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

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

Word count: Abstract – 205. Text – 7240.

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 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 polymerase and other viral proteins are recruited to the
37 DMV represents a major gap in our current knowledge. Here we present a novel, 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,(2-8) 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 principal drivers of membrane remodeling and
53 DMV formation.(9) The general picture that has emerged from this work is that nsp3 and nsp4
54 are primarily responsible for membrane doubling and DMV formation, while nsp6 is responsible
55 for membrane zippering and the tethering of DMV clusters to lipid droplets and the ER.(10)
56 Initial images of severe acute respiratory syndrome (SARS) DMVs suggested a closed
57 organelle,(11) however more detailed cryo-electron tomography (cryo-ET) images have
58 captured multiple distinct pores embedded in the membranes of each DMV, providing a
59 necessary mode of egress for newly synthesized RNA. Low resolution cryo-ET reconstructions
60 for both murine hepatitis virus (MHV)(12) and SARS-CoV-2(13) clearly revealed pore structures
61 with distinct six-fold symmetric cytosolic crowns terminating in prominent prong-like features.
62 In the case of SARS-CoV-2, this cryo-ET work also demonstrated that DMV formation and pore

63 formation go hand in hand, and that nsp3 and nsp4 are the minimal viral components required.
64 A recent 4.2 Å resolution cryo-ET structure further revealed the pore in even greater detail,
65 specifically showing the crown to be composed of twelve subunits of nsp3, anchored to a
66 dodecameric arrangement of nsp4 spanning both inner and outer membranes.(14)

67 Nsp3 and nsp4 can be found on the first open reading frame (ORF1) of the coronavirus genome.
68 Encompassing over 70% of the genome, ORF1 encodes two viral polyproteins, pp1a (nsp1-
69 nsp10) and pp1ab (nsp1-nsp16). They are distinguished by a ribosomal frameshift located
70 between nsp10 and nsp1*2*.(15) Both polyproteins are cleaved by viral proteases embedded in
71 nsp3 and nsp5. Studies of MHV and SARS established that the first three non-structural proteins
72 (nsp1-nsp3) are efficiently cleaved by the papain-like proteases of nsp3 (PLpro in SARS-CoV-
73 2),(16, 17) while the remainder of the polyprotein (nsp4-nsp10, referred to here as pp1a', and
74 nsp4-nsp16, referred to as pp1ab') are cleaved in a more regulated manner by the main
75 protease (Mpro) of nsp5.(18, 19) Notably, biochemical studies of cleaved nsp5 have shown the
76 protease to be significantly more active when dimerized,(20) but the factors that initiate nsp5
77 maturation and polyprotein processing have yet to be elucidated.(21) Considering that
78 uncleaved nsp5 is sandwiched between the membrane binding proteins nsp4 and nsp6,
79 additional constraints on its initial activity should be expected, but these constraints have not
80 been explored and remain poorly understood. With respect to the DMV pore, it has been
81 clearly demonstrated that cleavage between nsp3 and nsp4 is required prior to pore
82 formation,(13, 22) but no comparable study has established when in the process of pore
83 formation cleavage between nsp4 and nsp5 occurs. Thus, whether nsp5 maturation occurs
84 prior to pore formation or in concert with it is an open question.

85 Once the pp1a' and pp1ab' polyproteins are fully cleaved, the resulting individual proteins,
86 nsp7-nsp16, constitute the viral replication machinery.(23) At a minimum, RNA polymerization
87 is carried out by a complex of (nsp12)(nsp7)$(nsp8)_2$.(24) Additional enzymatic domains of
88 nsp12,(25) nsp13,(26, 27) nsp14(28, 29) and nsp16,(30) aided by the cofactors nsp9 and nsp10,
89 are responsible for other critical viral RNA replication functions including proofreading and
90 mRNA capping. Previously, we proposed these viral proteins assemble around the
91 endonuclease nsp15 to form a hexameric complex of over 60 subunits.(31) Our modeling
92 suggested this complex efficiently coordinates the various functions just described. Implicitly,
93 these proteins must be recruited to the DMV prior to its closure for replication inside the
94 organelle to occur, although how that might be achieved is unknown, representing a major gap
95 in our knowledge of coronavirus replication.(6-8)

96 Here we tackle the question of pore formation and cleavage of the polyprotein from a
97 structural modeling approach. We first review our attempts at constructing a model of the
98 nsp3/nsp4 pore structure, highlighting the successes and limitations of AlphaFold in such an
99 endeavor. We ultimately propose modest changes to the modeled prongs in the published
100 structure of Huang, et al.(14) We then examine constraints on nsp5 protease activity imposed
101 by the flanking membrane binding proteins nsp4 and nsp6 and describe our discovery of what
102 appears to be a polyprotein precursor to pore formation which overcomes those constraints.
103 We propose that when this pore precursor is formed from six heterodimers of pp1a' and
104 pp1ab', the cleaved products will be trapped within the DMV and have the proper
105 stoichiometry to form the previously proposed hexameric replication complex. The implications
106 of the proposed model are far ranging, with direct relevance to membrane remodeling,

107 regulation of nsp5 protease activity, and formation of the replication and transcription complex
108 within the closed vesicles.

109 **Results**

110 *SARS-CoV-2 nsp3 monomer*

111 At 1945 residues, nsp3 is the largest protein produced by the SARS-CoV-2 virus. While most of
112 the individual domains have been solved by X-ray crystallography, AlphaFold2.1 was used to
113 predict the full-length structure (Figure 1a). The majority of the protein, with some notable
114 exceptions, was predicted with high confidence (Figure S1), as measured by the predicted Local
115 Distance Difference Test (pLDDT),(32) with most domains scoring in the 70-90 range. X-ray
116 structures for Ubl1 (ubiquitin-like domain 1)(33), Mac1 (macrodomain 1)(34), Mac3(35), DPUP
117 (domain preceding Ubl2 and PLpro)(36), Ubl2/PLpro(37), NAB (nucleic acid binding domain)(38)
118 and βSM (betacoronavirus specific marker)(39) were clearly reproduced in the AlphaFold
119 model. Underscoring the success of the method, following completion of this work, a structure
120 (PDB 8ILC) became available for the SARS-CoV-2 C-terminal CoV-Y domain (Y2, Y3 and Y4)(40)
121 that matched the AlphaFold prediction as well. Additional domains with no available structures
122 for any coronavirus were predicted for TM1, 3Ecto, TM2 (a bundle more accurately described
123 as transmembrane domain 2 or TMD2) and Y1. The pLDDT score for the region between βSM
124 and Y1, which includes TM1-3Ecto-TMD2, was comparatively lower, at 64.5, but notably the
125 residues immediately following βSM (1318-1364) were predicted to be α-helical, with residues
126 1318-1349 having amphipathic character. The residues prior to TM1 (1385-1400) were
127 predicted to form a second amphipatheic helix.

128 The most notable failure in the AlphaFold prediction was for the Mac2 domain, which deviated
129 significantly from X-ray structures for the highly homologous SARS Mac2 domain (71% identity,
130 85% similarity)(41). This was unexpected, given the SARS structure's inclusion in the AlphaFold
131 training dataset, but it appeared to be unique to the prediction of full-length nsp3. Predictions
132 of the isolated domain or the Mac2 and flanking domains, were consistent with the SARS
133 structures, putting its pLDDT score in line with other domains (Figure S1). Prediction of the
134 heterodimer of Mac2 with human Paip1 was also consistent with the existing SARS Mac2/Paip1
135 X-ray crystal structure (PDB 6YXJ).(42)

136 In general, variations in the predicted structure conformations, coupled with the pLDDT
137 analysis, reflect significant flexibility between the domains (see Figure S2). This is most striking
138 for the highly variable region (HVR) between Ubl1 and Mac1, which was completely disordered
139 in the AlphaFold prediction. In addition, the loops connecting Mac1 and Mac2 (residues 388-
140 415) and connecting NAB and βSM (residues 1197-1240) are particularly long. Ultimately, these
141 linkers provided a set of constraints on the maximum distance and range of motion between
142 domains that we considered when examining the nsp3 component of the pore.

143 *SARS-CoV-2 nsp3 dodecamer*

144 Early analysis of the SARS-CoV-2 and MHV pore structures determined from cryo-ET suggested
145 that the cystosolic crown is composed of a hexamer of nsp3, with the N-terminus residing in the
146 prongs(12, 13). Due to practical limitations in the size of structures that could be predicted with
147 AlphaFold (about 5000 residues), it wasn't possible to generate a model for an entire nsp3
148 hexamer (11,670 residues) and exploration of truncated constructs had only limited success.

149 The publication of the 4.2 Å cryo-ET structure however provided surprising clarity.(14) The six-
150 fold symmetric crown is instead a dodecamer of nsp3, although much of the protein structure
151 was not resolved. Specifically, the C-terminal Y1/CoV-Y domains form the dodecameric base of
152 the crown and the cytosolic portion of the pore. It is constructed from an inner hexameric ring
153 of Y1/Y2 (forming the pore), which is surrounded by another hexameric outer Y1/Y2 ring to
154 create the base of the crown. The twelve CoV-Y Y3/Y4 domains interlace above the base,
155 forming the bulk of the crown. The crown is embedded in the outer membrane through TMD2,
156 interacting with nsp4. Here we note that the clearest success among the AlphaFold predictions
157 of multimeric nsp3 constructs was that of the inner Y1/Y2 hexameric ring, shown in Figure S3 in
158 comparison to the cryo-ET structure.

159 The N-terminal domains preceding TM1 are only partially resolved for six subunits and not
160 resolved at all for the other six, implying a degree of disorder, which may reflect real dynamics
161 of the pore or the lack of other key stabilizing elements such as RNA or additional host or viral
162 proteins. Alternatively, it may simply be an artifact of how the structure was obtained and
163 refined (the cryo-ET structure was determined from DMVs derived from nsp3-nsp4 expression
164 in HEK293F cells, not DMVs derived from infected cells). What is definitively resolved is that a
165 hexameric ring of Ubl2/PLpro rests atop the CoV-Y dodecamer to build up the crown. Huang, et
166 al. also fit each of the six prongs with the Mac2, Mac3 and DPUP domains from one subunit and
167 the NAB domain from a neighboring subunit. However, the resolution of these prongs is
168 considerably weaker (>9 Å), making the fit ambiguous. Several considerations led us to remodel
169 this part of the structure. Huang, et al. assumed the prong density was primarily composed of
170 the Mac2, Mac3 and NAB domains and used AlphaFold to predict a complex of these domains,

171 which they docked into the density. This fit put the NAB domain implausibly far from the
172 assumed position of the βSM domain at the base of the crown. In contrast, they did not fit the
173 Ubl1, HVR or Mac1 domains, which Zimmermann, et al.(13) previously showed was critical to
174 the proper formation of the prongs. Considering that Mac1 corresponds to the only Mac
175 domain in MHV nsp3, we reasoned that it was an important component of the prong density
176 and was likely similarly positioned in both structures., Taking into account functional
177 constraints, similarities and differences between MHV and SARS-CoV-2, and all available cryo-
178 ET maps, including the prior maps for both MHV and SARS-CoV-2, we fit the prongs with Ubl1,
179 Mac1, Mac2, Mac3 and DPUP, omitting only the HVR domain for lack of structural information.
180 Furthermore, we positioned the NAB domain between PLpro domains in the rim of the crown.
181 We also modeled the βSM domain and two amphipathic helices on the outer membrane
182 cytosolic surface, following previously unfit density. As a proof of principle, we connected the
183 NAB and βSM domains with the 44-residue linker. A similar model was built for the MHV nsp3
184 dodecamer to establish consistency across diverse coronaviruses. The final remodeled nsp3
185 crown is shown in Figures 1b and 1c and described in greater detail in the Supplementary
186 Information and Figures S4, S5 and S6.

187 *SARS-CoV-2 nsp4 monomer*

188 Our model of nsp4 (Figure 2a) is consistent with previous AlphaFold models with four distinct
189 domains.(43) The N-terminal TM1 is a single α-helix which spans residues 1-27. This is followed
190 by the lumenal domain, a soluble region composed of two distinct lobes connected by a 9-
191 residue linker. The previously designated TM2/TM3/TM4 helices(44) were instead predicted by
192 AlphaFold to be a 5-helix bundle, spanning residues 277-398. For simplicity, we'll refer to this

193 region as transmembrane domain 2 (TMD2), with specific helices α1-α5. Finally, the AlphaFold
194 predicted C-terminal domain (CTD) is another soluble region which reproduces existing X-ray
195 structures(45). The average pLDDT score for the full protein was 83.0 (Figure S7).

196 In none of the predicted structures does TM1 interact with TMD2. The conformation is instead
197 consistent with nsp4 association with the DMV double membrane, where the nsp4 N-terminus
198 is cytosolic, TM1 sits in the outer membrane, the lumenal domain is positioned in the
199 membrane gap, TMD2 sits in the inner membrane, and the nsp4 C-terminal domain (CTD) is
200 exposed to the DMV interior. Thus, the overall topology of the AlphaFold predicted nsp4
201 structure is consistent with that seen in the detailed pore structure. Details of the predicted
202 structure, such as the lumenal domain and TMD2, were also confirmed by the cryo-ET
203 structure, with the only notable difference being a conformational shift of helix α1 in TMD2, to
204 be discussed below in more detail.

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

206 To investigate constraints on the nsp5 protease imposed by its position within the polyprotein,
207 we generated AlphaFold models for the uncleaved nsp4-nsp5-nsp6-nsp7 construct (Figure 2b).
208 The pLDDT declined slightly to an average of 77.0 (Figure S8), but X-ray structures of nsp5(46)
209 and nsp7(47) were well reproduced, and nsp6 was predicted to form an 8-helix bundle. The
210 overall organization of the polyprotein showed some flexibility between the individual domains.
211 However, in all predicted structures, the nsp4 TMD2 was aligned with nsp6 as if bound to a
212 common membrane. Unlike the prediction for cleaved nsp4, the lumenal domain of nsp4 in the
213 polyprotein adopted a conformation between its N- and C-lobes (linked by residues 124-132)

214 which oriented TM1 such that it associated with the nsp6 helix bundle (Figure S9). This would
215 be more consistent with association with the single ER membrane rather than the DMV double
216 membrane, making nsp5, nsp7, the nsp4 N-terminus and the nsp4 CTD cytosolic.(48) The
217 observation that this conformational change is driven by the nsp4 lumenal domain is consistent
218 with previous characterizations of this domain by Klatte, et al.(49)

219 The nsp5 protease dimer is known to be significantly more active than the monomer,(20, 21)
220 but a major conclusion from this work is that despite a relatively long flexible loop connecting
221 nsp4 and nsp5 (residues 488-510), *cis*-cleavage of these proteins does not appear to be
222 feasible. The cleavage site on the connecting loop (residues 500-501) simply cannot reach the
223 protease active site without compromising the integrity of the protease structure (Figure 3a).
224 Similarly, *cis*-cleavage of the nsp5 and nsp6 proteins at residues 807-808 does not appear to be
225 possible under any circumstances. Thus, in both cases, cleavage must proceed via interaction
226 with the protease of a second polyprotein or an already cleaved nsp5 dimer.

227 *SARS-CoV-2 nsp4-nsp5-nsp6-nsp7 dimer*

228 The cleaved nsp5 dimer has been exhaustively studied through X-ray crystallography for
229 multiple coronaviruses.(50-52) A key factor in the increased protease activity of the dimer as
230 compared to the monomer is how the N-terminal amine of one monomer engages with the S1
231 active site pocket of the other monomer. This has been shown to be critical to stabilizing the
232 active site in both SARS(50, 53, 54) and SARS-CoV-2,(52) enhancing the protease's catalytic
233 activity. We considered if a canonical nsp5 dimer could form within the uncleaved nsp4-nsp5-
234 nsp6-nsp7 polyprotein, and if so, would it be capable of self-cleavage. Using a truncated

235 construct, in which the nsp4 TM1 and lumenal domains were removed, we generated an
236 AlphaFold model of the uncleaved dimer possessing approximately C2 symmetry (Figure 3b). In
237 this model, nsp4 and nsp6 are aligned as if bound to a common membrane, with the canonical
238 nsp5 dimer oriented such that its domains 1 rest against the membrane and domains 3 are
239 most exposed to the cytosol.

240 The nsp4-nsp5 cleavage site P1 Q500 residue can be seen to occupy the S1 pocket of the active
241 site. Yet the full nsp4-nsp5 cleavage peptide is not properly aligned, and critically the P1' S501
242 cleavage site residue cannot be positioned in the S1' pocket due to constraints from the nsp5
243 dimer itself (Figure 3c). Thus, the dimer is incapable of cleaving nsp4/nsp5. To stress this point,
244 a 100ns molecular dynamics (MD) simulation of the membrane bound dimer maintained this
245 interaction between the glutamine residue and the S1 site over the entire simulation (Figure
246 S10). We conclude from this analysis that not only does the uncleaved dimer not have the
247 advantage of the cleaved nsp5 N-terminus to stabilize the S1 site and activate the protease it
248 has the uncleaved nsp4 C-terminal glutamine residue blocking the site from acting on any other
249 substrate.

250 If additional steric and conformational constraints from the extended polyprotein or potentially
251 an nsp3 interaction prove to be more dominant, it should be considered that this canonical
252 dimer may not even form until some cleavage has already occurred via a different route.
253 Underscoring this point, the nsp5 protein of MERS only weakly dimerizes, depending instead on
254 a ligand-induced dimerization mechanism.(51) Indeed, as seen in Figure 3d, taking the same
255 nsp4-nsp5-nsp6-nsp7 dimer and assuming at least one nsp4/nsp5 cleavage has occurred via
256 some yet to be determined mechanism, the resulting nsp5 N-terminal amine then activates the

257 protease of the uncleaved subunit. The nsp5-nsp6 and nsp6-nsp7 cleavage sites of the cleaved
258 monomer can be observed to be in close proximity to the activated protease and could be
259 potential substrates. This analysis suggests that an initial nsp4-nsp5 cleavage event is required
260 but does not proceed through the canonical nsp5 dimer. Some other association between
261 polyproteins must be responsible.

262 *SARS-CoV-2 nsp4 dodecamer*

263 As with nsp3, our initial assumption was that the nsp4 component of the pore was hexameric.
264 Based on our predicted structures for the nsp4 monomer and previous work suggesting nsp3
265 interacts with the nsp4 lumenal domain,(49, 55) we reasoned TM1 must sit in the outer
266 membrane, positioning the lumenal domain in the membrane gap. This suggested the portion
267 of the pore embedded in the inner membrane would be formed from a hexamer of the nsp4
268 TMD2 domains, with the nsp4 CTD exposed to the DMV interior. Thus, we worked with a
269 truncated construct which spanned residues 256-500 of nsp4 (TMD2 and CTD). AlphaFold
270 indeed predicted several plausible hexameric structures with pore-like properties, but none
271 was entirely consistent with the available cryo-ET maps. However, upon close examination of
272 the MHV map, we identified density within the membrane gap which was similar in size to the
273 nsp4 lumen domain, but which indicated nsp4 was in fact dodecameric, not hexameric.
274 Unfortunately, AlphaFold predictions of dodecameric nsp4 (residues 256-500) did not produce
275 structures with the expected C6 symmetry. The detailed pore structure from Huang, et al.
276 confirmed that the nsp4 component is indeed a dodecamer, characterized by a central
277 hexameric ring interlaced with an outer hexameric ring. While failing to produce the full

278 dodecameric nsp4 assembly, AlphaFold in fact accurately predicted the structure of the inner
279 hexameric ring, as shown in Figure S11 in comparison to the cryo-ET structure.

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

281 As the initial nsp5 cleavage events appeared to be unresolved, we considered the possibility of
282 pore formation prior to nsp4/nsp5 cleavage. Our preliminary exploration with AlphaFold also
283 considered hexamers of uncleaved nsp4-nsp5-nsp6 (truncated to residues 256-1042). AlphaFold
284 generated a particularly compelling model shown in Figure 4a and Figure S12, which consisted
285 of an inner ring of nsp4 TMD2 surrounded by an outer ring of nsp6. The two transmembrane
286 domains are linked via an inner, close-packed solvent exposed ring formed from the nsp4 CTD,
287 which is surrounded by a larger close-packed ring formed from nsp5. Consistent with the
288 emerging topology of the nsp3/nsp4 pore, both nsp5 and the nsp4 CTD would be exposed to
289 the interior of the DMV in this model.

290 As we came to understand the stoichiometry of the nsp4 component of the pore was
291 dodecameric instead of hexameric, we found it was straightforward to envision similar close-
292 packed polyprotein rings of increasing size. As an example, the chikungunya virus protein nsP1
293 has been shown via cryo-EM to form a dodecameric ring.(56) We used AlphaFold to look at a
294 series of smaller aggregates (hexamer, octamer, decamer), and a similarly arranged ring was
295 predicted in each case (see Figure S13a). Unfortunately, a dodecameric arrangement of nsp4-
296 nsp5-nsp6 or any ring larger than a hexamer was beyond the practical limits of AlphaFold. To
297 generate such a model, we instead resorted to a simple geometric expansion of the nsp4-nsp5-
298 nsp6 hexamer. This was executed by defining the center of mass of each subunit within the

299 hexamer, measuring the distance between adjacent subunits, and translating and rotating
300 duplicate sets of these subunits to form a dodecameric ring in which this distance was
301 maintained. Structural refinement from this starting point (as detailed below) led to the fully
302 optimized construct shown in Figure 4b. A similar approach was applied to the AlphaFold model
303 of hexameric chikungunya nsP1 to yield a dodecameric model which reproduces the cryo-EM
304 structure well (Figure S13b).

305 In comparison to the earlier MHV and SARS-CoV-2 cryo-ET maps, this dodecameric
306 arrangement appeared to be consistent with the overall diameter of the inner membrane pore
307 and provided the first hint as to how these pores may form. Once the detailed pore structure of
308 Huang, et al became available, the arrangement of the final fully cleaved pore was clear, and
309 our attention turned to whether the uncleaved model was consistent with this structure.

310 As described above, the nsp4 dodecamer component of the pore structure can be described as
311 an inner hexameric ring interlaced with an outer hexameric ring. Based on the locations of the
312 C-terminal domains, a clear requirement for this structure to form is that the inner nsp4
313 hexamer must be cleaved from nsp5. However, it's equally clear that the outer hexameric ring
314 can remain part of the uncleaved polyprotein and retain the integrity of the pore structure.
315 Thus, maturation of the pore within the framework of this model would at minimum require
316 the cleavage of alternating nsp4/nsp5 sites.

317 This presented an interesting topological question and an opportunity to reverse engineer the
318 process. In the nsp4-nsp5-nsp6 hexamers, we saw a domain swapping occur in which TMD2
319 adopts an open conformation with the α1 helix jutting out at a significant angle (~60°) to

320 interact with a neighboring TMD2 bundle of α2-α5. This open conformation (see Figure S14a) is
321 also seen in both the inner and outer hexamers of the structure solved by Huang, et al., where
322 domain swapping is clearly observed in the inner hexamer. Furthermore, examining an
323 interlaced pair of nsp4 subunits from this structure, the open conformation has the effect of
324 projecting one subunit inward and the other outward, causing the two to cross each other
325 (Figure S14b).

326 We recognized that there was some ambiguity as to how this may manifest itself when
327 converting the AlphaFold predicted nsp4-nsp5-nsp6 hexamer to a dodecamer. We considered
328 three possibilities: one in which all twelve subunits adopt the open conformation, one in which
329 they all adopt the closed conformation and one in which only alternating subunits adopt the
330 open conformation (Figures S14c). The overall structure was unchanged except for whether the
331 α1 helix was open or closed. From these three possibilities, it could be seen that the structure
332 with all twelve subunits in the open conformation does not have a direct path to the interlaced
333 final pore structure upon nsp4/nsp5 cleavage. However, in an arrangement in which only
334 alternating subunits adopt the open conformation, upon cleavage of the nsp4-nsp5 linkages
335 from those same six subunits, the structure would have a simple pathway to adopt the mature
336 pore conformation (see the movie in the Supplementary Information). From this perspective,
337 the model shown in Figure 4b may be considered a hexameric ring of polyprotein dimers,
338 mirroring the situation for the nsp3 dodecamer.

339 This model was ultimately merged with the nsp4 lumenal domains of the Huang, et al. structure
340 and our rebuilt model of the nsp3 dodecamer. The nsp4 TM1 helices, missing from the cryo-ET

341 structure, were also added in parallel to the nsp3 TM1 helices, consistent with discernible, but
342 previously unmodelled, density in the maps.

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

344 Having created a model of dodecameric nsp4-nsp6, we considered modeling the remainder of
345 the uncleaved polyprotein. This extended model could be derived from either pp1a' (nsp4-
346 nsp10) or pp1ab' (nsp4-nsp16) or a combination of both. We considered that certain known
347 protein-protein interactions may be preserved in this situation, creating a degree of
348 preorganization that would facilitate efficient formation of the replication complex upon
349 polyprotein cleavage. From cryo-EM and X-ray structures, interactions between nsp7 and nsp8,
350 nsp7 and nsp12, nsp8 and nsp12, nsp8 and nsp13, nsp9 and nsp12, nsp10 and nsp14, and
351 nsp10 and nsp16 are well characterized(57-60), as is nsp15 hexamerization(61).

352 Working from a dimer of nsp4-nsp5-nsp6 extracted from the optimized dodecameric ring, it
353 appeared that extending both subunits to the full pp1ab' polyprotein would be too bulky and
354 provide only limited interactions between the subunits. In contrast, extending both to the
355 smaller pp1a' polyprotein would be well tolerated, but offer little in additional protein-protein
356 interactions. However, a heterodimer formed from pp1a' and pp1ab' offered the maximum
357 opportunity to incorporate the known protein-protein interactions.

358 Initial exploration of polyprotein flexibility was conducted with AlphaFold on multiple
359 constructs. The uncleaved protein dimers (eg nsp9-nsp10, nsp10-nsp12, etc) revealed
360 conformational flexibility between all pairs of proteins (Figure S15), driven by typically
361 unstructured linkers. This led to exploration with larger fragments, with particular attention to

362 forming interactions that are already known. We extended the initial nsp4-nsp5-nsp6 dimer to
363 include nsp7, consistent with a conformation seen in the AlphaFold prediction of monomeric
364 nsp4-nsp5-nsp6-nsp7. We then found we could extend one of the subunits to include nsp8 and
365 have that protein interact with nsp7 from the other subunit just as it does in the cryo-EM
366 structures of the replication complex (eg 6XEZ).(57) Furthermore, we found we could associate
367 nsp13 with this nsp7/nsp8 pair, again as seen in the replication complex structure (Figure S16a).
368 This immediately established constraints on where nsp9, nsp12, nsp14 and the second nsp8
369 would need to connect.

370 AlphaFold predictions of uncleaved nsp8-nsp9-nsp10-nsp12 led to one of the largest building
371 blocks of the polyprotein heterodimer. The predictions showed variability in the positioning of
372 nsp9 and nsp10, but consistency in the interaction between nsp8 and nsp12 (Figure S16b). This
373 mirrored the replication complex as well and put constraints on how this string of proteins from
374 pp1ab' would be placed relative to nsp7 and nsp13. As both nsp14 and nsp16 are known to
375 coordinate to nsp10, we further aimed to find a conformation that would allow both protein
376 pairs to be satisfied. We proposed that one of these proteins would coordinate *cis* to nsp10
377 within the larger pp1ab' polyprotein and the other would coordinate *trans* to the terminal
378 nsp10 of pp1a'. We examined both possibilities with respect to binding to the nsp8-nsp9-nsp10-
379 nsp12 polyprotein. We concluded that only coordination of nsp16 following existing X-ray
380 crystal structures (6W4H)(60) had the potential to satisfy the constraints of linking all the
381 subunits. We then coordinated nsp14 and a second subunit of nsp10 to nsp12, following our
382 previous model for the full hexameric replication complex (see Figure S16c).(31)

383 These two large complexes (nsp4-nsp5-nsp6-nsp7-nsp8/nsp4-nsp5-nsp6-nsp7/nsp13 and nsp8-
384 nsp9-nsp10-nsp12/nsp10/nsp14/nsp16) were reasonably consistent with each other. We found
385 they could be easily linked and optimized, connecting nsp7 to nsp8, nsp12 to nsp13 and nsp13
386 to nsp14. This meant only nsp9 was needed to link nsp8 and nsp10 of pp1a' and nsp15 was
387 needed to link nsp14 and nsp16 of pp1ab' (see Figure S16d). Protein-protein docking with
388 Piper(62) was used to explore initial placement of the final proteins. The most promising
389 structures were then optimized with constraints to minimize the linkage distances before finally
390 linking the completed polyproteins and doing a refinement of the full protein heterodimer. In
391 the case of the nsp14-nsp15 and nsp15-nsp16 linkage, significant freedom was given to nsp9-
392 nsp10/nsp16 to adopt a conformation that would facilitate the final linkage. The final model is
393 shown in Figure 4c. While the conformational space of the two polyproteins is enormous and
394 could not be exhaustively studied, in the end, we established that multiple known protein-
395 protein interactions could be preserved in the polyprotein heterodimer. These include: pp1a'-
396 nsp8/pp1ab'-nsp7; pp1a'-nsp8/pp1ab'-nsp13; pp1a'-nsp10/pp1ab'-nsp14; pp1ab'-nsp8/pp1ab'-
397 nsp12; and pp1ab'-nsp10/pp1ab'-nsp16. This heterodimer was then incorporated into the
398 dodecameric model where it went through an additional round of refinement to remove any
399 clashes between the various components (Figure 4d).

400 *SARS-CoV-2 polyprotein cleavage*

401 The close-packed arrangement of nsp5 in the polyprotein dodecameric ring establishes a
402 favorable environment for the protease to make the initial nsp4-nsp5 cleavage. A key
403 observation we made was that the nsp4-nsp5 linker (residues 488-510) was long enough that it
404 could conceivably connect nsp4 to either of two nsp5 subunits in the ring (at a distance of 35 Å

405 or 52 Å). As originally constructed from the AlphaFold hexamer, the connection was to the
406 nearer nsp5-nsp6 subunit, comparable to how the proteins are arranged in the monomeric
407 polyprotein. However, connecting nsp4 instead to the neighboring nsp5-nsp6 subunit allowed
408 the nsp4-nsp5 cleavage site to pass over the proximal nsp5 protease active site. A positional
409 shift of the nsp5 proteins would still be required for cleavage to occur, and this was
410 investigated for both subunits. We found that simply by shifting alternating nsp5 subunits while
411 retaining the positions of everything else, we could identify a conformation that was capable of
412 nsp4-nsp5 cleavage, consistent with available crystal structures of bound substrate.(63) As
413 depicted in Figures 5a and 5b, the shifted nsp5 subunit functions as the substrate, while the
414 nsp5 subunit which retained its original position functions as the protease. Given the
415 constraints imposed by the full $(pp1a'/pp1ab')_6$ model, a repositioning of nsp5 associated with
416 pp1ab' appears more likely than that of pp1a' suggesting the larger polyprotein is the initial
417 substrate for cleavage.

418 As previously described, the protease activity of nsp5 is known to be significantly enhanced
419 when the protein is in its canonical dimer form, in large part due to stabilization of the S1
420 pocket by the N-terminal amine of the paired subunit.(52) A striking observation from these
421 AlphaFold derived models was that the C-terminal domain of nsp4 forms an interaction with
422 nsp5 which appears to mimic this aspect of nsp5 dimerization (Figure 5c). Specifically, a
423 conserved nsp4 basic residue (K452) forms a salt bridge with the conserved nsp5 glutamic acid
424 (E666), replicating the interaction seen in the dimer between the same glutamic acid and the N-
425 terminus of the paired subunit. The similar interaction seen here suggests formation of the

426 pore prior to polyprotein cleavage both activates the nsp5 protease and positions the nsp4-
427 nsp5 substrate to carry out the first cleavage event.

428 This proposed arrangement of alternating positions of nsp5 in the dodecameric ring would lead
429 to the cleavage of half of the nsp4-nsp5 linkages. Once this occurs, the reduced constraints
430 allow the nsp3/nsp4 pore to adopt a conformation consistent with that of the cryo-ET
431 structure. That is, the six cleaved nsp4 proteins rotate inward to form the central hexameric
432 pore. With this reconfiguration, the six remaining uncleaved nsp4 subunits can rotate outward
433 to adopt the conformations seen in the cryo-ET structure. At the same time, the six partially
434 cleaved nsp5 subunits can now form canonical nsp5 dimers with the six uncleaved nsp5
435 subunits, as shown in Figures 5d and 5e. In this case, the protease site in the uncleaved nsp5
436 subunit would be considered activated by the N-terminal amine from the cleaved subunit, as
437 previously described. From this configuration, cleavage of additional linkages could take several
438 pathways, with nsp5-nsp6, nsp6-nsp7, and nsp8-nsp9 linkages all within immediate range
439 (Figure S17).

440 **Discussion**

441 The events that lead to nsp5 protease cleavage of the coronavirus ORF1 polyproteins pp1a and
442 pp1ab are at present ill-defined. Furthermore, there is currently no clear hypothesis for how
443 the resulting cleavage products, which constitute the replication machinery, are recruited to
444 the closed DMV replication organelle. Cleavage of nsp1, nsp2 and nsp3 from the polyprotein
445 occurs very quickly, facilitated by the nsp3 PL proteases(17). Cleavage of the remainder of the
446 polyprotein by nsp5 appears to be a slower process, possibly requiring dimer formation or

447 some other regulatory event to initiate(64). As part of an investigation into the structure of the
448 SARS-CoV-2 DMV nsp3/nsp4 pore, we were struck by a particular AlphaFold generated model of
449 a ring formed by a multimer of uncleaved nsp4-nsp5-nsp6 polyproteins. We considered the
450 possibility that such a structure may represent a precursor to formation of the DMV pore in an
451 infected cell. From this starting point, we generated a series of structural models outlining a
452 plausible path for DMV and pore formation stemming from the interaction of twelve nsp3
453 subunits with six pp1a' and six pp1ab' uncleaved polyproteins. We argue that it is the formation
454 of an immature pore from these components that initiates the protease activity of nsp5 and
455 creates a situation where the cleaved products would be positioned within the DMV prior to its
456 closure.

457 We made several key observations from this modeling effort. First, analysis of the nsp4-nsp5-
458 nsp6-nsp7 monomer made clear that *cis*-cleavage of the nsp4-nsp5 and nsp5-nsp6 linkages is
459 not possible, implying cleavage is facilitated by some interaction between polyproteins. Based
460 on a great deal of biochemical and crystallographic work on cleaved constructs,(18, 19, 21, 50,
461 52, 54) it is well established that nsp5 is significantly more active as a dimer. But further
462 analysis of the nsp4-nsp5-nsp6-nsp7 dimer indicated that the constraints imposed by the
463 flanking membrane binding proteins nsp4 and nsp6 effectively render the canonical nsp5 dimer
464 inactive within the context of the polyprotein. Such a dimer can only be considered active
465 following nsp4-nsp5 cleavage of at least one of the two subunits. This leads to the conclusion
466 that some alternative interaction between the polyproteins is responsible for the initial nsp4-
467 nsp5 cleavage event.

468 Second, the uncleaved polyproteins pp1a' and pp1ab' can form a favorable heterodimer in
469 which known protein-protein interactions are established. While pp1a' and pp1ab' homodimers
470 could also be imagined, neither would have the advantages of the complementarity seen in the
471 heterodimer. As driven by a conformational change in the lumenal domain of nsp4, the
472 pp1a'/pp1ab' dimer is capable of binding to either a single membrane or a double membrane.
473 When bound to the single membrane of the ER, the nsp4 N-terminus and CTD, nsp5, nsp7-
474 nsp10 of pp1a' and nsp7-nsp16 of pp1ab' are cytosolic, while only the nsp4 lumen domain is in
475 the ER interior. When bound to the double membrane of a DMV, the nsp4 N-terminus is the
476 only element remaining exposed to the cytosol. With the nsp4 lumen domain sitting in the
477 membrane gap, the bulk of the polyprotein dimer then resides in the DMV interior.

478 These heterodimers, when in the double membrane spanning conformation, are then capable
479 of forming a dodecameric ring comprised of alternating pp1a' and ppa1b' polyproteins. This
480 ring has a diameter comparable to that seen for the nsp4 dodecamer in the recent nsp3/nsp4
481 pore structure reported by Huang, et al.(14) and can associate with the nsp3 dodecamer in the
482 same manner. Thus, we propose that this $(nsp3)_{12}(pp1a')_6(pp1ab')_6$ complex represents an
483 immature form of the pore.

484 The polyprotein dodecameric ring leads to close-packing of nsp5 and the nsp4 CTD, providing
485 the necessary scaffolding to facilitate *trans*-cleavage of the nsp4-nsp5 linkage. We identified a
486 conformation within this framework in which the 23-residue nsp4-nsp5 linker of alternating
487 subunits engages the proximal nsp5 protease active site. Critically, we also noticed that a
488 conserved basic residue from the nsp4 C-terminal domain (K452) forms a salt bridge with the
489 nsp5 glutamate residue (E666) located in the protease active site S1 pocket. This interaction

490 mimics the situation in the nsp5 dimer, in which the same glutamate residue interacts with the
491 N-terminal amine of the other subunit. It is thought to be critical to stabilization of the protease
492 active site, explaining the significant enhancement of protease activity associated with
493 dimerization(54). Thus, we conclude that the formation of this immature pore activates the
494 nsp5 protease and sets it up for the initial nsp4-nsp5 cleavage.

495 Following this initial nsp4-nsp5 cleavage event, the six cleaved nsp4 proteins and the six still
496 uncleaved nsp4 proteins are free to adopt the conformations seen in the mature pore
497 structure. The reduced structural constraints also allow formation of canonical nsp5 dimers in
498 which one of the two protease sites would be activated. With these nsp5 dimers having
499 proximity to additional cleavage sites within the polyprotein, the remainder of the polyprotein
500 can be processed, leading to the final fully cleaved $(nsp3)_{12}(nsp4)_{12}$ pore.

501 With this model, we found that pore formation prior to nsp5 polyprotein processing is a
502 reasonable possibility and may in fact be the regulatory step needed to ensure proper
503 localization of the cleaved proteins that make up the replication machinery. A notable
504 observation of the full model presented here is that the stoichiometry is consistent with our
505 previous model for the hexameric replication complex(31). That model included 60+ subunits
506 centered around nsp15. Specifically, the model includes six subunits of nsp7, nsp12, nsp13,
507 nsp14, nsp15, and nsp16, and twelve subunits of nsp8 and nsp10, with up to six transiently
508 associated nsp9 subunits. This complex is most efficiently derived from six pp1a and six pp1ab
509 polyproteins, meaning the precursor presented here simultaneously produces one pore
510 structure and generates one full replication complex. When viewed together, the two models

511 imply a mechanism not only for DMV and pore formation, but for recruitment of the replicase
512 to the DMV.

513 As depicted in Figure 6, the following picture emerges from these models. Upon cleavage by the
514 PL proteases, nsp3 then associates with the nsp4 component of either pp1a' or pp1ab' on the
515 ER membrane (Figure 6a). The nsp3 3Ecto and nsp4 lumenal domains are in the ER lumen, with
516 the bulk of the proteins exposed to the cytosol. The association between nsp3 and nsp4 then
517 induces double membrane formation, with aggregation into a tetrameric unit
518 $(nsp3)_2$(pp1a')(pp1ab') following (Figure 6b). The exact sequence of events is unclear, but this
519 pathway seems likely given the immediate proximity between nsp3 and either pp1a' or pp1ab'
520 following PLpro cleavage and the observation that the nsp3/nsp4 interaction drives membrane
521 rearrangement.(55) Once the double membrane forms, the nsp3 3Ecto and the nsp4 lumen
522 domains are positioned in the gap between the two membranes, and the remainder of nsp3 is
523 on one side of the double membrane and nsp5, nsp7 and the rest of the polyprotein are on the
524 other side. Further aggregation of these tetrameric building blocks leads to the immature pore
525 and may facilitate the observed increased curvature of the double membrane (Figure 6c).
526 Formation of additional pores will ultimately yield a sealed, spherical DMV.

527 The picture of membrane remodeling presented here raises the question of the relative timing
528 of nsp5 polyprotein processing and DMV formation. We see no justification for a strict
529 requirement that a DMV be fully formed for cleavage to begin. Should cleavage occur prior to
530 DMV formation, the nsp products would be released into the cytosol. We note that in this
531 situation, cleaved nsp6 could migrate back to the ER surface where it would be available for
532 membrane zippering, a key step in forming the connectors for DMV clusters.(10) However,

533 should cleavage follow closure of the DMV, the resulting products would be trapped inside
534 (Figure 6d). This is a scenario that would likely follow formation of the final pore of the DMV,
535 suggesting each DMV contains at minimum one hexameric replication complex. Interestingly,
536 this process may ultimately become self-limiting, as free nsp5 dimers build up in the cytosol,
537 cleavage of new polyproteins at late stages may occur without the requirement of new pore
538 formation.

539 The proposed mechanism of regulated protease activation to ensure proper localization of the
540 viral proteins is similar to the well-established mechanism of HIV protease maturation.(65, 66)
541 In that case, the protease resides on the gag-pol polyprotein and requires dimerization to be
542 enzymatically active. While the cleaved protease rapidly dimerizes in a biochemical setting,
543 dimerization and protease activation of the polyprotein only occur within the immature viral
544 particle. This ensures that the products (which include the reverse transcriptase and the
545 integrase) are contained within the virion and not released into the cytosol.

546 The implications of this model require careful consideration. Within the protected environment
547 of the DMV, the proposed stoichiometry, pre-organization and high protein concentration (we
548 estimate at least 1-2 µM based on a DMV diameter of 250-350 nm(11, 12, 67)) would be ideal
549 for formation of a long-lived full hexameric replicase complex. As previously described, this
550 complex should ensure that replication, proofreading, capping and dsRNA unwinding are
551 coordinated and efficient processes.(31) Confirmation of the exact contents of DMVs has been
552 notoriously difficult, however. Multiple studies have shown that the DMVs contain RNA,
553 specifically identifying both dsRNA and newly synthesized RNA.(3, 7, 8, 11, 67) Identification of

554 viral protein components of the replicase within DMVs has been less conclusive, although
555 increasing evidence supports their association with DMVs.(7, 8)

556 In contrast, the less controlled environment outside of DMVs may mean the full complex does
557 not form or that complexes are transient. The minimally known replication competent complex
558 is formed from nsp12, nsp7 and two subunits of nsp8.(68, 69) Association with nsp9(58, 70) and
559 two subunits of nsp13(26, 57) have been documented as well, but characterization of other
560 viral protein interactions has proven more elusive. Indeed, the proposed dual role of nsp15
561 suggests a full hexameric complex forms within the DMV but likely does not in the cytosol.
562 Within the DMV, endonuclease activity would need to be regulated to prevent degradation of
563 the RNA. This would be the case for the hexameric replication complex, where nsp15 is
564 proposed to function primarily as scaffolding, with its endonuclease activity highly
565 curtailed.(31) However, in the cytosol, nsp15's endonuclease activity serves as an effective
566 defense against the immune response by degrading cytosolic RNA,(61, 71) suggesting the it
567 exists as a free hexamer and is not regulated by the larger replication complex.

568 In the absence of full complex formation, some functions, such as proofreading and mRNA
569 capping, might be compromised or lost entirely. Yet plenty of evidence remains that some
570 replication is occurring outside of the DMVs. At minimum, there is a temporal argument for
571 this, in that replication necessarily precedes DMV formation.(72) However, other lines of
572 evidence are consistent with a two-compartment model persisting throughout infection.
573 Notably, the groundbreaking work of Sawicki and Sawicki(73) established that coronavirus RNA
574 replication is dependent on continuous ORF1 translation and proteolytic processing. Further
575 work on several temperature sensitive (TS) mutations showed that synthesis of negative-sense

576 RNA was halted at higher temperatures but positive-sense RNA was not.(74) Donaldson, et
577 al.(74) showed that this was a result of inhibited nsp5 protease activity at the higher
578 temperatures. More recently, Schmidt, et al.(75) demonstrated that negative strand synthesis,
579 was dependent on UMPylated-nsp9 priming, facilitated by the host factor SND1. This did not
580 appear to be the case for positive strand synthesis. With no proposed mechanism for
581 recruitment of additional proteins to an already formed DMV, the dependence of negative-
582 sense RNA synthesis on host factors and the continuous generation of viral replicase proteins
583 (specifically nsp9) suggest transcription of the minus strand occurs exclusively in the cytosol. In
584 contrast, this model would be consistent with positive strand synthesis occurring in the cytosol
585 at early stages but predominantly occurring within the protected environment of DMVs at later
586 stages.

587 In summary, in asking the question whether the nsp3/nsp4 DMV pore forms prior to nsp5
588 cleavage of the polyprotein, we have developed a simple framework for understanding how the
589 replication complex is recruited to DMVs in coronavirus infected cells. Our model directly links
590 the dodecameric nsp3/nsp4 pore structure and the previously proposed 60+ subunit hexameric
591 replication complex to the same polyprotein intermediate, elucidating how the nsp5 protease is
592 regulated to ensure proper localization of the replicase in the DMV. From a close examination
593 of structural constraints within monomeric and multimeric forms of the polyprotein, aided in
594 large part by AlphaFold predictions, we conclude that formation of an immature pore from
595 twelve nsp3 subunits, and six subunits each of the intermediate polyproteins pp1a' and pp1ab'
596 initiates the protease activity of nsp5. Following cleavage of half of the nsp4-nsp5 linkages, the
597 mature nsp3/nsp4 pore can form, and protease cleavage of the remainder of the polyproteins

598 will release the components necessary to form the full replication complex into the DMV. While

599 the detailed mechanism is specific to coronaviruses, we expect similar processes are taking

600 place with other single-stranded (+) RNA viral families.

601

602 **Materials and Methods**

603 AlphaFold2.1(76) (with a template cutoff date of 2/15/2021) was used to predict the structures
604 of individual monomers of nsp3, nsp4, and uncleaved nsp4-nsp5-nsp6-nsp7 from SARS-CoV-2,
605 as well as hexameric complexes of the C-terminal Y1/CoV-Y domain of nsp3, hexameric
606 complexes of the TMD2/CTD domains of nsp4 (residues 256-500), hexameric complexes of
607 uncleaved nsp4-nsp5-nsp6 (residues 256-1042), dimeric complexes of uncleaved nsp4-nsp5-
608 nsp6-nsp7, and various fragments and heterodimers of nsp3 and the pp1a' and pp1ab'
609 polyproteins. The SARS-CoV-2 sequences used are from the original Wuhan strain (NCBI
610 accession number QHD43415.1(77)). Nsp3 spans residues 819-2763 of the polyprotein, nsp4
611 spans 2764-3263, pp1a' (nsp4-nsp10) spans 2764-4392, and pp1ab' (nsp4-nsp16) spans 2764-
612 7096. The MHV sequence used in construction of the nsp3 dodecamer was from NCBI sequence
613 AAX23975.1.(78) Additional detail on the constructs explored can be found in the
614 Supplementary Information.

615 In all cases, AMBER relaxation was turned off and structures were instead refined with software
616 from the Schrödinger Suite, including Prime, Macromodel and Desmond(79). In all cases, the
617 OPLS4 force field was employed.(80) A typical workflow involved construction of an initial
618 model either from AlphaFold directly or via protein-protein docking with Piper(81) or manual
619 manipulation of individual subunits, followed by sidechain prediction in Prime. This was
620 followed by a stepwise minimization with Prime and/or Macromodel, in which residues at the
621 protein-protein interface were optimized first, building up to a minimization of the full
622 structure. In cases where large movements from the starting point occurred, a second round of
623 optimization would be carried out starting from a structure in which the original protein

624 subunits were re-aligned to their new positions. Stability was tested at several stages of model
625 development with molecular dynamics (MD) using Desmond, typically run for 100ns at constant
626 temperature and pressure (NPT) with suitable constraints. SPC water was used in these
627 simulations, as well as a POPC membrane, if appropriate. Macromodel and Desmond were also
628 used to predict conformations of some long loops. Construction of the pp1a'/pp1ab' dimer
629 involved preservation of known interactions as much as possible, using AlphaFold and Piper
630 protein-protein docking to explore possible arrangements of various subunits. Model quality
631 was assessed with Schrödinger's Protein Reports and by MolProbity.(82)

632 Models were guided by the cryo-ET maps deposited in the Electron Microscopy Data Bank with
633 accession codes EMD-11514 (MHV(12)), EMD-15963 (SARS-CoV-2(13)), and EMD-39107 and
634 EMD-39109 (SARS-CoV-2(14)). Ultimately, the models were merged with the structures of
635 Huang, et al. (PDB 8YAX and 8YB5)(14) and further refined using the approaches described
636 above. While the final pore models have approximate C6 rotational symmetry, strict symmetry
637 constraints were not imposed.

638 **Data Availibility**

639 All data needed to evaluate the conclusions of this work are present in the paper and/or the
640 Supplementary Materials. Structure models in CIF format are available upon request.

641 **References**

642 1. Wolff G, Barcena M. 2021. Multiscale Electron Microscopy for the Study of Viral Replication
643 Organelles. Viruses 13.
644 2. Wolff G, Melia CE, Snijder EJ, Barcena M. 2020. Double-Membrane Vesicles as Platforms for Viral
645 Replication. Trends Microbiol 28:1022-1033.
646 3. Snijder EJ, Limpens R, de Wilde AH, de Jong AWM, Zevenhoven-Dobbe JC, Maier HJ, Faas F,
647 Koster AJ, Barcena M. 2020. A unifying structural and functional model of the coronavirus
648 replication organelle: Tracking down RNA synthesis. PLoS Biol 18:e3000715.
649 4. Sergio MC, Ricciardi S, Guarino AM, Giaquinto L, De Matteis MA. 2024. Membrane remodeling
650 and trafficking piloted by SARS-CoV-2. Trends Cell Biol doi:10.1016/j.tcb.2023.12.006.
651 5. Roingeard P, Eymieux S, Burlaud-Gaillard J, Hourioux C, Patient R, Blanchard E. 2022. The
652 double-membrane vesicle (DMV): a virus-induced organelle dedicated to the replication of
653 SARS-CoV-2 and other positive-sense single-stranded RNA viruses. Cell Mol Life Sci 79:425.
654 6. Wang J, Han M, Roy AR, Wang H, Mockl L, Zeng L, Moerner WE, Qi LS. 2022. Multi-color super-
655 resolution imaging to study human coronavirus RNA during cellular infection. Cell Rep Methods
656 2:100170.
657 7. Chen A, Lupan AM, Quek RT, Stanciu SG, Asaftei M, Stanciu GA, Hardy KS, de Almeida Magalhaes
658 T, Silver PA, Mitchison TJ, Salic A. 2024. A coronaviral pore-replicase complex links RNA synthesis
659 and export from double-membrane vesicles. Sci Adv 10:eadq9580.
660 8. Andronov L, Han M, Zhu Y, Balaji A, Roy AR, Barentine AES, Patel P, Garhyan J, Qi LS, Moerner
661 WE. 2024. Nanoscale cellular organization of viral RNA and proteins in SARS-CoV-2 replication
662 organelles. Nat Commun 15:4644.
663 9. Angelini MM, Akhlaghpour M, Neuman BW, Buchmeier MJ. 2013. Severe acute respiratory
664 syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles.
665 mBio 4.
666 10. Ricciardi S, Guarino AM, Giaquinto L, Polishchuk EV, Santoro M, Di Tullio G, Wilson C, Panariello
667 F, Soares VC, Dias SSG, Santos JC, Souza TML, Fusco G, Viscardi M, Brandi S, Bozza PT, Polishchuk
668 RS, Venditti R, De Matteis MA. 2022. The role of NSP6 in the biogenesis of the SARS-CoV-2
669 replication organelle. Nature 606:761-768.
670 11. Knoops K, Kikkert M, Worm SH, Zevenhoven-Dobbe JC, van der Meer Y, Koster AJ, Mommaas
671 AM, Snijder EJ. 2008. SARS-coronavirus replication is supported by a reticulovesicular network of
672 modified endoplasmic reticulum. PLoS Biol 6:e226.
673 12. Wolff G, Limpens R, Zevenhoven-Dobbe JC, Laugks U, Zheng S, de Jong AWM, Koning RI, Agard
674 DA, Grunewald K, Koster AJ, Snijder EJ, Barcena M. 2020. A molecular pore spans the double
675 membrane of the coronavirus replication organelle. Science 369:1395-1398.
676 13. Zimmermann L, Zhao X, Makroczyova J, Wachsmuth-Melm M, Prasad V, Hensel Z,
677 Bartenschlager R, Chlanda P. 2023. SARS-CoV-2 nsp3 and nsp4 are minimal constituents of a
678 pore spanning replication organelle. Nat Commun 14:7894.
679 14. Huang Y, Wang T, Zhong L, Zhang W, Zhang Y, Yu X, Yuan S, Ni T. 2024. Molecular architecture of
680 coronavirus double-membrane vesicle pore complex. Nature 633:224-231.
681 15. Sawicki SG, Sawicki DL, Siddell SG. 2007. A contemporary view of coronavirus transcription. J
682 Virol 81:20-9.
683 16. Harcourt BH, Jukneliene D, Kanjanahaluethai A, Bechill J, Severson KM, Smith CM, Rota PA,
684 Baker SC. 2004. Identification of severe acute respiratory syndrome coronavirus replicase
685 products and characterization of papain-like protease activity. J Virol 78:13600-12.
686 17. Graham RL, Denison MR. 2006. Replication of murine hepatitis virus is regulated by papain-like
687 proteinase 1 processing of nonstructural proteins 1, 2, and 3. J Virol 80:11610-20.

688 18. Li C, Qi Y, Teng X, Yang Z, Wei P, Zhang C, Tan L, Zhou L, Liu Y, Lai L. 2010. Maturation
689 mechanism of severe acute respiratory syndrome (SARS) coronavirus 3C-like proteinase. J Biol
690 Chem 285:28134-40.
691 19. Chen S, Jonas F, Shen C, Hilgenfeld R. 2010. Liberation of SARS-CoV main protease from the viral
692 polyprotein: N-terminal autocleavage does not depend on the mature dimerization mode.
693 Protein Cell 1:59-74.
694 20. Roe MK, Junod NA, Young AR, Beachboard DC, Stobart CC. 2021. Targeting novel structural and
695 functional features of coronavirus protease nsp5 (3CL(pro), M(pro)) in the age of COVID-19. J
696 Gen Virol 102.
697 21. Noske GD, Nakamura AM, Gawriljuk VO, Fernandes RS, Lima GMA, Rosa HVD, Pereira HD, Zeri
698 ACM, Nascimento AFZ, Freire M, Fearon D, Douangamath A, von Delft F, Oliva G, Godoy AS.
699 2021. A Crystallographic Snapshot of SARS-CoV-2 Main Protease Maturation Process. J Mol Biol
700 433:167118.
701 22. Oudshoorn D, Rijs K, Limpens R, Groen K, Koster AJ, Snijder EJ, Kikkert M, Barcena M. 2017.
702 Expression and Cleavage of Middle East Respiratory Syndrome Coronavirus nsp3-4 Polyprotein
703 Induce the Formation of Double-Membrane Vesicles That Mimic Those Associated with
704 Coronaviral RNA Replication. mBio 8.
705 23. Malone B, Urakova N, Snijder EJ, Campbell EA. 2022. Structures and functions of coronavirus
706 replication-transcription complexes and their relevance for SARS-CoV-2 drug design. Nat Rev
707 Mol Cell Biol 23:21-39.
708 24. Hillen HS, Kokic G, Farnung L, Dienemann C, Tegunov D, Cramer P. 2020. Structure of replicating
709 SARS-CoV-2 polymerase. Nature 584:154-156.
710 25. Park GJ, Osinski A, Hernandez G, Eitson JL, Majumdar A, Tonelli M, Henzler-Wildman K,
711 Pawlowski K, Chen Z, Li Y, Schoggins JW, Tagliabracci VS. 2022. The mechanism of RNA capping
712 by SARS-CoV-2. Nature 609:793-800.
713 26. Chen J, Wang Q, Malone B, Llewellyn E, Pechersky Y, Maruthi K, Eng ET, Perry JK, Campbell EA,
714 Shaw DE, Darst SA. 2021. Ensemble cryo-electron microscopy reveals conformational states of
715 the SARS-CoV-2 helicase replication-transcription complex. To be published.
716 27. Malone B, Chen J, Wang Q, Llewellyn E, Choi YJ, Olinares PDB, Cao X, Hernandez C, Eng ET, Chait
717 BT, Shaw DE, Landick R, Darst SA, Campbell EA. 2021. Structural basis for backtracking by the
718 SARS-CoV-2 replication-transcription complex. Proc Natl Acad Sci U S A 118.
719 28. Bouvet M, Imbert I, Subissi L, Gluais L, Canard B, Decroly E. 2012. RNA 3'-end mismatch excision
720 by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14
721 exoribonuclease complex. Proc Natl Acad Sci U S A 109:9372-7.
722 29. Ogando NS, El Kazzi P, Zevenhoven-Dobbe JC, Bontes BW, Decombe A, Posthuma CC, Thiel V,
723 Canard B, Ferron F, Decroly E, Snijder EJ. 2021. Structure-function analysis of the nsp14 N7-
724 guanine methyltransferase reveals an essential role in Betacoronavirus replication. Proc Natl
725 Acad Sci U S A 118.
726 30. Decroly E, Debarnot C, Ferron F, Bouvet M, Coutard B, Imbert I, Gluais L, Papageorgiou N, Sharff
727 A, Bricogne G, Ortiz-Lombardia M, Lescar J, Canard B. 2011. Crystal structure and functional
728 analysis of the SARS-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. PLoS
729 Pathog 7:e1002059.
730 31. Perry JK, Appleby TC, Bilello JP, Feng JY, Schmitz U, Campbell EA. 2021. An atomistic model of
731 the coronavirus replication-transcription complex as a hexamer assembled around nsp15. J Biol
732 Chem 297:101218.
733 32. Mariani V, Biasini M, Barbato A, Schwede T. 2013. lDDT: a local superposition-free score for
734 comparing protein structures and models using distance difference tests. Bioinformatics
735 29:2722-8.

736 33. Stogios,P.J., et al., PDB 7KAG, unpublished, 2020.
737 34. Brosey CA, Houl JH, Katsonis P, Balapiti-Modarage LPF, Bommagani S, Arvai A, Moiani D, Bacolla
738 A, Link T, Warden LS, Lichtarge O, Jones DE, Ahmed Z, Tainer JA. 2021. Targeting SARS-CoV-2
739 Nsp3 macrodomain structure with insights from human poly(ADP-ribose) glycohydrolase (PARG)
740 structures with inhibitors. Prog Biophys Mol Biol 163:171-186.
741 35. 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.
742 2022. Two Binding Sites of SARS-CoV-2 Macrodomain 3 Probed by Oxaprozin and Meclomen. J
743 Med Chem 65:15227-15237.
744 36. Gallo, A., et al., PDB 7P2O, unpublished, 2022.
745 37. Calleja DJ, Kuchel N, Lu BGC, Birkinshaw RW, Klemm T, Doerflinger M, Cooney JP, Mackiewicz L,
746 Au AE, Yap YQ, Blackmore TR, Katneni K, Crighton E, Newman J, Jarman KE, Call MJ, Lechtenberg
747 BC, Czabotar PE, Pellegrini M, Charman SA, Lowes KN, Mitchell JP, Nachbur U, Lessene G,
748 Komander D. 2022. Insights Into Drug Repurposing, as Well as Specificity and Compound
749 Properties of Piperidine-Based SARS-CoV-2 PLpro Inhibitors. Front Chem 10:861209.
750 38. Stogios, P.J., et al., PDB 7LGO, unpublished, 2021.
751 39. Stogios, P.J., et al., PDB 7T9W, unpublished, 2021.
752 40. Li Y, Pustovalova Y, Shi W, Gorbatyuk O, Sreeramulu S, Schwalbe H, Hoch JC, Hao B. 2023. Crystal
753 structure of the CoV-Y domain of SARS-CoV-2 nonstructural protein 3. Sci Rep 13:2890.
754 41. Tan J, Vonrhein C, Smart OS, Bricogne G, Bollati M, Kusov Y, Hansen G, Mesters JR, Schmidt CL,
755 Hilgenfeld R. 2009. The SARS-unique domain (SUD) of SARS coronavirus contains two
756 macrodomains that bind G-quadruplexes. PLoS Pathog 5:e1000428.
757 42. Lei J, Ma-Lauer Y, Han Y, Thoms M, Buschauer R, Jores J, Thiel V, Beckmann R, Deng W,
758 Leonhardt H, Hilgenfeld R, von Brunn A. 2021. The SARS-unique domain (SUD) of SARS-CoV and
759 SARS-CoV-2 interacts with human Paip1 to enhance viral RNA translation. EMBO J 40:e102277.
760 43. Yan W, Zheng Y, Zeng X, He B, Cheng W. 2022. Structural biology of SARS-CoV-2: open the door
761 for novel therapies. Signal Transduct Target Ther 7:26.
762 44. Sparks JS, Lu X, Denison MR. 2007. Genetic analysis of Murine hepatitis virus nsp4 in virus
763 replication. J Virol 81:12554-63.
764 45. Xu X, Lou Z, Ma Y, Chen X, Yang Z, Tong X, Zhao Q, Xu Y, Deng H, Bartlam M, Rao Z. 2009. Crystal
765 structure of the C-terminal cytoplasmic domain of non-structural protein 4 from mouse hepatitis
766 virus A59. PLoS One 4:e6217.
767 46. 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,
768 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,
769 Qin C, Shi Z, Jiang H, Rao Z, Yang H. 2020. Structure of M(pro) from SARS-CoV-2 and discovery of
770 its inhibitors. Nature 582:289-293.
771 47. Zhang C, Li L, He J, Chen C, Su D. 2021. Nonstructural protein 7 and 8 complexes of SARS-CoV-2.
772 Protein Sci 30:873-881.
773 48. Oostra M, te Lintelo EG, Deijs M, Verheije MH, Rottier PJ, de Haan CA. 2007. Localization and
774 membrane topology of coronavirus nonstructural protein 4: involvement of the early secretory
775 pathway in replication. J Virol 81:12323-36.
776 49. Klatte N, Shields DC, Agoni C. 2022. Modelling the Transitioning of SARS-CoV-2 nsp3 and nsp4
777 Lumenal Regions towards a More Stable State on Complex Formation. Int J Mol Sci 24.
778 50. Yang H, Yang M, Ding Y, Liu Y, Lou Z, Zhou Z, Sun L, Mo L, Ye S, Pang H, Gao GF, Anand K, Bartlam
779 M, Hilgenfeld R, Rao Z. 2003. The crystal structures of severe acute respiratory syndrome virus
780 main protease and its complex with an inhibitor. Proc Natl Acad Sci U S A 100:13190-5.
781 51. Tomar S, Johnston ML, St John SE, Osswald HL, Nyalapatla PR, Paul LN, Ghosh AK, Denison MR,
782 Mesecar AD. 2015. Ligand-induced Dimerization of Middle East Respiratory Syndrome (MERS)

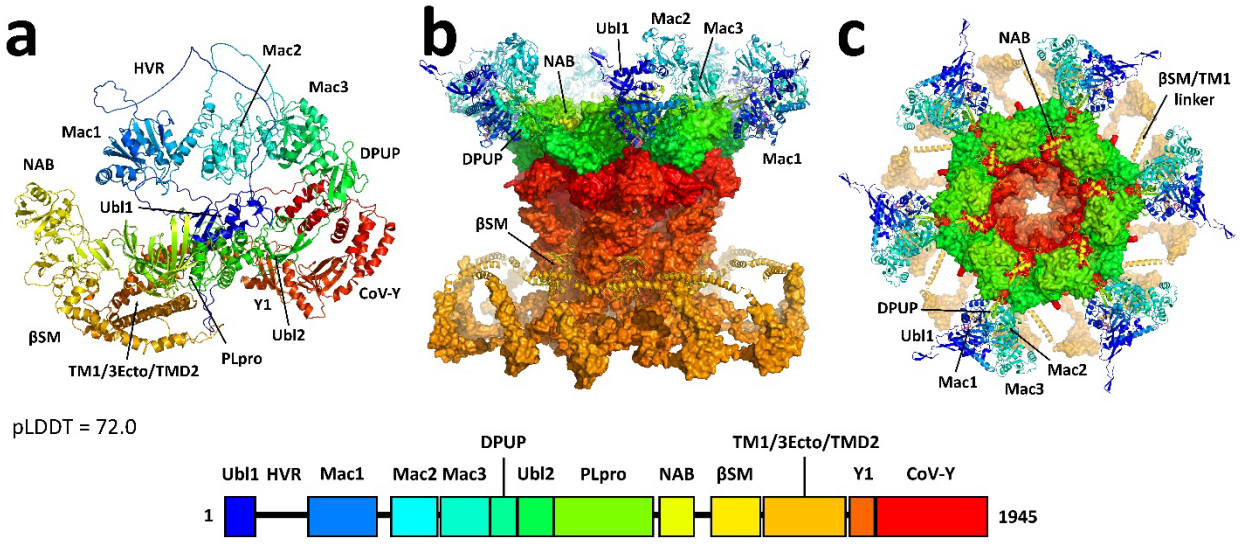
783 Coronavirus nsp5 Protease (3CLpro): IMPLICATIONS FOR nsp5 REGULATION AND THE
784 DEVELOPMENT OF ANTIVIRALS. J Biol Chem 290:19403-22.
785 52. Lee J, Worrall LJ, Vuckovic M, Rosell FI, Gentile F, Ton AT, Caveney NA, Ban F, Cherkasov A,
786 Paetzel M, Strynadka NCJ. 2020. Crystallographic structure of wild-type SARS-CoV-2 main
787 protease acyl-enzyme intermediate with physiological C-terminal autoprocessing site. Nat
788 Commun 11:5877.
789 53. Anand K, Ziebuhr J, Wadhwani P, Mesters JR, Hilgenfeld R. 2003. Coronavirus main proteinase
790 (3CLpro) structure: basis for design of anti-SARS drugs. Science 300:1763-7.
791 54. Chen S, Chen L, Tan J, Chen J, Du L, Sun T, Shen J, Chen K, Jiang H, Shen X. 2005. Severe acute
792 respiratory syndrome coronavirus 3C-like proteinase N terminus is indispensable for proteolytic
793 activity but not for enzyme dimerization. Biochemical and thermodynamic investigation in
794 conjunction with molecular dynamics simulations. J Biol Chem 280:164-73.
795 55. Hagemeijer MC, Monastyrska I, Griffith J, van der Sluijs P, Voortman J, van Bergen en
796 Henegouwen PM, Vonk AM, Rottier PJ, Reggiori F, de Haan CA. 2014. Membrane
797 rearrangements mediated by coronavirus nonstructural proteins 3 and 4. Virology 458-459:125-
798 35.
799 56. Jones R, Bragagnolo G, Arranz R, Reguera J. 2021. Capping pores of alphavirus nsP1 gate
800 membranous viral replication factories. Nature 589:615-619.
801 57. Chen J, Malone B, Llewellyn E, Grasso M, Shelton PMM, Olinares PDB, Maruthi K, Eng ET,
802 Vatandaslar H, Chait BT, Kapoor TM, Darst SA, Campbell EA. 2020. Structural Basis for Helicase-
803 Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. Cell 182:1560-1573
804 e13.
805 58. 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,
806 Chen Y, Guddat LW, Wang Q, Rao Z, Lou Z. 2021. Cryo-EM Structure of an Extended SARS-CoV-2
807 Replication and Transcription Complex Reveals an Intermediate State in Cap Synthesis. Cell
808 184:184-193 e10.
809 59. Liu C, Shi W, Becker ST, Schatz DG, Liu B, Yang Y. 2021. Structural basis of mismatch recognition
810 by a SARS-CoV-2 proofreading enzyme. Science doi:10.1126/science.abi9310.
811 60. Rosas-Lemus M, Minasov G, Shuvalova L, Inniss NL, Kiryukhina O, Brunzelle J, Satchell KJF. 2020.
812 High-resolution structures of the SARS-CoV-2 2'-O-methyltransferase reveal strategies for
813 structure-based inhibitor design. Sci Signal 13.
814 61. Frazier MN, Wilson IM, Krahn JM, Butay KJ, Dillard LB, Borgnia MJ, Stanley RE. 2022. Flipped
815 over U: structural basis for dsRNA cleavage by the SARS-CoV-2 endoribonuclease. Nucleic Acids
816 Res 50:8290-8301.
817 62. Schrödinger Release 2025-2: PIPER, Schrödinger, LLC, New York, NY, 2025.
818 63. Shaqra AM, Zvornicanin SN, Huang QYJ, Lockbaum GJ, Knapp M, Tandeske L, Bakan DT, Flynn J,
819 Bolon DNA, Moquin S, Dovala D, Kurt Yilmaz N, Schiffer CA. 2022. Defining the substrate
820 envelope of SARS-CoV-2 main protease to predict and avoid drug resistance. Nat Commun
821 13:3556.
822 64. Kaptan S, Girych M, Enkavi G, Kulig W, Sharma V, Vuorio J, Rog T, Vattulainen I. 2022.
823 Maturation of the SARS-CoV-2 virus is regulated by dimerization of its main protease. Comput
824 Struct Biotechnol J 20:3336-3346.
825 65. Tabler CO, Wegman SJ, Chen J, Shroff H, Alhusaini N, Tilton JC. 2022. The HIV-1 Viral Protease Is
826 Activated during Assembly and Budding Prior to Particle Release. J Virol 96:e0219821.
827 66. Harrison J, Passos DO, Bruhn JF, Bauman JD, Tuberty L, DeStefano JJ, Ruiz FX, Lyumkis D, Arnold
828 E. 2022. Cryo-EM structure of the HIV-1 Pol polyprotein provides insights into virion maturation.
829 Sci Adv 8:eabn9874.

830 67. Klein S, Cortese M, Winter SL, Wachsmuth-Melm M, Neufeldt CJ, Cerikan B, Stanifer ML, Boulant
831 S, Bartenschlager R, Chlanda P. 2020. SARS-CoV-2 structure and replication characterized by in
832 situ cryo-electron tomography. Nat Commun 11:5885.
833 68. Subissi L, Posthuma CC, Collet A, Zevenhoven-Dobbe JC, Gorbalenya AE, Decroly E, Snijder EJ,
834 Canard B, Imbert I. 2014. One severe acute respiratory syndrome coronavirus protein complex
835 integrates processive RNA polymerase and exonuclease activities. Proc Natl Acad Sci U S A
836 111:E3900-9.
837 69. Kirchdoerfer RN, Ward AB. 2019. Structure of the SARS-CoV nsp12 polymerase bound to nsp7
838 and nsp8 co-factors. Nat Commun 10:2342.
839 70. Slanina H, Madhugiri R, Bylapudi G, Schultheiss K, Karl N, Gulyaeva A, Gorbalenya AE, Linne U,
840 Ziebuhr J. 2021. Coronavirus replication-transcription complex: Vital and selective NMPylation of
841 a conserved site in nsp9 by the NiRAN-RdRp subunit. Proc Natl Acad Sci U S A 118.
842 71. Deng X, Hackbart M, Mettelman RC, O'Brien A, Mielech AM, Yi G, Kao CC, Baker SC. 2017.
843 Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in
844 macrophages. Proc Natl Acad Sci U S A 114:E4251-E4260.
845 72. Pathak R, Eliscovich C, Mena I, Cupic A, Rutkowska M, Chandran K, Jangra RK, Garcia-Sastre A,
846 Singer RH, Kalpana GV. 2024. Visualization of Early RNA Replication Kinetics of SARS-CoV-2 by
847 Using Single Molecule RNA-FISH Combined with Immunofluorescence. Viruses 16.
848 73. Sawicki SG, Sawicki DL. 1986. Coronavirus minus-strand RNA synthesis and effect of
849 cycloheximide on coronavirus RNA synthesis. J Virol 57:328-34.
850 74. Donaldson EF, Graham RL, Sims AC, Denison MR, Baric RS. 2007. Analysis of murine hepatitis
851 virus strain A59 temperature-sensitive mutant TS-LA6 suggests that nsp10 plays a critical role in
852 polyprotein processing. J Virol 81:7086-98.
853 75. Schmidt N, Ganskih S, Wei Y, Gabel A, Zielinski S, Keshishian H, Lareau CA, Zimmermann L,
854 Makroczyova J, Pearce C, Krey K, Hennig T, Stegmaier S, Moyon L, Horlacher M, Werner S, Aydin
855 J, Olguin-Nava M, Potabattula R, Kibe A, Dolken L, Smyth RP, Caliskan N, Marsico A, Krempl C,
856 Bodem J, Pichlmair A, Carr SA, Chlanda P, Erhard F, Munschauer M. 2023. SND1 binds SARS-CoV-
857 2 negative-sense RNA and promotes viral RNA synthesis through NSP9. Cell 186:4834-4850 e23.
858 76. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R,
859 Zidek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B,
860 Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M,
861 Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli
862 P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. Nature
863 596:583-589.
864 77. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang
865 YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. 2020. A new coronavirus
866 associated with human respiratory disease in China. Nature 579:265-269.
867 78. Sperry SM, Kazi L, Graham RL, Baric RS, Weiss SR, Denison MR. 2005. Single-amino-acid
868 substitutions in open reading frame (ORF) 1b-nsp14 and ORF 2a proteins of the coronavirus
869 mouse hepatitis virus are attenuating in mice. J Virol 79:3391-400.
870 79. Schrödinger Release 2023-1: Schrödinger, LLC, New York, NY, 2023.
871 80. Lu C, Wu C, Ghoreishi D, Chen W, Wang L, Damm W, Ross GA, Dahlgren MK, Russell E, Von
872 Bargen CD, Abel R, Friesner RA, Harder ED. 2021. OPLS4: Improving Force Field Accuracy on
873 Challenging Regimes of Chemical Space. J Chem Theory Comput 17:4291-4300.
874 81. Kozakov D, Brenke R, Comeau SR, Vajda S. 2006. PIPER: an FFT-based protein docking program
875 with pairwise potentials. Proteins 65:392-406.

876 82. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, 3rd,
877 Snoeyink J, Richardson JS, Richardson DC. 2007. MolProbity: all-atom contacts and structure
878 validation for proteins and nucleic acids. Nucleic Acids Res 35:W375-83.

879

880 **Figures**



881

882 **Figure 1.** a) AlphaFold predicted structure of monomeric SARS-CoV-2 nsp3. The domain
883 organization of nsp3 is indicated below, colored blue to red across its 1945 residues. The nsp3
884 structure was largely well predicted, with a few exceptions, including a long, disordered loop
885 corresponding to the highly variable region (HVR) and a mis-folded Mac2 domain, as discussed
886 in the text. The average pLDDT score was 72.0. b) Side and c) top views of the remodeled nsp3
887 dodecameric component of the DMV pore, based on the cryo-ET structure and maps of Huang,
888 et al.(14) High resolution portions of the structure were preserved from the original fit and
889 rendered here with surface representation. Low resolution portions of the structure were
890 remodeled, introducing several domains that were missing from the original fit. The newly
891 modeled domains are labeled and shown in ribbon representation here. Ubl: Ubiquitin-like
892 domain. HVR: Highly variable region. Mac: Macrodomain. DPUP: Domain preceding Ubl2 and
893 PLpro. PLpro: Papain-like protease. NAB: Nucleic acid binding domain. βSM: Betacoronavirus
894 specific marker.

895

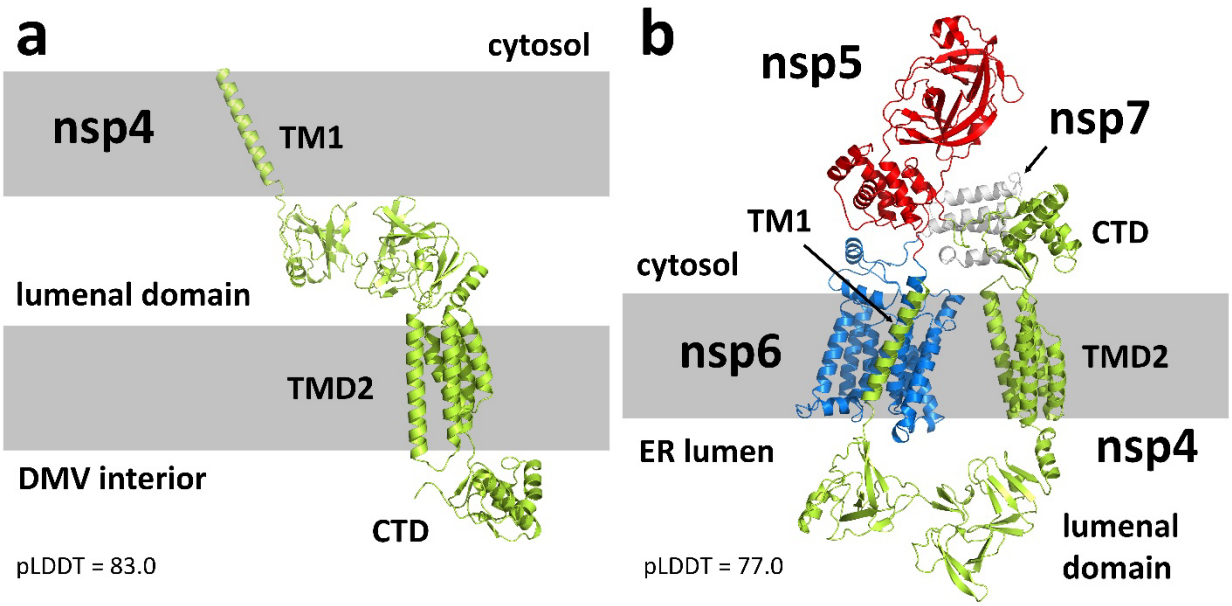


Figure 2. a) AlphaFold predicted structure of nsp4. The model reflects a conformation consistent with association to a double membrane (indicated by the gray bars), where the lumenal domain resides in the membrane gap. b) Predicted structure of the uncleaved nsp4-nsp7 polyprotein, with nsp4 in lime green, nsp5 in red, nsp6 in marine blue and nsp7 in white. The model adopts a conformation consistent with binding to the ER membrane, where the nsp4 N-terminus and CTD, nsp5 and nsp7 are cytosolic and the nsp4 lumenal domain is in the ER lumen. A conformational change between the N-lobe and C-lobe of the nsp4 lumenal domain determines whether the protein binds to a single membrane or a double membrane.

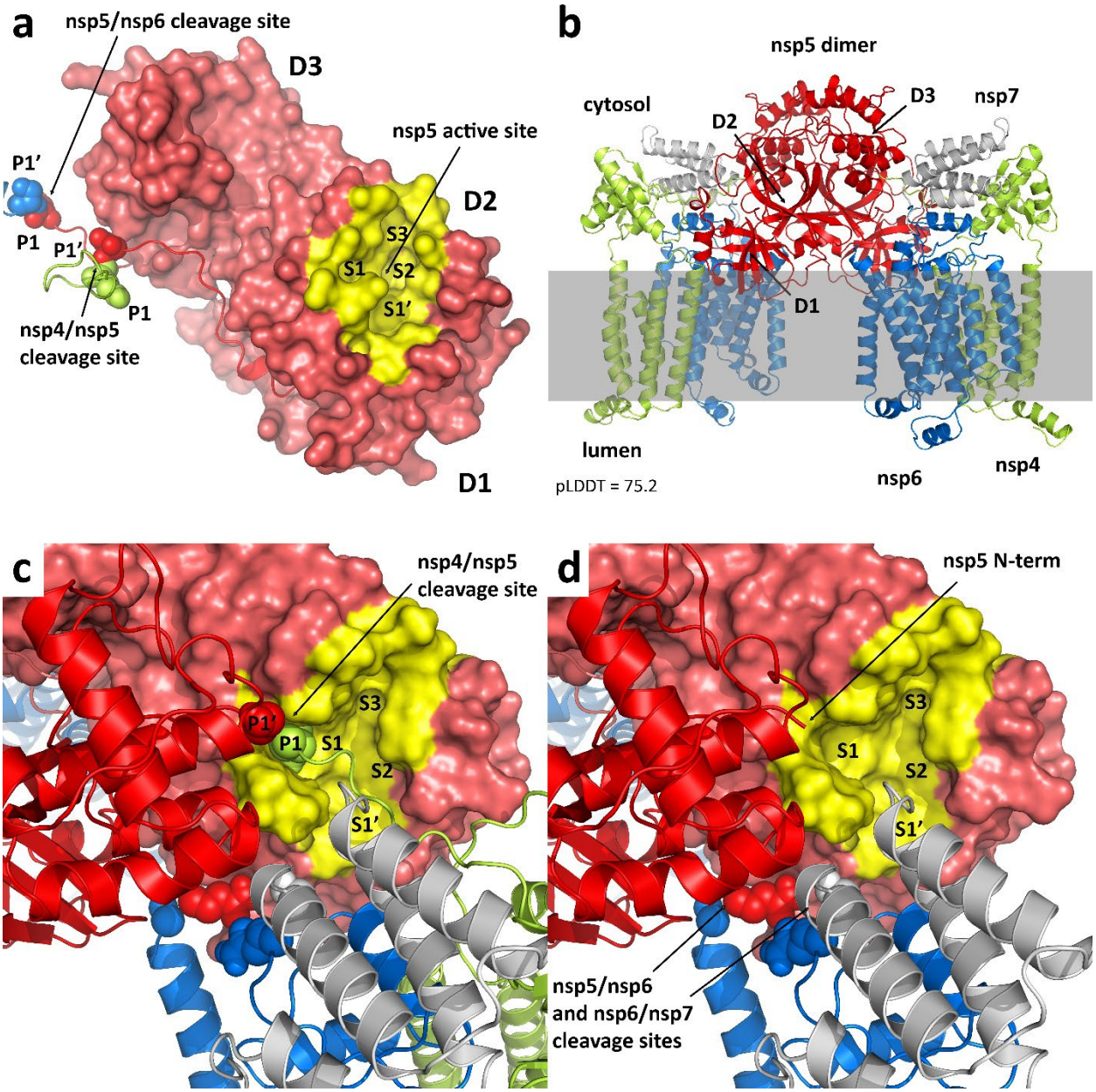
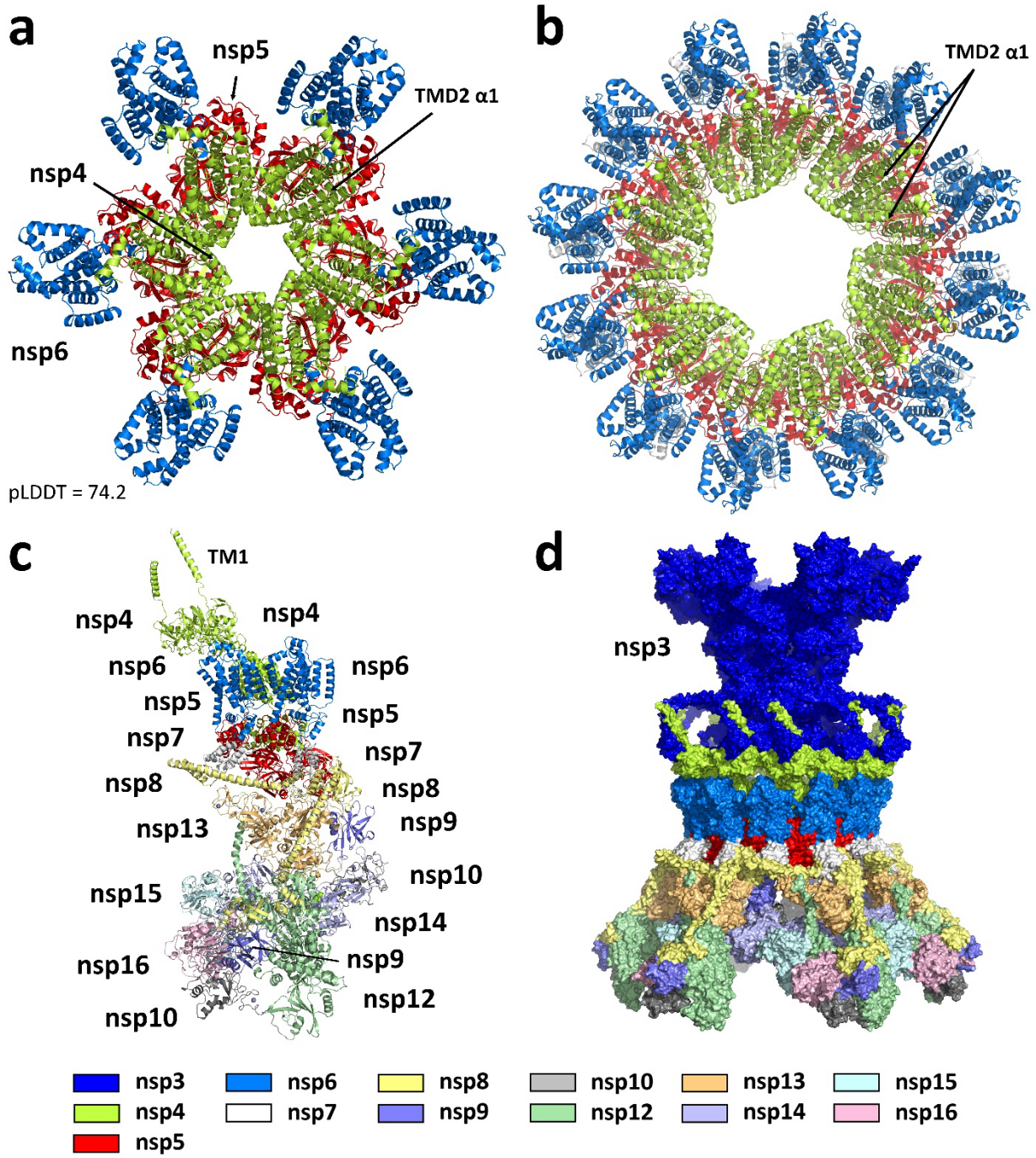


Figure 3. a) Location of the nsp4/nsp5 and nsp5/nsp6 cleavage sites on a single subunit relative to the nsp5 protease active site. The nsp4/nsp5 cleavage site (between residues Q500 and S501) sits on a loop that is not long enough to reasonably engage the active site. This is also true of the nsp5/nsp6 cleavage site. We can conclude that *cis*-cleavage is not a viable pathway. The three nsp5 domains are labeled D1, D2 and D3. The protease substrate recognition pockets

911 are labeled S1’, S1, S2 and S3. The peptide cleavage sites are labeled P1’ and P1. b) AlphaFold
912 prediction of the nsp4-nsp5-nsp6-nsp7 dimer. The predicted structure is anchored by a
913 canonical nsp5 dimer and associated with a single membrane through nsp4 and nsp6. It has an
914 average pLDDT score of 75.2. c) Detail of the nsp5 active site in the predicted uncleaved dimer.
915 The nsp4/nsp5 cleavage site engages with the nsp5 active site from the other subunit, but not
916 in a productive conformation. This effectively renders the protease inactive and implies the
917 initial cleavage event must proceed through a different mechanism. d) Assuming nsp4/nsp5
918 cleavage has already occurred, the protease site becomes active, potentially capable of cleaving
919 nsp5/nsp6 or nsp6/nsp7. The nsp5 N-terminal amine has previously been implicated in the
920 increased activity of the dimer vs. the monomer.

921



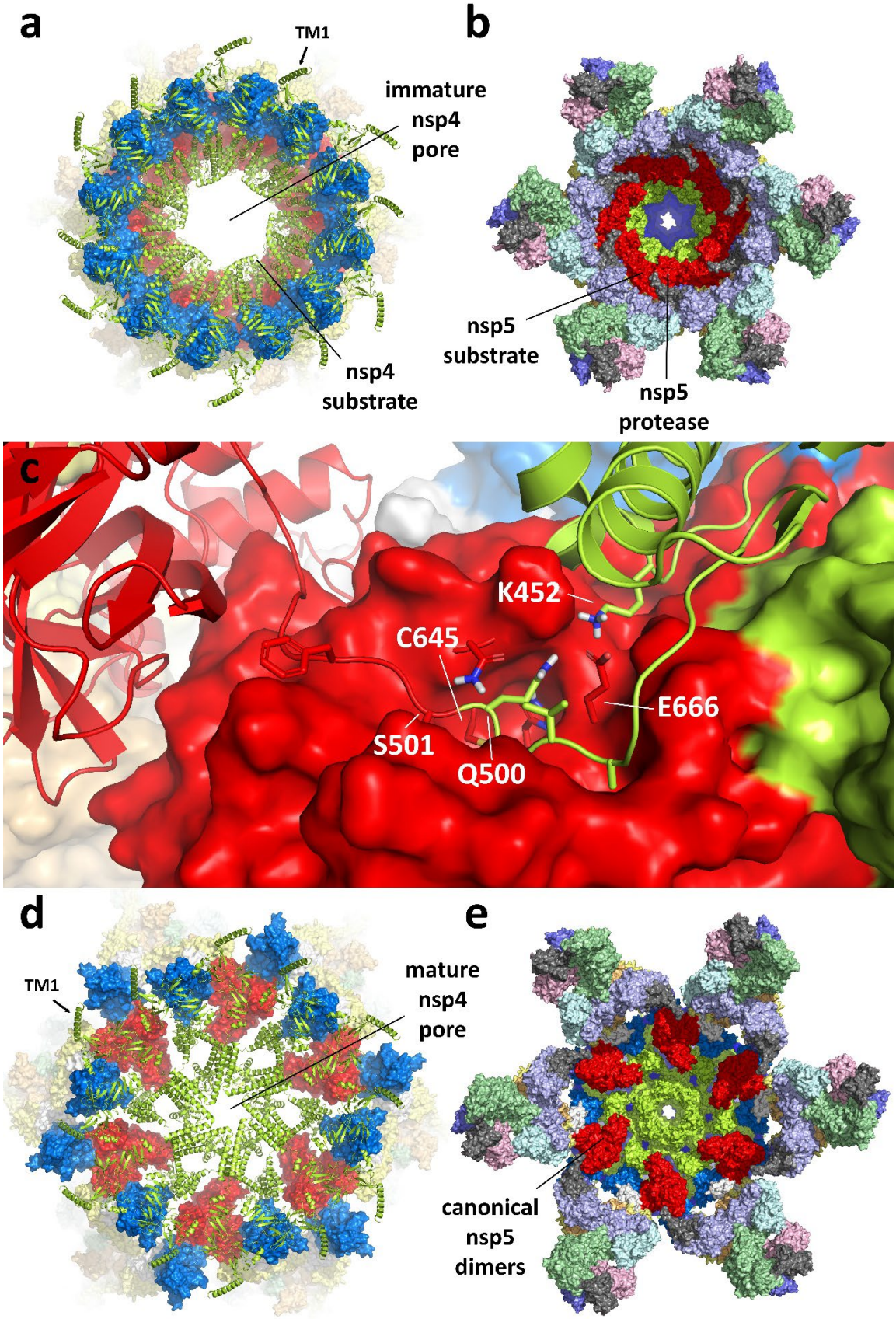
922

923 **Figure 4.** a) AlphaFold prediction of the hexameric SARS-CoV-2 nsp4-nsp6 polyprotein (a

924 truncated construct covering residues 256-1042). Nsp4 (lime green) forms a pore-like interior,

925 flanked by nsp6 (marine blue). Nsp5 (red) forms a close packed ring-like arrangement on the

926 underside. b) The nsp4-nsp6 hexamer transformed into a dodecamer. The nsp4 TMD2
927 alternates between open and closed conformations, facilitating pore maturation upon nsp4-
928 nsp5 cleavage. c) The pp1a'/pp1ab' heterodimer. The dimer incorporates known interactions
929 between the nsp proteins to the maximum extent. The nsp4 conformation is consistent with
930 coordination to a double membrane and formation of the dodecameric ring. d) Model of the
931 complete DMV immature pore, prior to nsp4-nsp5 cleavage. The precursor to the pore is
932 formed from twelve subunits of nsp3 (dark blue) and six subunits each of uncleaved pp1a' and
933 pp1ab'.


a
TM1
immature
nsp4
pore
nsp4
substrate
b
nsp5
substrate
nsp5
protease
c
K452
C645
E666
S501
Q500
d
TM1
mature
nsp4
pore
e
canonical
nsp5
dimers

934

935

936 **Figure 5.** Proposed mechanism for nsp5 protease activation and pore maturation. a) View from
937 above and b) view from below the immature pore in its pre-cleavage state. Six alternating nsp5
938 subunits within the dodecamer formed from pp1a' and pp1ab' assume a conformation which
939 allows *trans*-cleavage of the nsp4-nsp5 linker by the remaining nsp5 subunits. c) Detail of the
940 nsp5 protease active site shows cleavage occurs between Q500 on nsp4 and S501 on nsp5
941 (numbering consistent with the uncleaved polyprotein pp1ab'). The interaction between nsp4
942 K452 and nsp5 E666 mimics the activating interaction between the nsp5 N-terminus and the
943 same glutamic acid residue in the canonical nsp5 dimer. d) View from above and e) view from
944 below the post-cleavage mature pore. The reduced structural constraints from the nsp4-nsp5
945 cleavage of ppa1b' allows the nsp4 subunits to adopt the conformation seen in the structure by
946 Huang, et al.(14) Despite pp1a' remaining uncleaved at this point, canonical nsp5 dimers can
947 form in which one of the two protease active sites should be considered active and well
948 positioned to cleave additional linkages.

949

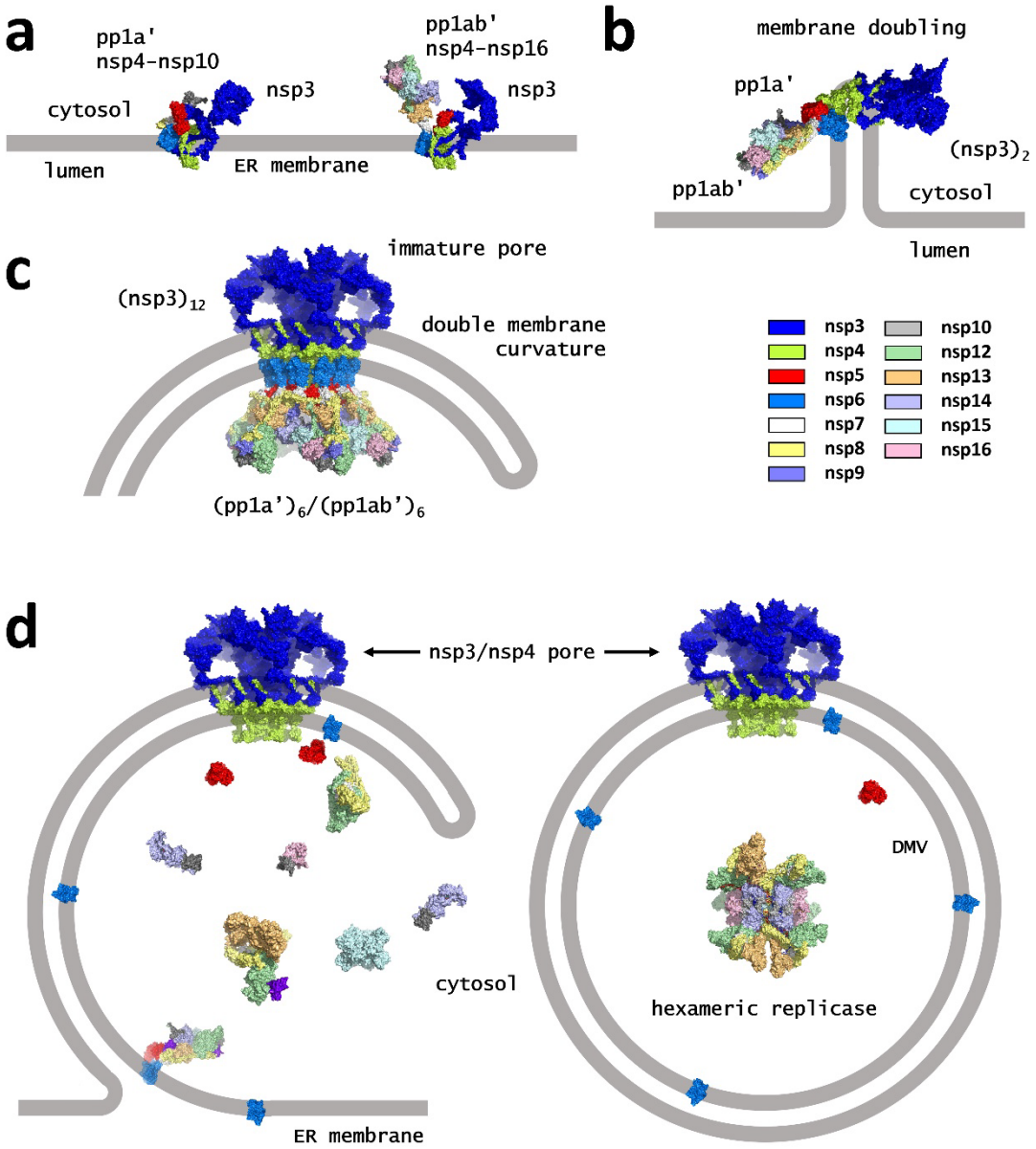


Figure 6. Schematic of pore formation, membrane remodeling and polyprotein cleavage. a) Following protease cleavage by PLpro, nsp3 and the polyprotein intermediates pp1a' (nsp4-nsp10) and ppa1b' (nsp4-nsp16) are expressed on the ER membrane. Given their initial proximity, nsp3/pp1a' and nsp3/pp1ab' dimers may form first. b) Interaction between nsp3 and nsp4 leads to a conformational change in the nsp4 lumen domain which drives membrane

956 doubling, while association between pp1a' and pp1ab' creates the basic $(nsp3)_2$(pp1a')(pp1ab')
957 building block for construction of the pore. c) Further aggregation leads to formation of the
958 dodecameric pore precursor and membrane curvature. The close packing of nsp5 in this
959 complex activates its protease. d) Cleavage of the polyprotein prior to DMV closure releases
960 viral proteins into the cytosol (left). Cleavage following DMV closure traps the viral proteins
961 within the DMV, creating ideal conditions for the previously predicted hexameric replicase to
962 form.(31)