A structural roadmap for the formation of the coronavirus nsp3/nsp4 double membrane vesicle pore and its implications for polyprotein processing and replication/transcription

Jason K. Perry,[a]# Samuel Itskanov,[a] John P. Bilello,[b] Eric B. Lansdon[a]

[a]Structural Biology & Chemistry, Gilead Sciences, Foster City, CA, USA

[b]Discovery Virology, Gilead Sciences, Foster City, CA, USA

Running Title: Model of coronavirus DMV pore formation

#Addess correspondence to Jason K. Perry, jason.perry@gilead.com.

Word count: Abstract – 205. Text – 5657.

Abstract

Coronavirus replication is understood to occur within double membrane vesicles (DMVs) that arise during viral infection. Prior work has determined that these DMVs have characteristic pores formed from the non-structural viral proteins nsp3 and nsp4, which facilitate export of newly synthesized viral RNA. Yet how the replication machinery, which is minimally comprised of the non-structural proteins nsp7 to nsp16, is recruited to the DMV remains a mystery. Working from AlphaFold and previously determined structures, we constructed a series of models that link formation of the DMV pore to nsp5 protease processing of the polyprotein and trapping of the cleaved products within the DMV itself. We argue that the initial pore is formed from twelve subunits of nsp3 and six subunits each of the intermediate uncleaved polyproteins pp1a' (nsp4-nsp10) and pp1ab' (nsp4-nsp16). Formation of this initial structure activates the protease function of alternating nsp5 subunits within a close-packed ring, facilitating the initial *trans*-cleavage of the nsp4-nsp5 linkage. Maturation of the pore follows, as does formation of canonical nsp5 dimers, which can process the remainder of the polyproteins. When protease activation occurs subsequent to closure of the DMV, the cleavage products, whose stoichiometry is consistent with the previously proposed nsp15-centered hexameric replication complex, will be trapped inside.

30 **Importance**

31 Coronaviruses, like many other positive-sense single-stranded RNA viruses, rely on intracellular
32 membrane remodeling to create a protected environment for efficient replication to occur. The
33 resulting double membrane vesicles (DMVs) have characteristic pores formed from the non-
34 structural viral proteins nsp3 and nsp4, while enzymes responsible for RNA replication,
35 including the polymerase nsp12, are contained inside. Recent structural work has elucidated
36 the nature of the pores, but how the pores form and how the polymerase and other viral
37 proteins are recruited to the DMV has not yet been addressed. Here we present a step-by-step
38 structural model of how the pores initially form from the uncleaved polyprotein, how protease
39 activity is initiated, and how the viral replication machinery is trapped within the DMV.

40

41 **Introduction**

42 A unifying feature of the cytosolic replication of single-stranded (+) RNA viruses is a remodeling
43 of host intracellular membranes to create local protective environments which facilitate the
44 generation and storage of nascent viral RNA.(1) Chief among these modifications is the creation
45 of replication organelles (ROs) which can be categorized as either invaginated spherules, as in
46 the case of flaviviruses, or double membrane vesicles (DMVs), as in the case of enteroviruses.
47 Coronaviruses are an example of the latter, where remodeling of the endoplasmic reticulum
48 (ER) membrane produces a network of DMVs. Accumulating evidence points to these vesicles
49 as housing both viral RNA and the proteins necessary for replication,(2-6) thus serving the dual
50 purposes of shielding the RNA from triggering an immune response and protecting the RNA and
51 viral proteins from degradation. Research has focused on the viral transmembrane non-
52 structural proteins nsp3, nsp4 and nsp6 as the primary drivers of membrane remodeling and
53 DMV formation.(7) Initial images of these DMVs suggested a closed organelle,(8) however more
54 detailed cryo-electron tomography (cryo-ET) images have captured multiple distinct pores
55 embedded in the membranes of each DMV, providing a necessary mode of egress for newly
56 synthesized RNA. Low resolution cryo-ET reconstructions for both murine hepatitis virus
57 (MHV)(9) and SARS-CoV-2(10) clearly revealed pore structures with distinct six-fold symmetric
58 cytosolic crowns terminating in prominent prong-like features. This cryo-ET work also
59 demonstrated that DMV formation and pore formation go hand in hand, and that nsp3 and
60 nsp4 are the minimal viral components required. A recent 4.2 Å resolution cryo-ET structure
61 further revealed the pore in even greater detail, specifically showing the crown to be composed

62 of twelve subunits of nsp3, anchored to a dodecameric arrangement of nsp4 spanning both
63 inner and outer membranes.(11)

64 Nsp3 and nsp4 can be found on the first open reading frame (ORF1) of the coronavirus genome.
65 Encompassing over 70% of the genome, ORF1 encodes two viral polyproteins, pp1a (nsp1-
66 nsp10) and pp1ab (nsp1-nsp16). They are distinguished by a ribosomal frameshift located
67 between nsp10 and nsp12.(12) Both polyproteins are cleaved by embedded viral proteases in
68 nsp3 and nsp5. The first three non-structural proteins (nsp1-nsp3) are efficiently cleaved by the
69 papain-like proteases of nsp3 (PLpro in SARS-CoV-2),(13, 14) while the remainder of the
70 polyprotein (nsp4-nsp10, referred to here as pp1a', and nsp4-nsp16, referred to as pp1ab') are
71 cleaved in a more regulated manner by the main protease (Mpro) of nsp5.(15, 16) With respect
72 to DMV and pore formation, it has been clearly demonstrated that cleavage between nsp3 and
73 nsp4 is required,(10, 17) but no comparable study has established if cleavage between nsp4
74 and nsp5 is also required. Notably, nsp5, sandwiched between the membrane binding proteins
75 nsp4 and nsp6, is significantly more active when in a dimeric form,(18) but the factors that
76 initiate nsp5 maturation and polyprotein processing remain poorly understood.

77 Once the pp1a' and pp1ab' polyproteins are fully cleaved, the resulting individual proteins,
78 nsp7-nsp16, constitute the viral replication machinery. At a minimum, RNA polymerization is
79 carried out by a complex of (nsp12)(nsp7)$(nsp8)_2$.(19) Additional enzymatic domains of
80 nsp12,(20) nsp13,(21, 22) nsp14(23, 24) and nsp16,(25) aided by the cofactors nsp9 and nsp10,
81 are responsible for other critical viral RNA replication functions including proofreading and
82 mRNA capping. Previously, we proposed these viral proteins assemble around the
83 endonuclease nsp15 to form a hexameric complex of over 60 subunits.(26) Our modeling

84 suggested this complex efficiently coordinates the various functions just described. Implicitly,
85 some or all of these proteins must be recruited to the DMV for replication to occur, although
86 how that might be achieved has been a major unaddressed question.(27)

87 In this work, we address the question of pore formation and cleavage of the polyprotein from a
88 structural modeling approach. Significantly, we link formation of the pore to formation and
89 trapping of the previously proposed hexameric replication complex within the DMV. Much of
90 the work presented here was performed prior to publication of the high-resolution cryo-ET
91 structures by Huang, et al.,(11) but our initial models were largely consistent with these new
92 structures, and we found it relatively straightforward to incorporate those structures into our
93 models. The implications of the proposed model are far ranging, with direct relevance to
94 membrane remodeling, regulation of nsp5 protease activity, and formation of the replication
95 and transcription complex within the closed vesicles.

96 **Results**

97 *SARS-CoV-2 nsp3 monomer*

98 At 1945 residues, nsp3 is the largest protein produced by the SARS-CoV-2 virus. While most of
99 the individual domains have been solved by X-ray crystallography, AlphaFold was used to
100 predict the full-length structure (Figure 1a). The majority of the protein, with some notable
101 exceptions, was predicted with high confidence (Figure S1), as measured by the predicted Local
102 Distance Difference Test (pLDDT),(28) with most domains scoring in the 70-90 range. X-ray
103 structures for Ubl1 (ubiquitin-like domain 1)(29), Mac1 (macrodomain 1)(30), Mac3(31), DPUP
104 (domain preceding Ubl2 and PLpro)(32), Ubl2/PLpro(33), NAB (nucleic acid binding domain)(34)

105 and βSM (betacoronavirus specific marker)(35) were clearly reproduced in the AlphaFold
106 model. Underscoring the success of the method, following completion of this work, a structure
107 became available for the SARS-CoV-2 C-terminal CoV-Y domain(36) that matched very well the
108 AlphaFold prediction. Additional domains with no available structures for any coronavirus were
109 predicted for TM1, 3Ecto, TM2 (a bundle more accurately described as transmembrane domain
110 2 or TMD2) and Y1. The pLDDT score for the region between βSM and Y1, which includes TM1-
111 3Ecto-TMD2, was comparatively lower, at 64.5, but the linker between βSM and TM1 was
112 clearly seen to form two amphipathic helices. The most notable failure in the AlphaFold
113 prediction was for the Mac2 domain, which deviated significantly from X-ray structures for the
114 highly homologous SARS Mac2 domain (71% identity, 85% similarity)(37). This appeared to be
115 unique to the prediction of full-length nsp3 though. Predictions of the isolated domain or the
116 Mac2 and flanking domains, were consistent with the SARS structures, as was prediction of the
117 heterodimer of Mac2 with human Paip1(38). In addition, the highly variable region (HVR)
118 following Ubl1 was completely disordered, as were several loops connecting various domains.
119 In general, variations in the predicted structure conformations, coupled with the pLDDT
120 analysis, reflected significant flexibility between the domains.

121 *SARS-CoV-2 nsp3 dodecamer*

122 Early analysis of the pore structure determined from cryo-ET suggested that the cystosolic
123 crown is composed of a hexamer of nsp3, with the N-terminus residing in the prongs(9, 10).
124 Due to practical limitations in the size of structures that could be predicted with AlphaFold
125 (about 5000 residues), it wasn't possible to generate a model for an entire nsp3 hexamer
126 (11,670 residues) and exploration of truncated constructs had only limited success. The

127 publication of the 4.2 Å cryo-ET structure however provided surprising clarity.(11) The six-fold
128 symmetric crown is instead a dodecamer of nsp3, although much of the protein structure was
129 not resolved. Specifically, the C-terminal Y1/CoV-Y domains form the dodecameric base of the
130 crown and the cytosolic portion of the pore. It is constructed from an inner hexameric ring of Y1
131 (forming the pore), which is surrounded by another hexameric outer Y1 ring to create the base
132 of the crown. The twelve CoV-Y domains interlace above the base, forming the bulk of the
133 crown. The crown is embedded in the outer membrane through TMD2, interacting with nsp4
134 through the 3Ecto domain and TM1. Here we note that the clearest success among the
135 AlphaFold predictions of multimeric nsp3 constructs was that of the inner Y1 hexameric ring,
136 shown in Figure S2 in comparison to the cryo-ET structure.

137 The N-terminal domains preceding TM1 are only partially resolved for six subunits and not
138 resolved at all for the other six, implying a degree of disorder, which may reflect real dynamics
139 of the pore or the lack of other key stabilizing elements such as RNA or additional host or viral
140 proteins. Alternatively, it may simply be an artifact of how the structure was obtained and
141 refined. What is definitively resolved is that a hexameric ring of Ubl2/PLpro rests atop the CoV-
142 Y dodecamer to build up the crown. Huang, et al. also fit each of the six prongs with the Mac2,
143 Mac3 and DPUP domains from one subunit and the NAB domain from a neighboring subunit.
144 However, the resolution of these prongs is considerably weaker (>9 Å) and the fit more
145 speculative by their own metrics. We chose to remodel this part of the structure, taking into
146 account functional constraints, similarities and differences between MHV and SARS-CoV-2, and
147 all available cryo-ET maps, including the prior maps for both MHV and SARS-CoV-2.(9, 10) Our
148 final model adds previously unfit Ubl1 and Mac1 domains in the prongs, positions the NAB

149 domain between PLpro domains in the rim of the crown, and connects the N-terminal domains
150 to TM1 via the βSM domain and two amphipathic helices on the outer membrane cytosolic
151 surface. It is shown in Figure 1b and described in greater detail in the Supplementary
152 Information and Figure S3.

153 *SARS-CoV-2 nsp4 monomer*

154 Our model of nsp4 (Figure 2a) is consistent with previous AlphaFold models with four distinct
155 domains.(39) The N-terminal TM1 is a single α-helix which spans residues 1-27. This is followed
156 by the lumenal domain, a soluble region composed of two distinct lobes connected by a 9-
157 residue linker. The previously designated TM2/TM3/TM4 helices(40) were instead predicted by
158 AlphaFold to be a 5-helix bundle, spanning residues 277-398. For simplicity, we'll refer to this
159 region as transmembrane domain 2 (TMD2), with specific helices α1-α5. Finally, the AlphaFold
160 predicted C-terminal domain (CTD) is another soluble region which reproduces existing X-ray
161 structures(41). The average pLDDT score for the full protein was 83.0 (Figure S4).

162 In none of the predicted structures does TM1 interact with TMD2. The conformation is instead
163 consistent with nsp4 association with the DMV zippered double membrane, where the nsp4 N-
164 terminus is cytosolic, TM1 sits in the outer membrane, the lumenal domain is positioned in the
165 membrane gap, TMD2 sits in the inner membrane, and the nsp4 C-terminal domain (CTD) is
166 exposed to the DMV interior. Thus, the overall topology of the AlphaFold predicted nsp4
167 structure is consistent with that seen in the detailed pore structure.(11)

168 *SARS-CoV-2 nsp4-nsp5-nsp6-nsp7 monomer*

169 To investigate constraints on the nsp5 protease imposed by its position within the polyprotein,
170 we generated AlphaFold models for the uncleaved nsp4-nsp5-nsp6-nsp7 construct (Figure 2b).
171 The pLDDT declined slightly to an average of 77.0 (Figure S5), but X-ray structures of nsp5(42)
172 and nsp7(43) were well reproduced, and nsp6 was predicted to form an 8-helix bundle. The
173 overall organization of the polyprotein showed some flexibility between the individual domains.
174 However, in all predicted structures, the nsp4 TMD2 was aligned with nsp6 as if bound to a
175 common membrane. Unlike the prediction for cleaved nsp4, the nsp4 lumenal domain adopted
176 a conformation between its N- and C-lobes which oriented TM1 such that it associated with the
177 nsp6 helix bundle. This would be more consistent with association with the ER membrane
178 rather than the zippered double membrane, making nsp5, nsp7, the nsp4 N-terminus and the
179 nsp4 CTD cytosolic.(44) The observation that this conformational change is driven by the nsp4
180 lumenal domain is consistent with previous characterizations of this domain by Klatte, et al.(45)

181 A major conclusion from this work is that despite a relatively long flexible loop connecting nsp4
182 and nsp5 (residues 488-510), *cis*-cleavage of these proteins does not appear to be feasible. The
183 cleavage site on the connecting loop (residues 500-501) can only reach the protease active site
184 if the nsp5 helix at residues 510-515 is unwound, compromising the integrity of the protease
185 structure (Figure S6). Similarly, *cis*-cleavage of the nsp5 and nsp6 proteins at residues 807-808
186 does not appear to be possible under any circumstances. Thus, in both cases, cleavage must
187 proceed via interaction with the protease of a second polyprotein or an already cleaved nsp5.

188 *SARS-CoV-2 nsp4 dodecamer*

189 As with nsp3, our initial assumption was that the nsp4 component of the pore was hexameric.
190 Based on our predicted structures for the nsp4 monomer and previous work suggesting nsp3
191 interacts with the nsp4 lumenal domain,(45, 46) we reasoned TM1 must sit in the outer
192 membrane, positioning the lumenal domain in the membrane gap. This suggested the portion
193 of the pore embedded in the inner membrane would be formed from a hexamer of the nsp4
194 TMD2 domains, with the nsp4 CTD exposed to the DMV interior. Thus, we worked with a
195 truncated construct which spanned residues 256-500 of nsp4 (TMD2 and CTD). AlphaFold
196 indeed predicted several plausible hexameric structures with pore-like properties, but none
197 was entirely consistent with the available cryo-ET maps. However, upon close examination of
198 the MHV map, we identified density within the membrane gap which was similar in size to the
199 nsp4 lumen domain, but which indicated nsp4 was in fact dodecameric, not hexameric.
200 Unfortunately, AlphaFold predictions of dodecameric nsp4 (residues 256-500) did not produce
201 structures with the expected C6 symmetry. The detailed pore structure from Huang, et al.(11)
202 confirmed that the nsp4 component is indeed a dodecamer, characterized by a central
203 hexameric ring surrounded by an outer hexameric ring. While we didn't initially recognize it to
204 be the case, AlphaFold in fact accurately predicted the central component of the hexameric
205 inner pore, as shown in Figure S7 in comparison to the cryo-ET structure.

206 *SARS-CoV-2 nsp4-nsp5-nsp6 dodecamer*

207 Considering the possibility of pore formation prior to nsp5 cleavage, our preliminary
208 exploration with AlphaFold also considered hexamers of uncleaved nsp4-nsp5-nsp6 (truncated
209 to residues 256-1042). AlphaFold generated a particularly compelling model shown in Figure 3a,
210 which consisted of an inner ring of nsp4 TMD2 surrounded by an outer ring of nsp6. The two

211 transmembrane domains are linked via an inner, solvent exposed ring formed from the nsp4
212 CTD, which is surrounded by a larger ring formed from nsp5. Consistent with the emerging
213 topology of the nsp3/nsp4 pore, both nsp5 and the nsp4 CTD would be exposed to the interior
214 of the DMV in this model.

215 We found it was straightforward to envision similar rings of varying sizes, but a dodecameric
216 arrangement of nsp4-nsp5-nsp6 was beyond the practical limits of AlphaFold. To generate such
217 a model, we instead resorted to a simple geometric expansion of the nsp4-nsp5-nsp6 hexamer.
218 This was executed by defining the center of mass of each subunit within the hexamer,
219 measuring the distance between adjacent subunits, and translating and rotating duplicate sets
220 of these subunits to form a dodecameric ring in which this distance was maintained. Structural
221 refinement from this starting point led to the fully optimized construct shown in Figure 3b.

222 In comparison to the earlier MHV and SARS-CoV-2 cryo-ET maps, this dodecameric
223 arrangement appeared to be consistent with the overall diameter of the inner membrane pore
224 and provided the first hint as to how these pores may form. Once the detailed pore structure of
225 Huang, et al(11) became available, the arrangement of the final fully cleaved pore was clear,
226 and our attention turned to whether the uncleaved model was topologically consistent with
227 this structure. A clear requirement was that the inner nsp4 hexamer in the final pore structure
228 requires cleavage from nsp5; however, the outer nsp4 hexamer does not. Interestingly, in the
229 nsp4-nsp5-nsp6 hexamers, we saw a domain swapping occur in which TMD2 adopts an open
230 conformation with α1 jutting out at a significant angle (~60°) to interact with a neighboring
231 TMD2 bundle of α2-α5. This is also seen in both the inner and outer hexamers of the structure
232 solved by Huang, et al. We recognized however that there was some ambiguity as to how this

233 may manifest itself in the nsp4-nsp5-nsp6 dodecamer. We concluded that if all subunits
234 undergo the domain swap, then the final pore could only form following cleavage of all twelve
235 nsp4-nsp5 linkages. However, in an arrangement in which only alternating subunits adopt the
236 open conformation, upon cleavage of the nsp4-nsp5 linkages from those same six subunits, the
237 structure would have a simple pathway to adopt the mature pore conformation (Figure 3b).
238 This model was merged with the nsp4 lumenal domains of the Huang, et al. structure and our
239 rebuilt model of the nsp3 dodecamer. The nsp4 TM1 helices, missing from the cryo-ET
240 structure, were also added in parallel to the nsp3 TM1 helices, consistent with discernible, but
241 previously unmodelled, density in the maps.

242 *SARS-CoV-2 uncleaved pp1a'/pp1ab' model*

243 Having created a model of dodecameric nsp4-nsp6, we considered modeling the remainder of
244 the uncleaved polyprotein. This extended model could be derived from either pp1a' (nsp4-
245 nsp10) or pp1ab' (nsp4-nsp16) or a combination of both. We considered that certain known
246 protein-protein interactions may be preserved in this situation, creating a degree of
247 preorganization that would facilitate efficient formation of the replication complex upon
248 polyprotein cleavage. From cryo-EM and X-ray structures, interactions between nsp7 and nsp8,
249 nsp7 and nsp12, nsp8 and nsp12, nsp8 and nsp13, nsp9 and nsp12, nsp10 and nsp14, and
250 nsp10 and nsp16 are well characterized(47-50), as is nsp15 hexamerization(51).

251 Working from a dimer of nsp4-nsp5-nsp6 extracted from the optimized dodecameric ring, it
252 appeared that extending both subunits to the full pp1ab' polyprotein would be too bulky. In
253 contrast, extending both to the smaller pp1a' polyprotein would be well tolerated, but offer

254 little in additional protein-protein interactions. However, a heterodimer formed from pp1a' and
255 pp1ab' offered the maximum opportunity to incorporate the known protein-protein
256 interactions. Following exploration with AlphaFold and Piper, we identified an arrangement
257 (Figure 3c) in which multiple protein-protein interactions are preserved: pp1a'-nsp8/pp1ab'-
258 nsp7; pp1a'-nsp8/pp1ab'-nsp13; pp1a'-nsp10/pp1ab'-nsp14; pp1ab'-nsp8/pp1ab'-nsp12; and
259 pp1ab'-nsp10/pp1ab'-nsp16. In addition, nsp14 adopts a conformation and position relative to
260 nsp12 consistent with our previous model for the full hexameric replication complex(26).
261 Significantly, cleavage sites were all accessible on the surface of the polyprotein heterodimer.
262 This dimer was then refined and incorporated into the dodecameric model where it went
263 through a final refinement (Figure 3d).

264 We note that constraints imposed by the heterodimer provide a second topological argument
265 for the alternating conformations of the TMD2 domains in the dodecameric ring. In considering
266 that the ring may be derived from preformed heterodimers which originate on a single ER
267 membrane, the dimer would need to introduce a twist at the nsp4 TMD2-lumen domain linkage
268 when adopting the double membrane binding conformation found in the ring. This twist forces
269 one of the α1 helices to open up and cross over the other nsp4 subunit (see Figure S8). Overall,
270 the model provides a consistent evolution from heterodimer formation on the ER membrane,
271 to adopting a double membrane spanning conformation upon association with nsp3, to
272 formation of the mature pore upon nsp4 cleavage.

273 *SARS-CoV-2 polyprotein cleavage*

274 The close-packed arrangement of nsp5 in the polyprotein dodecameric ring establishes a
275 favorable environment for the protease to make the initial nsp4-nsp5 cleavage. A key
276 observation we made was that the nsp4-nsp5 linker (residues 488-510) was long enough that it
277 could conceivably connect nsp4 to either of two nsp5 subunits in the ring (at a distance of 35 Å
278 or 52 Å). As originally constructed from the AlphaFold hexamer, the connection was to the
279 nearer nsp5-nsp6 subunit, comparable to how the proteins are arranged in the monomeric
280 polyprotein. However, connecting nsp4 instead to the neighboring nsp5-nsp6 subunit allowed
281 the nsp4-nsp5 cleavage site to pass over the proximal nsp5 protease active site. A positional
282 shift of the nsp5 proteins would still be required for cleavage to occur, and this was
283 investigated for both subunits. We found that simply by flipping alternating nsp5 subunits while
284 retaining the positions of everything else, we could identify a conformation that was capable of
285 nsp4-nsp5 cleavage, consistent with available crystal structures of bound substrate.(52) As
286 depicted in Figures 4a and 4b, the flipped nsp5 subunit functions as the substrate, while the
287 nsp5 subunit which retained its original position functions as the protease. Given the
288 constraints imposed by the full $(pp1a'/pp1ab')_6$ model, a repositioning of nsp5 associated with
289 pp1ab' appears more likely than that of pp1a' suggesting the larger polyprotein is the initial
290 substrate for cleavage.

291 The protease activity of nsp5 is known to be significantly enhanced when the protein is in its
292 dimer form(53). Formation of such a dimer does not appear to be possible prior to nsp5's
293 cleavage from either nsp4 or nsp6. However, a striking observation from these AlphaFold
294 derived models is that the C-terminal domain of nsp4 forms an interaction with nsp5 which
295 appears to mimic important aspects of nsp5 dimerization (Figure 4c). Specifically, a conserved

296 nsp4 basic residue (K452) forms a salt bridge with a conserved nsp5 glutamic acid (E666),
297 replicating the interaction seen in the dimer between the same glutamic acid and the N-
298 terminus of the paired subunit. This interaction has previously been implicated in the enhanced
299 dimer activity, with the conclusion that it stabilizes the P1 pocket(54). The similar interaction
300 seen here suggests formation of the pore prior to polyprotein cleavage both activates the nsp5
301 protease and positions the nsp4-nsp5 substrate to carry out the first cleavage event.

302 This proposed arrangement of alternating positions of nsp5 in the dodecameric ring would lead
303 to the cleavage of half of the nsp4-nsp5 linkages. Once this occurs, the reduced constraints
304 allow the nsp3/nsp4 pore to adopt a conformation consistent with that of the cryo-ET
305 structure. That is, the six cleaved nsp4 proteins rotate inward to form the lumenal hexameric
306 pore. With this reconfiguration, the six remaining uncleaved nsp4 subunits can also adopt the
307 conformations seen in the cryo-ET structure. At the same time, the six partially cleaved nsp5
308 subunits can now form canonical nsp5 dimers with the six uncleaved nsp5 subunits, as shown in
309 Figures 4d and 4e. From this configuration, cleavage of additional linkages could take several
310 pathways, with nsp5-nsp6, nsp6-nsp7, and nsp8-nsp9 linkages all within range (Figure S9).

311 **Discussion**

312 It has been established that cleavage between nsp3 and nsp4 is a prerequisite for pore
313 formation and membrane remodeling(10, 17). However, no such requirement has been shown
314 with respect to cleavage of nsp4 from the remaining intermediate polyproteins pp1a' and
315 pp1ab'. Cleavage of nsp1, nsp2 and nsp3 from the polyprotein occurs very quickly, facilitated by
316 the nsp3 PL proteases(14). Cleavage of the remainder of the polyprotein by nsp5 remains more

317 enigmatic, but appears to be a slower process, possibly requiring dimer formation or some
318 other regulatory event to initiate(55). Our own analysis of the nsp4-nsp5-nsp6-nsp7 monomer
319 made clear that *cis*-cleavage of the nsp4-nsp5 and nsp5-nsp6 linkages is highly unlikely,
320 implying cleavage follows some interaction between polyproteins. In the work presented here,
321 we generated a series of structural models outlining a plausible path for DMV and pore
322 formation stemming from the interaction of twelve nsp3 subunits with six pp1a' and six pp1ab'
323 uncleaved polyproteins. We argue that it is the formation of an immature pore from these
324 components that initiates the protease activity of nsp5.

325 We made several key observations from this model. First, nsp4 can adopt conformations
326 consistent with binding either to a single membrane or to a double membrane. This is
327 facilitated by a conformational change in the lumen domain and is consistent with prior
328 conclusions made by Klatte, et al.(45) Empirically, it is the interaction between nsp3 and nsp4
329 that drives this conformational change, as nsp4 by itself does not induce membrane
330 remodeling.(46)

331 Second, the uncleaved polyproteins pp1a' and pp1ab' can form a favorable heterodimer in
332 which known protein-protein interactions are established. While pp1a' and pp1ab' homodimers
333 could also be imagined, neither would have the advantages of the complementarity seen in the
334 heterodimer. This dimer is also capable of binding to either a single membrane or a double
335 membrane. When bound to the single membrane of the ER, the nsp4 N-terminus and CTD,
336 nsp5, nsp7-nsp10 of pp1a' and nsp7-nsp16 of pp1ab' are cytosolic, while only the nsp4 lumen
337 domain is in the ER interior. When bound to the double membrane of a DMV, the nsp4 N-
338 terminus is the only element remaining exposed to the cytosol. With the nsp4 lumen domain

339 sitting in the membrane gap, the bulk of the polyprotein dimer then resides in the DMV
340 interior. As with isolated nsp4, this conformation change would necessarily be driven by
341 association with nsp3.

342 These heterodimers, when in the double membrane spanning conformation, are then capable
343 of forming a ring comprised of twelve alternating pp1a' and ppa1b' polyproteins. This ring has a
344 diameter comparable to that seen for the nsp4 dodecamer in the recent nsp3/nsp4 pore
345 structure reported by Huang, et al.(11) and can associate with the nsp3 dodecamer in the same
346 manner. Thus, we propose that this $(nsp3)_{12}(pp1a')_6(pp1ab')_6$ complex represents an immature
347 form of the pore. The polyprotein dodecameric ring leads to close packing of nsp5 and the nsp4
348 CTD, providing the necessary scaffolding to facilitate *trans*-cleavage. We identified a
349 conformation within this framework in which the 23-residue nsp4-nsp5 linker of alternating
350 subunits engages the proximal nsp5 protease active site. Critically, we also noticed that a basic
351 residue from the nsp4 C-terminal domain (K452) forms a salt bridge with the conserved nsp5
352 glutamate residue (E666) located in the protease active site P1 pocket. This interaction mimics
353 the situation in the nsp5 dimer, in which the same glutamate residue interacts with the N-
354 terminal amine of the other subunit. It is thought to be critical to stabilization of the protease
355 active site, explaining the significant enhancement of protease activity associated with
356 dimerization(54). Following this initial nsp4-nsp5 cleavage event, the six cleaved nsp4 proteins
357 are free to adopt the conformation seen in the mature pore structure. The reduced structural
358 constraints also allow formation of canonical nsp5 dimers, having proximity to additional
359 cleavage sites within the polyprotein.

360 With this model, we found that pore formation prior to nsp5 polyprotein processing is a
361 reasonable possibility and may in fact be the regulatory step needed to ensure proper
362 localization of the cleaved proteins that make up the replication machinery. A notable
363 observation of the full model presented here is that the stoichiometry is consistent with our
364 previous model for the hexameric replication complex(26). That model included 60+ subunits
365 centered around nsp15. Specifically, the model includes six subunits of nsp7, nsp12, nsp13,
366 nsp14, nsp15, and nsp16, and twelve subunits of nsp8 and nsp10, with up to six transiently
367 associated nsp9 subunits. This complex is most efficiently derived from six pp1a and six pp1ab
368 polyproteins, meaning one pore structure generates one full replication complex. When viewed
369 together, the two models imply a mechanism not only for DMV and pore formation, but for
370 recruitment of the replicase to the DMV.

371 As depicted in Figure 5, the following picture emerges from these models. Upon cleavage by the
372 PL proteases, the nsp3, pp1a' and pp1ab' products are associated with the ER membrane
373 (Figure 5a). The nsp3 3Ecto and nsp4 lumenal domains are in the ER lumen, with the bulk of the
374 proteins exposed to the cytosol. We suggest here that nsp3 likely associates with pp1a' and
375 pp1ab' first, before forming the tetrameric unit $(nsp3)_2$(pp1a')(pp1ab') and inducing membrane
376 zippering (Figure 5b). The exact sequence of events is unclear, but this pathway seems likely
377 given the immediate proximity between nsp3 and either pp1a' or pp1ab' following PLpro
378 cleavage. Once membrane zippering occurs, the nsp3 3Ecto and the nsp4 lumen domains are
379 positioned in the gap between the two membranes, and the remainder of nsp3 is on one side
380 of the double membrane and nsp5, nsp7 and the rest of the polyprotein are on the other side.
381 Aggregation of these tetrameric building blocks leads to the immature pore and may facilitate

382 the observed increased curvature of the zippered membrane (Figure 5c). Formation of
383 additional pores ultimately will yield a sealed, spherical DMV.

384 The picture of membrane remodeling presented here raises the question of the relative timing
385 of nsp5 polyprotein processing and DMV formation. We see no justification for a strict
386 requirement that a DMV be fully formed for cleavage to begin. Should cleavage occur prior to
387 DMV formation, the nsp products would be released into the cytosol. However, should cleavage
388 follow closure of the DMV, the resulting products would be trapped inside (Figure 5d). This is a
389 scenario that would likely follow formation of the final pore of the DMV, suggesting each DMV
390 contains at minimum one hexameric replication complex. Interestingly, this process may
391 ultimately become self-limiting, as free nsp5 dimers build up in the cytosol, cleavage of new
392 polyproteins at late stages may occur without the requirement of new pore formation.

393 The implications of this model require careful consideration. Within the protected environment
394 of the DMV, the proposed stoichiometry, pre-organization and high protein concentration (we
395 estimate at least 1-2 μM based on a DMV diameter of 250-350 nm(8, 9, 56)) would be ideal for
396 formation of a long-lived full hexameric replicase complex. As previously described, this
397 complex should ensure that replication, proofreading, capping and dsRNA unwinding are
398 coordinated and efficient processes.(26) Confirmation of the exact contents of DMVs has been
399 notoriously difficult, however. Multiple studies have shown that the DMVs contain RNA,
400 specifically identifying both dsRNA and newly synthesized RNA.(3, 8, 56-58) Identification of
401 viral protein components of the replicase within DMVs has been less conclusive, although
402 increasing evidence supports their association with DMVs.(57, 58)

403 In contrast, the less controlled environment outside of DMVs may mean the full complex does
404 not form or that complexes are transient. The minimally known replication competent complex
405 is formed from nsp12, nsp7 and two subunits of nsp8.(59, 60) Association with nsp9(48, 61) and
406 two subunits of nsp13(21, 47) have been documented as well, but characterization of other
407 viral protein interactions has proven more elusive. Indeed, the proposed dual role of nsp15
408 suggests a full hexameric complex forms within the DMV, but likely does not in the cytosol. As
409 the center of the hexameric replication complex, it is proposed to function primarily as
410 scaffolding, with its endonuclease activity highly curtailed.(26) Thus, with nsp15 present within
411 a DMV, formation of the complex would be a necessary regulatory mechanism to prevent
412 degradation of RNA. However, as a free nsp15 hexamer outside of the DMVs, its endonuclease
413 activity serves as an effective defense against the immune response by degrading cytosolic viral
414 RNA.(51, 62)

415 In the absence of full complex formation, some functions, such as proofreading and mRNA
416 capping, might be compromised or lost entirely. Yet plenty of evidence remains that some
417 replication is occurring outside of the DMVs. At minimum, there is a temporal argument for
418 this, in that replication necessarily precedes DMV formation.(63) However, other lines of
419 evidence are consistent with a two-compartment model persisting throughout infection.
420 Notably, the groundbreaking work of Sawicki and Sawicki(64) established that coronavirus RNA
421 replication is dependent on continuous ORF1 translation and proteolytic processing. Further
422 work on several temperature sensitive (TS) mutations showed that synthesis of negative-sense
423 RNA was halted at higher temperatures but positive-sense RNA was not.(65) Donaldson, et
424 al.(65) showed that this was a result of inhibited nsp5 protease activity at the higher

425 temperatures. More recently, Schmidt, et al.(66) demonstrated that negative strand synthesis,
426 was dependent on UMPylated-nsp9 priming, facilitated by the host factor SND1. This did not
427 appear to be the case for positive strand synthesis. With no proposed mechanism for
428 recruitment of additional proteins to an already formed DMV, the dependence of negative-
429 sense RNA synthesis on host factors and the continuous generation of viral replicase proteins
430 (specifically nsp9) suggest transcription of the minus strand occurs in the cytosol. In contrast,
431 this model would be consistent with positive strand synthesis occurring in the cytosol at early
432 stages but predominantly occurring within the protected environment of DMVs at later stages.

433 In summary, in asking the question whether the nsp3/nsp4 DMV pore forms prior to nsp5
434 cleavage of the polyprotein, we have developed a framework for understanding the formation
435 of DMVs and the replication complex in coronavirus infected cells. From a close examination of
436 structural constraints within monomeric and multimeric forms of the polyprotein, aided in large
437 part by AlphaFold predictions, we conclude that formation of an immature pore from twelve
438 nsp3 subunits, and six subunits each of the intermediate polyproteins pp1a' and pp1ab'
439 initiates the protease activity of nsp5. Following cleavage of half of the nsp4-nsp5 linkages, the
440 mature nsp3/nsp4 pore can form and protease cleavage of the remainder of the polyproteins
441 will release the components necessary to form the replication complex into the DMV. While the
442 detailed mechanism is specific to coronaviruses, we expect similar processes are taking place
443 with other single-stranded (+) RNA viral families.

444

445 **Materials and Methods**

446 AlphaFold2.1(67) (with a template cutoff date of 12/01/22) was used to predict the structures
447 of individual monomers of nsp3, nsp4, and uncleaved nsp4-nsp5-nsp6-nsp7 from SARS-CoV-2,
448 as well as hexameric complexes of the C-terminal Y1/CoV-Y domain of nsp3, hexameric
449 complexes of nsp4, and hexameric complexes of uncleaved nsp4-nsp5-nsp6. In all cases,
450 AMBER relaxation was turned off and structures were instead refined with software from the
451 Schrödinger Suite, including Prime, Macromodel and Desmond(68). A typical workflow involved
452 construction of an initial model either from AlphaFold directly or via manual manipulation of
453 individual subunits, followed by sidechain prediction in Prime. This was followed by a stepwise
454 minimization with Prime and/or Macromodel, in which residues at the protein-protein interface
455 were optimized first, building up to a minimization of the full structure. In cases where large
456 movements from the starting point occurred, a second round of optimization would be carried
457 out starting from a structure in which the original protein subunits were re-aligned to their new
458 positions. Stability was tested at several stages of model development with molecular dynamics
459 (MD) using Desmond, typically run for 100ns with suitable constraints. Macromodel and
460 Desmond were also used to predict conformations of some long loops. Construction of the
461 pp1a'/pp1ab' dimer involved preservation of known interactions as much as possible, using
462 AlphaFold and Piper(69) protein-protein docking to explore possible arrangements of various
463 subunits.

464 The SARS-CoV-2 sequences used are from the original Wuhan strain (NCBI accession number
465 QHD43415.1(70)). Nsp3 spans residues 819-2763 of the polyprotein, nsp4 spans 2764-3263,
466 pp1a' (nsp4-nsp10) spans 2764-4392, and pp1ab' (nsp4-nsp16) spans 2764-7096.

467 Models were guided by the cryo-ET maps deposited in the Electron Microscopy Data Bank with
468 accession codes EMD-11514 (MHV(9)), EMD-15963 (SARS-CoV-2(10)), and EMD-39107 and
469 EMD-39109 (SARS-CoV-2(11)). Ultimately, the models were merged with the structures of
470 Huang, et al. (PDB 8YAX and 8YB5)(11) and further refined using the approaches described
471 above. While the final pore models have approximate C6 rotational symmetry, strict symmetry
472 constraints were not imposed.

473 **References**

474 1. Wolff G, Barcena M. 2021. Multiscale Electron Microscopy for the Study of Viral Replication
475 Organelles. Viruses 13.
476 2. Wolff G, Melia CE, Snijder EJ, Barcena M. 2020. Double-Membrane Vesicles as Platforms for Viral
477 Replication. Trends Microbiol 28:1022-1033.
478 3. Snijder EJ, Limpens R, de Wilde AH, de Jong AWM, Zevenhoven-Dobbe JC, Maier HJ, Faas F,
479 Koster AJ, Barcena M. 2020. A unifying structural and functional model of the coronavirus
480 replication organelle: Tracking down RNA synthesis. PLoS Biol 18:e3000715.
481 4. Sergio MC, Ricciardi S, Guarino AM, Giaquinto L, De Matteis MA. 2024. Membrane remodeling
482 and trafficking piloted by SARS-CoV-2. Trends Cell Biol doi:10.1016/j.tcb.2023.12.006.
483 5. Roingeard P, Eymieux S, Burlaud-Gaillard J, Hourioux C, Patient R, Blanchard E. 2022. The
484 double-membrane vesicle (DMV): a virus-induced organelle dedicated to the replication of
485 SARS-CoV-2 and other positive-sense single-stranded RNA viruses. Cell Mol Life Sci 79:425.
486 6. Wang J, Han M, Roy AR, Wang H, Mockl L, Zeng L, Moerner WE, Qi LS. 2022. Multi-color super-
487 resolution imaging to study human coronavirus RNA during cellular infection. Cell Rep Methods
488 2:100170.
489 7. Angelini MM, Akhlaghpour M, Neuman BW, Buchmeier MJ. 2013. Severe acute respiratory
490 syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles.
491 mBio 4.
492 8. Knoops K, Kikkert M, Worm SH, Zevenhoven-Dobbe JC, van der Meer Y, Koster AJ, Mommaas
493 AM, Snijder EJ. 2008. SARS-coronavirus replication is supported by a reticulovesicular network of
494 modified endoplasmic reticulum. PLoS Biol 6:e226.
495 9. Wolff G, Limpens R, Zevenhoven-Dobbe JC, Laugks U, Zheng S, de Jong AWM, Koning RI, Agard
496 DA, Grunewald K, Koster AJ, Snijder EJ, Barcena M. 2020. A molecular pore spans the double
497 membrane of the coronavirus replication organelle. Science 369:1395-1398.
498 10. Zimmermann L, Zhao X, Makroczyova J, Wachsmuth-Melm M, Prasad V, Hensel Z,
499 Bartenschlager R, Chlanda P. 2023. SARS-CoV-2 nsp3 and nsp4 are minimal constituents of a
500 pore spanning replication organelle. Nat Commun 14:7894.
501 11. Huang Y, Wang T, Zhong L, Zhang W, Zhang Y, Yu X, Yuan S, Ni T. 2024. Molecular architecture of
502 coronavirus double-membrane vesicle pore complex. Nature 633:224-231.
503 12. Sawicki SG, Sawicki DL, Siddell SG. 2007. A contemporary view of coronavirus transcription. J
504 Virol 81:20-9.
505 13. Harcourt BH, Jukneliene D, Kanjanahaluethai A, Bechill J, Severson KM, Smith CM, Rota PA,
506 Baker SC. 2004. Identification of severe acute respiratory syndrome coronavirus replicase
507 products and characterization of papain-like protease activity. J Virol 78:13600-12.
508 14. Graham RL, Denison MR. 2006. Replication of murine hepatitis virus is regulated by papain-like
509 proteinase 1 processing of nonstructural proteins 1, 2, and 3. J Virol 80:11610-20.
510 15. Li C, Qi Y, Teng X, Yang Z, Wei P, Zhang C, Tan L, Zhou L, Liu Y, Lai L. 2010. Maturation
511 mechanism of severe acute respiratory syndrome (SARS) coronavirus 3C-like proteinase. J Biol
512 Chem 285:28134-40.
513 16. Chen S, Jonas F, Shen C, Hilgenfeld R. 2010. Liberation of SARS-CoV main protease from the viral
514 polyprotein: N-terminal autocleavage does not depend on the mature dimerization mode.
515 Protein Cell 1:59-74.
516 17. Oudshoorn D, Rijs K, Limpens R, Groen K, Koster AJ, Snijder EJ, Kikkert M, Barcena M. 2017.
517 Expression and Cleavage of Middle East Respiratory Syndrome Coronavirus nsp3-4 Polyprotein
518 Induce the Formation of Double-Membrane Vesicles That Mimic Those Associated with
519 Coronaviral RNA Replication. mBio 8.

520 18. Roe MK, Junod NA, Young AR, Beachboard DC, Stobart CC. 2021. Targeting novel structural and
521 functional features of coronavirus protease nsp5 (3CL(pro), M(pro)) in the age of COVID-19. J
522 Gen Virol 102.
523 19. Hillen HS, Kokic G, Farnung L, Dienemann C, Tegunov D, Cramer P. 2020. Structure of replicating
524 SARS-CoV-2 polymerase. Nature 584:154-156.
525 20. Park GJ, Osinski A, Hernandez G, Eitson JL, Majumdar A, Tonelli M, Henzler-Wildman K,
526 Pawlowski K, Chen Z, Li Y, Schoggins JW, Tagliabracci VS. 2022. The mechanism of RNA capping
527 by SARS-CoV-2. Nature 609:793-800.
528 21. Chen J, Wang Q, Malone B, Llewellyn E, Pechersky Y, Maruthi K, Eng ET, Perry JK, Campbell EA,
529 Shaw DE, Darst SA. 2021. Ensemble cryo-electron microscopy reveals conformational states of
530 the SARS-CoV-2 helicase replication-transcription complex. To be published.
531 22. Malone B, Chen J, Wang Q, Llewellyn E, Choi YJ, Olinares PDB, Cao X, Hernandez C, Eng ET, Chait
532 BT, Shaw DE, Landick R, Darst SA, Campbell EA. 2021. Structural basis for backtracking by the
533 SARS-CoV-2 replication-transcription complex. Proc Natl Acad Sci U S A 118.
534 23. Bouvet M, Imbert I, Subissi L, Gluais L, Canard B, Decroly E. 2012. RNA 3'-end mismatch excision
535 by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14
536 exoribonuclease complex. Proc Natl Acad Sci U S A 109:9372-7.
537 24. Ogando NS, El Kazzi P, Zevenhoven-Dobbe JC, Bontes BW, Decombe A, Posthuma CC, Thiel V,
538 Canard B, Ferron F, Decroly E, Snijder EJ. 2021. Structure-function analysis of the nsp14 N7-
539 guanine methyltransferase reveals an essential role in Betacoronavirus replication. Proc Natl
540 Acad Sci U S A 118.
541 25. Decroly E, Debarnot C, Ferron F, Bouvet M, Coutard B, Imbert I, Gluais L, Papageorgiou N, Sharff
542 A, Bricogne G, Ortiz-Lombardia M, Lescar J, Canard B. 2011. Crystal structure and functional
543 analysis of the SARS-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. PLoS
544 Pathog 7:e1002059.
545 26. Perry JK, Appleby TC, Bilello JP, Feng JY, Schmitz U, Campbell EA. 2021. An atomistic model of
546 the coronavirus replication-transcription complex as a hexamer assembled around nsp15. J Biol
547 Chem 297:101218.
548 27. Malone B, Urakova N, Snijder EJ, Campbell EA. 2022. Structures and functions of coronavirus
549 replication-transcription complexes and their relevance for SARS-CoV-2 drug design. Nat Rev
550 Mol Cell Biol 23:21-39.
551 28. Mariani V, Biasini M, Barbato A, Schwede T. 2013. lDDT: a local superposition-free score for
552 comparing protein structures and models using distance difference tests. Bioinformatics
553 29:2722-8.
554 29. Stogios,P.J., et al., PDB 7KAG, unpublished, 2020.
555 30. Brosey CA, Houl JH, Katsonis P, Balapiti-Modarage LPF, Bommagani S, Arvai A, Moiani D, Bacolla
556 A, Link T, Warden LS, Lichtarge O, Jones DE, Ahmed Z, Tainer JA. 2021. Targeting SARS-CoV-2
557 Nsp3 macrodomain structure with insights from human poly(ADP-ribose) glycohydrolase (PARG)
558 structures with inhibitors. Prog Biophys Mol Biol 163:171-186.
559 31. Li J, Zhong F, Li M, Liu Y, Wang L, Liu M, Li F, Zhang J, Wu J, Shi Y, Zhang Z, Tu X, Ruan K, Gao J.
560 2022. Two Binding Sites of SARS-CoV-2 Macrodomain 3 Probed by Oxaprozin and Meclomen. J
561 Med Chem 65:15227-15237.
562 32. Gallo, A., et al., PDB 7P2O, unpublished, 2022.
563 33. Calleja DJ, Kuchel N, Lu BGC, Birkinshaw RW, Klemm T, Doerflinger M, Cooney JP, Mackiewicz L,
564 Au AE, Yap YQ, Blackmore TR, Katneni K, Crighton E, Newman J, Jarman KE, Call MJ, Lechtenberg
565 BC, Czabotar PE, Pellegrini M, Charman SA, Lowes KN, Mitchell JP, Nachbur U, Lessene G,
566 Komander D. 2022. Insights Into Drug Repurposing, as Well as Specificity and Compound
567 Properties of Piperidine-Based SARS-CoV-2 PLpro Inhibitors. Front Chem 10:861209.

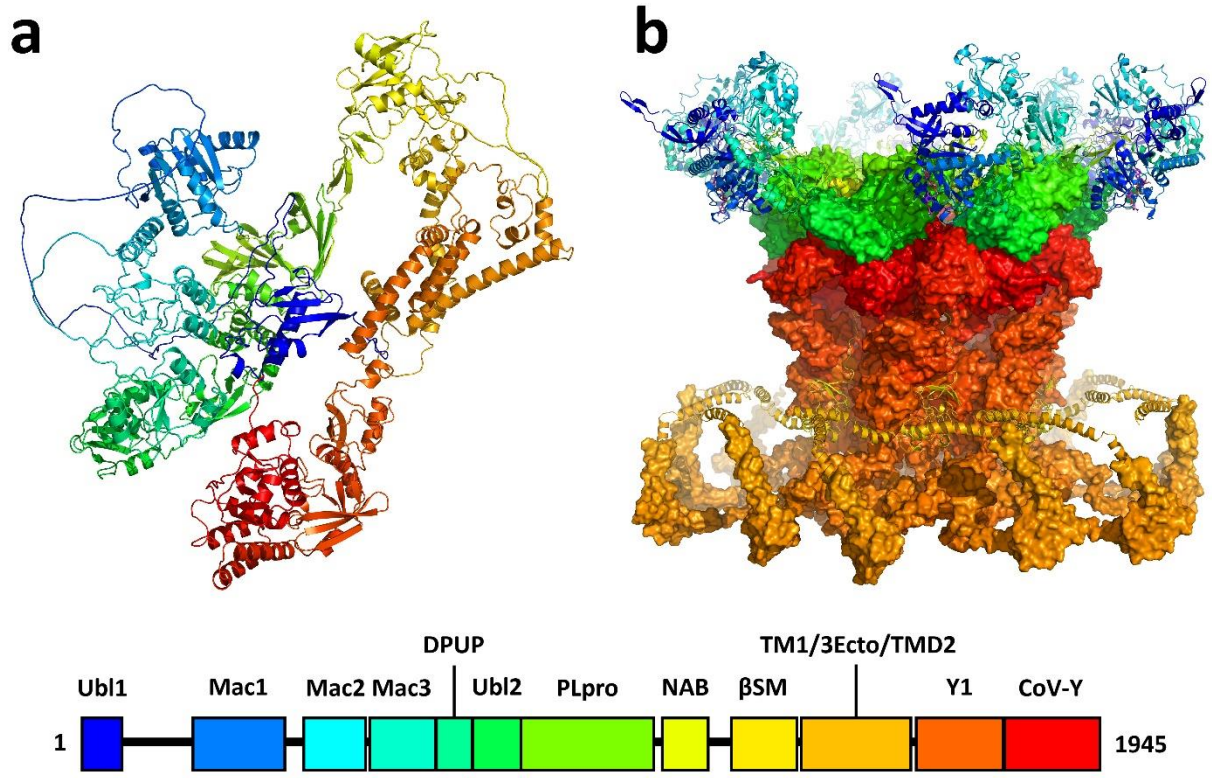
568 34. Stogios, P.J., et al., PDB 7LGO, unpublished, 2021.
569 35. Stogios, P.J., et al., PDB 7T9W, unpublished, 2021.
570 36. Li Y, Pustovalova Y, Shi W, Gorbatyuk O, Sreeramulu S, Schwalbe H, Hoch JC, Hao B. 2023. Crystal
571 structure of the CoV-Y domain of SARS-CoV-2 nonstructural protein 3. Sci Rep 13:2890.
572 37. Tan J, Vonrhein C, Smart OS, Bricogne G, Bollati M, Kusov Y, Hansen G, Mesters JR, Schmidt CL,
573 Hilgenfeld R. 2009. The SARS-unique domain (SUD) of SARS coronavirus contains two
574 macrodomains that bind G-quadruplexes. PLoS Pathog 5:e1000428.
575 38. Lei J, Ma-Lauer Y, Han Y, Thoms M, Buschauer R, Jores J, Thiel V, Beckmann R, Deng W,
576 Leonhardt H, Hilgenfeld R, von Brunn A. 2021. The SARS-unique domain (SUD) of SARS-CoV and
577 SARS-CoV-2 interacts with human Paip1 to enhance viral RNA translation. EMBO J 40:e102277.
578 39. Yan W, Zheng Y, Zeng X, He B, Cheng W. 2022. Structural biology of SARS-CoV-2: open the door
579 for novel therapies. Signal Transduct Target Ther 7:26.
580 40. Sparks JS, Lu X, Denison MR. 2007. Genetic analysis of Murine hepatitis virus nsp4 in virus
581 replication. J Virol 81:12554-63.
582 41. Xu X, Lou Z, Ma Y, Chen X, Yang Z, Tong X, Zhao Q, Xu Y, Deng H, Bartlam M, Rao Z. 2009. Crystal
583 structure of the C-terminal cytoplasmic domain of non-structural protein 4 from mouse hepatitis
584 virus A59. PLoS One 4:e6217.
585 42. Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, Zhang B, Li X, Zhang L, Peng C, Duan Y, Yu J, Wang L,
586 Yang K, Liu F, Jiang R, Yang X, You T, Liu X, Yang X, Bai F, Liu H, Liu X, Guddat LW, Xu W, Xiao G,
587 Qin C, Shi Z, Jiang H, Rao Z, Yang H. 2020. Structure of M(pro) from SARS-CoV-2 and discovery of
588 its inhibitors. Nature 582:289-293.
589 43. Zhang C, Li L, He J, Chen C, Su D. 2021. Nonstructural protein 7 and 8 complexes of SARS-CoV-2.
590 Protein Sci 30:873-881.
591 44. Oostra M, te Lintelo EG, Deijs M, Verheije MH, Rottier PJ, de Haan CA. 2007. Localization and
592 membrane topology of coronavirus nonstructural protein 4: involvement of the early secretory
593 pathway in replication. J Virol 81:12323-36.
594 45. Klatte N, Shields DC, Agoni C. 2022. Modelling the Transitioning of SARS-CoV-2 nsp3 and nsp4
595 Lumenal Regions towards a More Stable State on Complex Formation. Int J Mol Sci 24.
596 46. Hagemeijer MC, Monastyrska I, Griffith J, van der Sluijs P, Voortman J, van Bergen en
597 Henegouwen PM, Vonk AM, Rottier PJ, Reggiori F, de Haan CA. 2014. Membrane
598 rearrangements mediated by coronavirus nonstructural proteins 3 and 4. Virology 458-459:125-
599 35.
600 47. Chen J, Malone B, Llewellyn E, Grasso M, Shelton PMM, Olinares PDB, Maruthi K, Eng ET,
601 Vatandaslar H, Chait BT, Kapoor TM, Darst SA, Campbell EA. 2020. Structural Basis for Helicase-
602 Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. Cell 182:1560-1573
603 e13.
604 48. Yan L, Ge J, Zheng L, Zhang Y, Gao Y, Wang T, Huang Y, Yang Y, Gao S, Li M, Liu Z, Wang H, Li Y,
605 Chen Y, Guddat LW, Wang Q, Rao Z, Lou Z. 2021. Cryo-EM Structure of an Extended SARS-CoV-2
606 Replication and Transcription Complex Reveals an Intermediate State in Cap Synthesis. Cell
607 184:184-193 e10.
608 49. Liu C, Shi W, Becker ST, Schatz DG, Liu B, Yang Y. 2021. Structural basis of mismatch recognition
609 by a SARS-CoV-2 proofreading enzyme. Science doi:10.1126/science.abi9310.
610 50. Rosas-Lemus M, Minasov G, Shuvalova L, Inniss NL, Kiryukhina O, Brunzelle J, Satchell KJF. 2020.
611 High-resolution structures of the SARS-CoV-2 2'-O-methyltransferase reveal strategies for
612 structure-based inhibitor design. Sci Signal 13.
613 51. Frazier MN, Wilson IM, Krahn JM, Butay KJ, Dillard LB, Borgnia MJ, Stanley RE. 2022. Flipped
614 over U: structural basis for dsRNA cleavage by the SARS-CoV-2 endoribonuclease. Nucleic Acids
615 Res 50:8290-8301.

616 52. Shaqra AM, Zvornicanin SN, Huang QYJ, Lockbaum GJ, Knapp M, Tandeske L, Bakan DT, Flynn J,
617 Bolon DNA, Moquin S, Dovala D, Kurt Yilmaz N, Schiffer CA. 2022. Defining the substrate
618 envelope of SARS-CoV-2 main protease to predict and avoid drug resistance. Nat Commun
619 13:3556.
620 53. Goyal B, Goyal D. 2020. Targeting the Dimerization of the Main Protease of Coronaviruses: A
621 Potential Broad-Spectrum Therapeutic Strategy. ACS Comb Sci 22:297-305.
622 54. Chen S, Chen L, Tan J, Chen J, Du L, Sun T, Shen J, Chen K, Jiang H, Shen X. 2005. Severe acute
623 respiratory syndrome coronavirus 3C-like proteinase N terminus is indispensable for proteolytic
624 activity but not for enzyme dimerization. Biochemical and thermodynamic investigation in
625 conjunction with molecular dynamics simulations. J Biol Chem 280:164-73.
626 55. Kaptan S, Girych M, Enkavi G, Kulig W, Sharma V, Vuorio J, Rog T, Vattulainen I. 2022.
627 Maturation of the SARS-CoV-2 virus is regulated by dimerization of its main protease. Comput
628 Struct Biotechnol J 20:3336-3346.
629 56. Klein S, Cortese M, Winter SL, Wachsmuth-Melm M, Neufeldt CJ, Cerikan B, Stanifer ML, Boulant
630 S, Bartenschlager R, Chlanda P. 2020. SARS-CoV-2 structure and replication characterized by in
631 situ cryo-electron tomography. Nat Commun 11:5885.
632 57. Chen A, Lupan AM, Quek RT, Stanciu SG, Asaftei M, Stanciu GA, Hardy KS, de Almeida Magalhaes
633 T, Silver PA, Mitchison TJ, Salic A. 2024. A coronaviral pore-replicase complex links RNA synthesis
634 and export from double-membrane vesicles. Sci Adv 10:eadq9580.
635 58. Andronov L, Han M, Zhu Y, Balaji A, Roy AR, Barentine AES, Patel P, Garhyan J, Qi LS, Moerner
636 WE. 2024. Nanoscale cellular organization of viral RNA and proteins in SARS-CoV-2 replication
637 organelles. Nat Commun 15:4644.
638 59. Subissi L, Posthuma CC, Collet A, Zevenhoven-Dobbe JC, Gorbalenya AE, Decroly E, Snijder EJ,
639 Canard B, Imbert I. 2014. One severe acute respiratory syndrome coronavirus protein complex
640 integrates processive RNA polymerase and exonuclease activities. Proc Natl Acad Sci U S A
641 111:E3900-9.
642 60. Kirchdoerfer RN, Ward AB. 2019. Structure of the SARS-CoV nsp12 polymerase bound to nsp7
643 and nsp8 co-factors. Nat Commun 10:2342.
644 61. Slanina H, Madhugiri R, Bylapudi G, Schultheiss K, Karl N, Gulyaeva A, Gorbalenya AE, Linne U,
645 Ziebuhr J. 2021. Coronavirus replication-transcription complex: Vital and selective NMPylation of
646 a conserved site in nsp9 by the NiRAN-RdRp subunit. Proc Natl Acad Sci U S A 118.
647 62. Deng X, Hackbart M, Mettelman RC, O'Brien A, Mielech AM, Yi G, Kao CC, Baker SC. 2017.
648 Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in
649 macrophages. Proc Natl Acad Sci U S A 114:E4251-E4260.
650 63. Pathak R, Eliscovich C, Mena I, Cupic A, Rutkowska M, Chandran K, Jangra RK, Garcia-Sastre A,
651 Singer RH, Kalpana GV. 2024. Visualization of Early RNA Replication Kinetics of SARS-CoV-2 by
652 Using Single Molecule RNA-FISH Combined with Immunofluorescence. Viruses 16.
653 64. Sawicki SG, Sawicki DL. 1986. Coronavirus minus-strand RNA synthesis and effect of
654 cycloheximide on coronavirus RNA synthesis. J Virol 57:328-34.
655 65. Donaldson EF, Graham RL, Sims AC, Denison MR, Baric RS. 2007. Analysis of murine hepatitis
656 virus strain A59 temperature-sensitive mutant TS-LA6 suggests that nsp10 plays a critical role in
657 polyprotein processing. J Virol 81:7086-98.
658 66. Schmidt N, Ganskih S, Wei Y, Gabel A, Zielinski S, Keshishian H, Lareau CA, Zimmermann L,
659 Makroczyova J, Pearce C, Krey K, Hennig T, Stegmaier S, Moyon L, Horlacher M, Werner S, Aydin
660 J, Olguin-Nava M, Potabattula R, Kibe A, Dolken L, Smyth RP, Caliskan N, Marsico A, Krempl C,
661 Bodem J, Pichlmair A, Carr SA, Chlanda P, Erhard F, Munschauer M. 2023. SND1 binds SARS-CoV-
662 2 negative-sense RNA and promotes viral RNA synthesis through NSP9. Cell 186:4834-4850 e23.

663 67. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R,
664 Zidek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B,
665 Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M,
666 Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli
667 P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. Nature
668 596:583-589.
669 68. Schrödinger Release 2023-1: Schrödinger, LLC, New York, NY, 2023.
670 69. Kozakov D, Brenke R, Comeau SR, Vajda S. 2006. PIPER: an FFT-based protein docking program
671 with pairwise potentials. Proteins 65:392-406.
672 70. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang
673 YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. 2020. A new coronavirus
674 associated with human respiratory disease in China. Nature 579:265-269.
675 71. Eriksson KK, Cervantes-Barragan L, Ludewig B, Thiel V. 2008. Mouse hepatitis virus liver
676 pathology is dependent on ADP-ribose-1''-phosphatase, a viral function conserved in the alpha-
677 like supergroup. J Virol 82:12325-34.

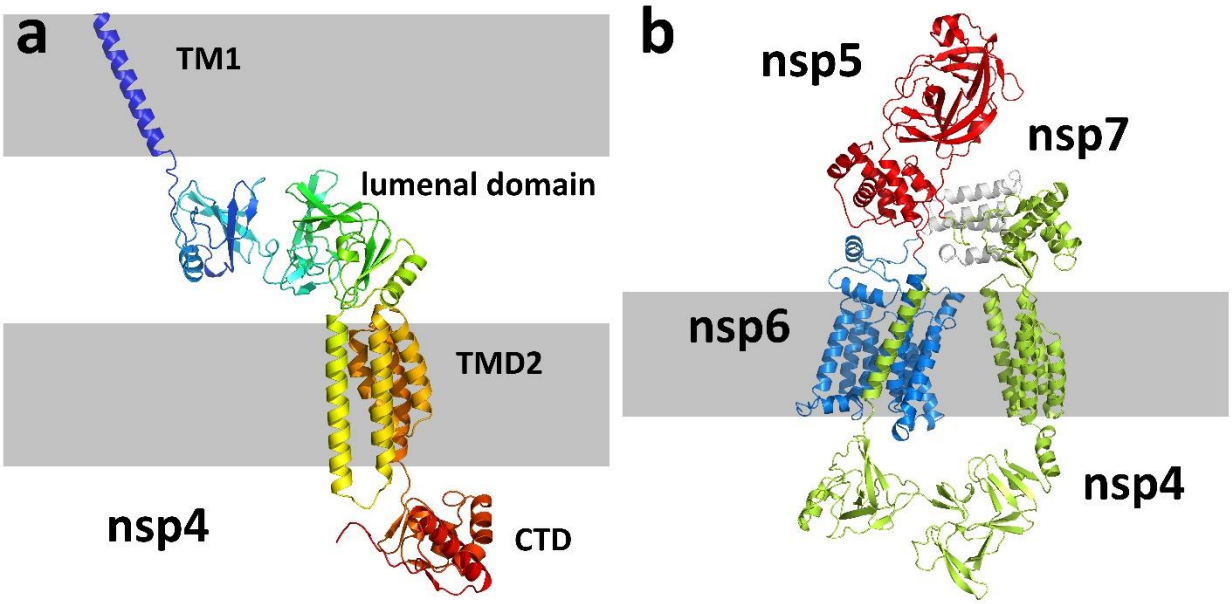
678

679 **Figures**



680

681 **Figure 1.** a) AlphaFold predicted structure of SARS-CoV-2 nsp3. The domain organization of
682 nsp3 is indicated below, colored blue to red across its 1945 residues. The nsp3 structure was
683 largely well predicted, with a few exceptions, including a long, disordered loop corresponding
684 to the highly variable region (HVR) following Ubl1. b) Remodeled nsp3 dodecameric component
685 of the DMV pore, based on the cryo-ET structure and maps of Huang, et al.(11) Portions of the
686 structure rendered with surface representation are preserved from the published structure.
687 Portions shown in cartoon representation have been added or remodeled.



688

689 **Figure 2.** a) Alphafold predicted structure of nsp4, colored blue to red across its 500 residues.
690 The model reflects a conformation consistent with association to a double membrane
691 (indicated by the gray bars), where the lumenal domain resides in the membrane gap. b)
692 Predicted structure of the uncleaved nsp4-nsp7 polyprotein, with nsp4 in green, nsp5 in red,
693 nsp6 in blue and nsp7 in white. The model adopts a conformation consistent with binding to
694 the ER membrane, where the nsp4 N-terminus and CTD, nsp5 and nsp7 are cytosolic and the
695 nsp4 lumenal domain is in the ER lumen.

696

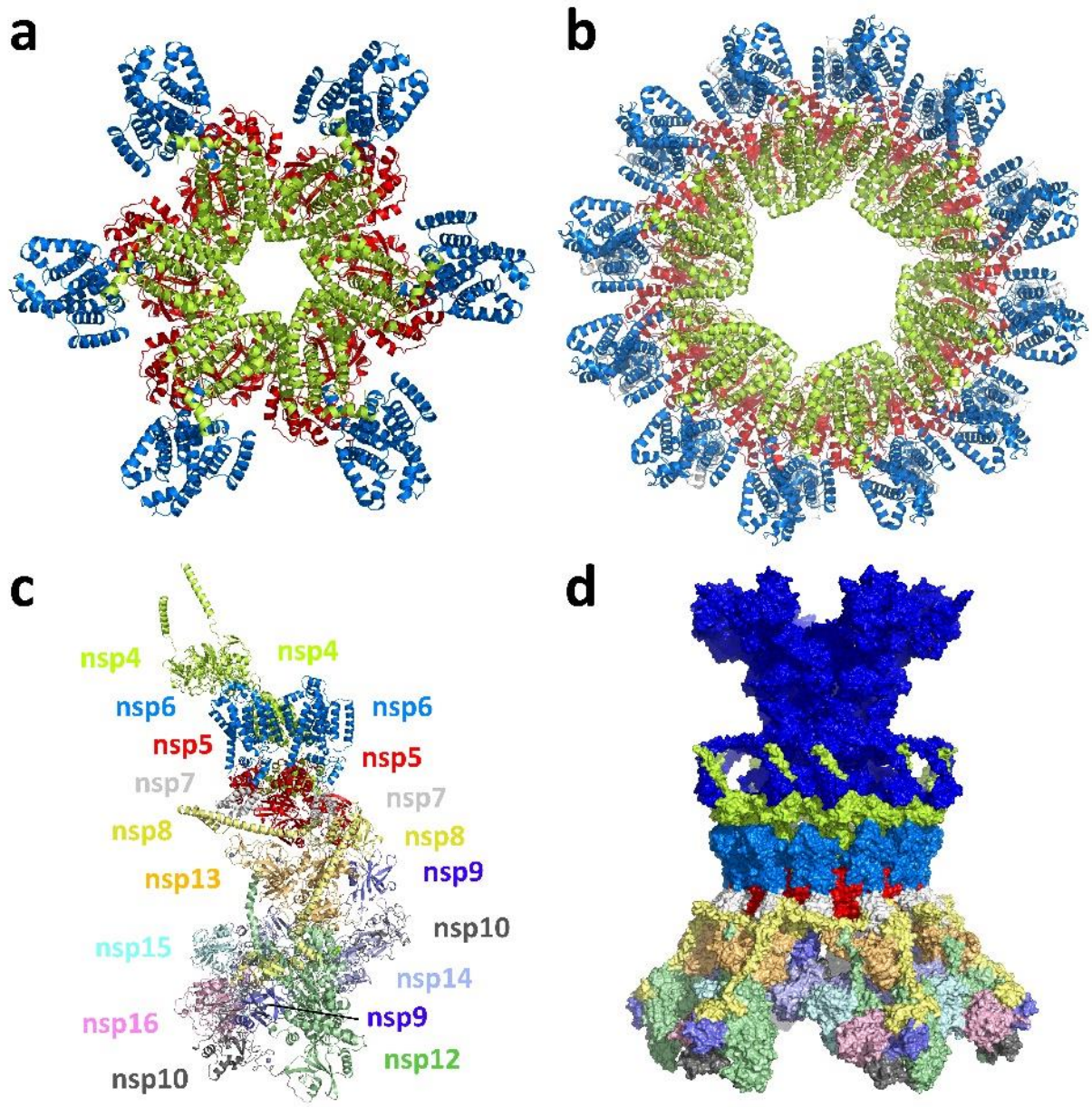
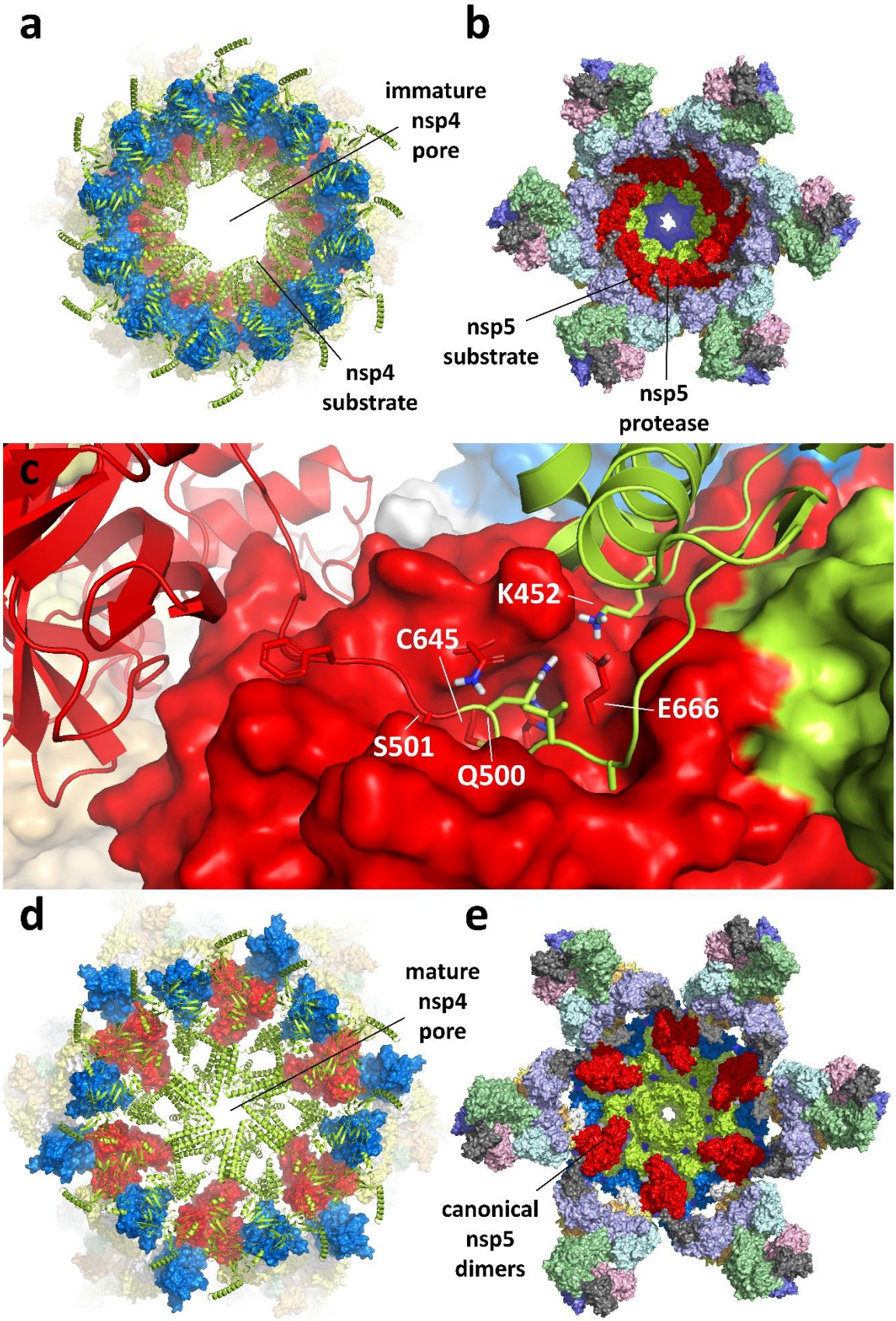


Figure 3. a) AlphaFold prediction of the hexameric SARS-CoV-2 nsp4-nsp6 polyprotein (a truncated construct covering residues 256-1042). Nsp4 (lime green) forms a pore-like interior, flanked by nsp6 (marine blue). Nsp5 (red) forms a close packed ring-like arrangement on the underside. b) The nsp4-nsp6 hexamer transformed into a dodecamer. The nsp4 TMD2 alternates between open and closed conformations, facilitating pore maturation upon nsp4-nsp5 cleavage. c) The pp1a'/pp1ab' heterodimer. The dimer incorporates known interactions

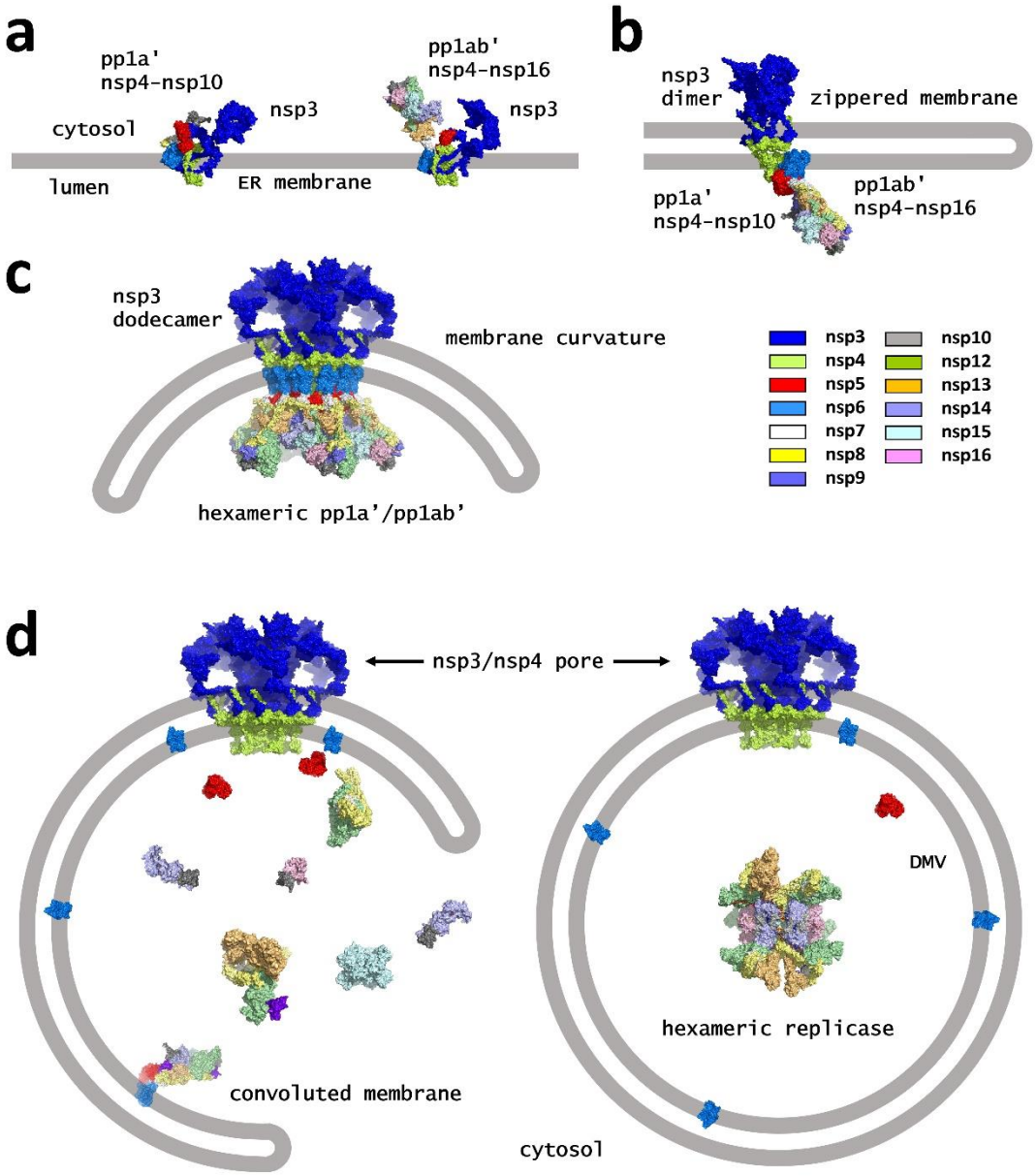
704 between the nsp proteins to the maximum extent. The nsp4 conformation is consistent with
705 coordination to a double membrane and formation of the dodecameric ring. d) Model of the
706 complete DMV immature pore, prior to nsp4-nsp5 cleavage. The precursor to the pore is
707 formed from twelve subunits of nsp3 (blue) and six subunits each of uncleaved pp1a' and
708 pp1ab'.


a
immature
nsp4
pore
nsp4
substrate
b
nsp5
substrate
nsp5
protease
c
K452
C645
E666
S501
Q500
d
mature
nsp4
pore
e
canonical
nsp5
dimers

709
710

711 **Figure 4.** Proposed mechanism for nsp5 protease activation and pore maturation. a) View from
712 above and b) view from below the immature pore in its pre-cleavage state. Six alternating nsp5
713 subunits within the dodecamer formed from pp1a' and pp1ab' assume a conformation which
714 allows *trans*-cleavage of the nsp4-nsp5 linker by the remaining nsp5 subunits. c) Detail of the
715 nsp5 protease active site shows cleavage occurs between Q500 on nsp4 and S501 on nsp5
716 (numbering consistent with the uncleaved polyprotein pp1ab'). The interaction between nsp4
717 K452 and nsp5 E666 mimics the interaction between the nsp5 N-terminus and the same
718 glutamic acid residue in the canonical nsp5 dimer. d) View from above and e) view from below
719 the post-cleavage mature pore. The reduced structural constraints from the nsp4-nsp5 cleavage
720 of ppa1b' allows the nsp4 subunits to adopt the conformation seen in the structure by Huang,
721 et al.(11) Despite pp1a' remaining uncleaved at this point, canonical nsp5 dimers can form in
722 which one of the two protease active sites should be considered active and well positioned to
723 cleave additional linkages.

724



725

726 **Figure 5.** Schematic of pore formation, membrane remodeling and polyprotein cleavage. a)
727 Following protease cleavage by PLpro, nsp3 and the polyprotein intermediates pp1a' (nsp4-
728 nsp10) and ppa1b' (nsp4-nsp16) are expressed on the ER membrane. Given their initial
729 proximity, nsp3/pp1a' and nsp3/pp1ab' dimers may form first. b) Interaction between nsp3 and
730 nsp4 leads to a conformational change in the nsp4 lumen domain which drives membrane

731 zippering, while association between pp1a' and pp1ab' creates the basic $(nsp3)_2$(pp1a')(pp1ab')
732 building block for construction of the pore. c) Aggregation leads to formation of the
733 dodecameric pore precursor and membrane curvature. The close packing of nsp5 in this
734 complex activates its protease. d) Cleavage of the polyprotein prior to DMV formation releases
735 viral proteins into the cytosol (left). Cleavage following DMV formation traps the viral proteins
736 within the DMV, creating ideal conditions for the previously predicted hexameric replicase to
737 form.