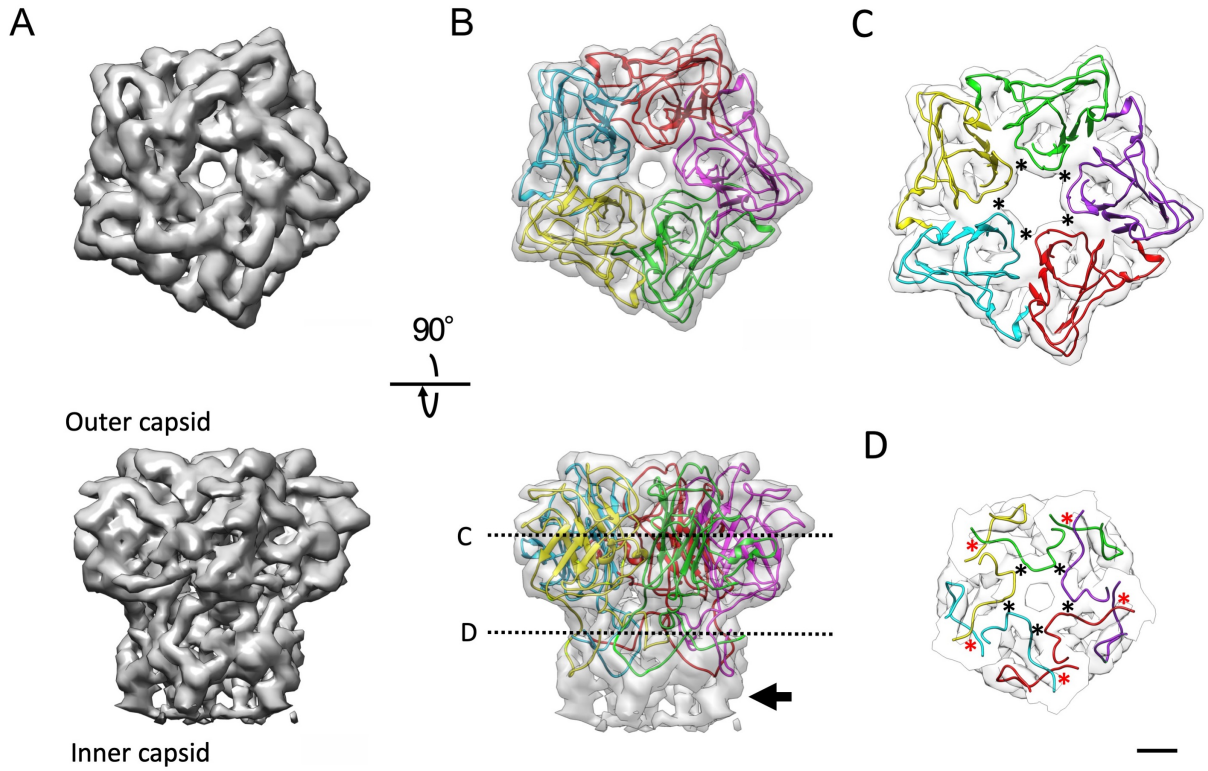


Figure 2. Segmentation of Melbournevirus, focussed on the asymmetric unit encompassing the five, three and twofold axes, and viewed from the inside of the capsid. The map was filtered by local resolution to permit visualisation of the flexible scaffold and support proteins, which are otherwise weak due to uniform b-factor sharpening of the map. Support, Scaffold, Zipper, Glue, Cement, Lattice, PC-α, PC-β, PC-γ, and Penton indicate the names of each mCPs mentioned in those of tokyovirus (Chihara et al., 2021). Fivefold, threefold, and twofold axes indicated by a black pentagon, triangle and double teardrop respectively. White arrow and arrowheads indicate the different orientations of the zipper protein. Bracket shows the "horseshoe" shaped head of the scaffold proteins.



578

579 **Figure 3.** Melbournevirus MCP, with the homology model fitted. A) An MCP trimer extracted
580 from the centre of the three-fold axis. B) Transparent density with the homology model fitted. C)
581 The MCP with the "cap" region, which the annotated MCP sequence used for building the MCP
582 does not fill highlighted. Top and bottom panels show the side and top view of the MCP. D) The
583 FG1 loops (dotted red circle), N-terminal domain (arrow), and HI1 loop (arrowhead) stabilize the
584 MCP trimer. The trimer models are viewed from the inner capsid side. E) The HI1 loop interacts
585 with the adjacent HI2 loop and forms a cup structure with the extended DE1 and FG1 loops to
586 accommodate the cap density (red in panel C). Scale bar equals 2 nm.



587

588 **Figure 4.** A pentamer of penton base proteins of Melbournevirus. A) The extracted penton
589 density. B) With a *de novo* poly-alanine chain model built into the individual symmetrical
590 densities. Top and bottom panels are viewed from the top and the side of the pentamer,
591 respectively. Arrow indicates the density does not fill the polyalanine model. C) The loops from
592 the individual penton base proteins externally stabilize the pentamer. The slice position is
593 indicated in panel B. The interaction points are labeled with asterisks. D) The interactions
594 between N- and C-terminals internally stabilize the pentamer. The slice position is indicated in
595 panel B. The interaction points between N-terminals and between N-terminal and C-terminal are
596 labeled with black and red asterisks, respectively. Scale bar equals 1 nm.

597

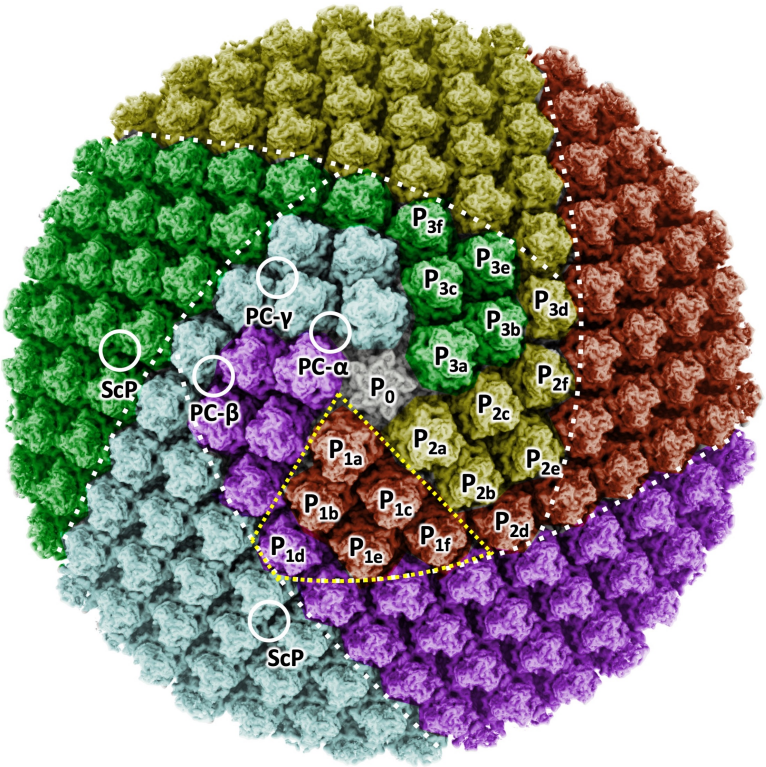


Figure 5. One of the pentasymmetron blocks, with MCPs coloured by orientation. The MCPs of Melbournevirus show the same orientation pattern as CroV and PVCB-1 (Xiao et al, 2017; Xian *et al.* 2020). The while dotted lines show the interfaces of trisymmetron and pentasymmetron. The yellow dotted triangle shows an asymmetric unit of the pentasymmetron. The white circles indicate the locations internal of the MCP layer where the pentasymmetron components are located (PC-α, PC-β and PC-γ respectively) and the "horseshoe" head of the scaffold protein (ScP). Three asymmetric units of the pentasymmetron are labelled $P_{1a\text{-}f}$, $P_{2a\text{-}f}$ and $P_{3a\text{-}f}$, showing that one of the MCPs in the asymmetric unit ($P_{1\text{-}3d}$) is rotated 60° to match that of the adjacent asymmetric unit.

609

610 **References**

611 Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J.,
612 Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a
613 comprehensive Python-based system for macromolecular structure solution. Acta
614 Crystallogr D Biol Crystallogr *66*, 213-221.

615 Aherfi, S., Colson, P., Audoly, G., Nappez, C., Xerri, L., Valensi, A., Million, M., Lepidi, H.,
616 Costello, R., and Raoult, D. (2016). Marseillevirus in lymphoma: a giant in the lymph node.
617 Lancet Infect Dis *16*, e225-e234.

618 Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local
619 alignment search tool. J Mol Biol *215*, 403-410.

620 Aoki, K., Hagiwara, R., Akashi, M., Sasaki, K., Murata, K., Ogata, H., and Takemura, M.
621 (2019). Fifteen Marseilleviruses Newly Isolated From Three Water Samples in Japan Reveal
622 Local Diversity of Marseilleviridae. Front Microbiol *10*, 1152.

623 Benson, S.D., Bamford, J.K., Bamford, D.H., and Burnett, R.M. (2004). Does common
624 architecture reveal a viral lineage spanning all three domains of life? Mol Cell *16*, 673-685.

625 Born, D., Reuter, L., Mersdorf, U., Mueller, M., Fischer, M.G., Meinhart, A., and Reinstein, J.
626 (2018). Capsid protein structure, self-assembly, and processing reveal morphogenesis of the
627 marine virophage mavirus. Proc Natl Acad Sci U S A *115*, 7332-7337.

628 Boyer, M., Yutin, N., Pagnier, I., Barrassi, L., Fournous, G., Espinosa, L., Robert, C., Azza, S.,
629 Sun, S., Rossmann, M.G., *et al.* (2009). Giant Marseillevirus highlights the role of amoebae
630 as a melting pot in emergence of chimeric microorganisms. Proc Natl Acad Sci U S A *106*,
631 21848-21853.

632 Brandes, N., and Linial, M. (2019). Giant Viruses-Big Surprises. Viruses *11*, 404.

633 Chen, S., McMullan, G., Faruqi, A.R., Murshudov, G.N., Short, J.M., Scheres, S.H., and
634 Henderson, R. (2013). High-resolution noise substitution to measure overfitting and validate
635 resolution in 3D structure determination by single particle electron cryomicroscopy.
636 Ultramicroscopy *135*, 24-35.

637 Chihara, A., Burton-Smith, R.N., Kajimura, N., Mitsuoka, K., Okamoto, K., Song, C., and
638 Murata, K. (2021). A novel capsid protein network allows the characteristic inner membrane
639 structure of *Marseilleviridae* giant viruses. bioRxiv, 2021.2002.2003.428533.

640 Colson, P., Pagnier, I., Yoosuf, N., Fournous, G., La Scola, B., and Raoult, D. (2013).
641 "Marseilleviridae", a new family of giant viruses infecting amoebae. Archives of Virology
642 *158*, 915-920.

643 Croll, T.I. (2018). ISOLDE: a physically realistic environment for model building into low-
644 resolution electron-density maps. Acta Crystallogr D Struct Biol *74*, 519-530.

645 De Castro, C., Klose, T., Speciale, I., Lanzetta, R., Molinaro, A., Van Etten, J.L., and Rossmann,
646 M.G. (2018). Structure of the chlorovirus PBCV-1 major capsid glycoprotein determined by
647 combining crystallographic and carbohydrate molecular modeling approaches. Proc Natl
648 Acad Sci U S A *115*, E44-E52.

649 Doutre, G., Philippe, N., Abergel, C., and Claverie, J.M. (2014). Genome analysis of the first
650 Marseilleviridae representative from Australia indicates that most of its genes contribute to
651 virus fitness. J Virol *88*, 14340-14349.

652 Fabre, E., Jeudy, S., Santini, S., Legendre, M., Trauchessec, M., Couté, Y., Claverie, J.-M., and
653 Abergel, C. (2017). Noumeavirus replication relies on a transient remote control of the host
654 nucleus. Nat Commun *8*, 15087.

655 Fang, Q., Zhu, D., Agarkova, I., Adhikari, J., Klose, T., Liu, Y., Chen, Z., Sun, Y., Gross, M.L.,
656 Van Etten, J.L., *et al.* (2019). Near-atomic structure of a giant virus. Nat Commun *10*, 388.
657 Fernandez-Leiro, R., and Scheres, S.H.W. (2017). A pipeline approach to single-particle
658 processing in RELION. Acta Crystallogr D Struct Biol *73*, 496-502.
659 Iyer, L.M., Aravind, L., and Koonin, E.V. (2001). Common origin of four diverse families of
660 large eukaryotic DNA viruses. J Virol *75*, 11720-11734.
661 Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T.L.
662 (2008). NCBI BLAST: a better web interface. Nucleic Acids Res *36*, W5-9.
663 Kayama, Y., Burton-Smith, R.N., Song, C., Terahara, N., Kato, T., and Murata, K. (2021).
664 Below 3 A structure of apoferritin using a multipurpose TEM with a side entry cryoholder.
665 Sci Rep *11*, 8395.
666 Klose, T., Reteno, D.G., Benamar, S., Hollerbach, A., Colson, P., La Scola, B., and Rossmann,
667 M.G. (2016). Structure of faustovirus, a large dsDNA virus. Proc Natl Acad Sci U S A *113*,
668 6206-6211.
669 Kutikhin, A.G., Yuzhalin, A.E., and Brusina, E.B. (2014). Mimiviridae, Marseilleviridae, and
670 virophages as emerging human pathogens causing healthcare-associated infections. GMS
671 hygiene and infection control *9*, Doc16.
672 Liu, S., Luo, Y., Wang, Y., Li, S., Zhao, Z., Bi, Y., Sun, J., Peng, R., Song, H., Zhu, D., *et al.*
673 (2019). Cryo-EM Structure of the African Swine Fever Virus. Cell Host Microbe *26*, 836-
674 843 e833.
675 Okamoto, K., Miyazaki, N., Reddy, H.K.N., Hantke, M.F., Maia, F., Larsson, D.S.D., Abergel,
676 C., Claverie, J.M., Hajdu, J., Murata, K., *et al.* (2018). Cryo-EM structure of a

677 Marseilleviridae virus particle reveals a large internal microassembly. Virology *516*, 239-
678 245.
679 Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and
680 Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and
681 analysis. J Comput Chem *25*, 1605-1612.
682 Pfab, J., Phan, N.M., and Si, D. (2021). DeepTracer for fast de novo cryo-EM protein structure
683 modeling and special studies on CoV-related complexes. Proc Natl Acad Sci U S A *118*.
684 Pintilie, G., and Chiu, W. (2012). Comparison of Segger and other methods for segmentation and
685 rigid-body docking of molecular components in cryo-EM density maps. Biopolymers *97*,
686 742-760.
687 Rohou, A., and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from
688 electron micrographs. J Struct Biol *192*, 216-221.
689 Scheres, S.H.W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure
690 determination. J Struct Biol *180*, 519-530.
691 Subramaniam, K., Behringer, D.C., Bojko, J., Yutin, N., Clark, A.S., Bateman, K.S., van Aerle,
692 R., Bass, D., Kerr, R.C., Koonin, E.V.*, et al.* (2020). A New Family of DNA Viruses
693 Causing Disease in Crustaceans from Diverse Aquatic Biomes. mBio *11*, e02938.
694 Wang, N., Zhao, D., Wang, J., Zhang, Y., Wang, M., Gao, Y., Li, F., Wang, J., Bu, Z., Rao, Z.*,*
695 *et al.* (2019). Architecture of African swine fever virus and implications for viral assembly.
696 Science *366*, 640-644.
697 Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de
698 Beer, T.A.P., Rempfer, C., Bordoli, L.*, et al.* (2018). SWISS-MODEL: homology modelling
699 of protein structures and complexes. Nucleic Acids Res *46*, W296-W303.

700 Xian, Y., Avila, R., Pant, A., Yang, Z., and Xiao, C. (2020). The Role of Tape Measure Protein
701 in Nucleocytoplasmic Large DNA Virus Capsid Assembly. Viral Immunol *34*, 41-48.
702 Xiao, C., Fischer, M.G., Bolotaulo, D.M., Ulloa-Rondeau, N., Avila, G.A., and Suttle, C.A.
703 (2017). Cryo-EM reconstruction of the Cafeteria roenbergensis virus capsid suggests novel
704 assembly pathway for giant viruses. Sci. Rep. 7, 5484.
705 Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017).
706 MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron
707 microscopy. Nat Methods *14*, 331-332.
708 Zhu, D., Wang, X., Fang, Q., Van Etten, J.L., Rossmann, M.G., Rao, Z., and Zhang, X. (2018).
709 Pushing the resolution limit by correcting the Ewald sphere effect in single-particle Cryo-
710 EM reconstructions. Nat Commun *9*, 1552.
711 Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., and Scheres,
712 S.H. (2018). New tools for automated high-resolution cryo-EM structure determination in
713 RELION-3. Elife *7*, e42166.
714 Zivanov, J., Nakane, T., and Scheres, S.H.W. (2020). Estimation of high-order aberrations and
715 anisotropic magnification from cryo-EM data sets in RELION-3.1. IUCrJ *7*, 253-267.
716