



**HAL**  
open science

# De novo modelling of HEV replication polyprotein: Five-domain breakdown and involvement of flexibility in functional regulation

Sonia Fieulaine, Thibault Tubiana, Stéphane Bressanelli

## ► To cite this version:

Sonia Fieulaine, Thibault Tubiana, Stéphane Bressanelli. De novo modelling of HEV replication polyprotein: Five-domain breakdown and involvement of flexibility in functional regulation. *Virology*, 2023, 578, pp.128-140. 10.1016/j.virol.2022.12.002 . hal-04023059

**HAL Id: hal-04023059**

**<https://hal.science/hal-04023059v1>**

Submitted on 10 Mar 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 ***De novo* modelling of HEV replication polyprotein: Five-domain breakdown**  
2 **and involvement of flexibility in functional regulation**

3

4 **Sonia FIEULAINÉ\*, Thibault TUBIANA\*, Stéphane BRESSANELLI**

5

6 Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198, Gif-sur-  
7 Yvette, France

8

9 \* equal contributions

10

11 [sonia.fieulaine@i2bc.paris-saclay.fr](mailto:sonia.fieulaine@i2bc.paris-saclay.fr)

12 [thibault.tubiana@i2bc.paris-saclay.fr](mailto:thibault.tubiana@i2bc.paris-saclay.fr)

13 [stephane.bressanelli@i2bc.paris-saclay.fr](mailto:stephane.bressanelli@i2bc.paris-saclay.fr)

14

15 **CORRESPONDENCE**

16 Sonia Fieulaine & Stéphane Bressanelli

17 Email: [sonia.fieulaine@i2bc.paris-saclay.fr](mailto:sonia.fieulaine@i2bc.paris-saclay.fr) (SF); [stephane.bressanelli@i2bc.paris-saclay.fr](mailto:stephane.bressanelli@i2bc.paris-saclay.fr) (SB)

18

19 **CONFLICT OF INTEREST**

20 The authors declare no conflict of interest.

21

22 **FUNDING**

23 This research received specific grants from ANRS (ECTZ105819 and ECTZ188022) and a postdoctoral  
24 fellowship (ECTZ189696 to TT).

25

26 **ABSTRACT**

27 Hepatitis E virus (HEV), a major cause of acute viral hepatitis, is a single-stranded, positive-sense RNA  
28 virus. As such, it encodes a 1700-residue replication polyprotein pORF1 that directs synthesis of new  
29 viral RNA in infected cells. Here we report extensive modeling with AlphaFold2 of the full-length  
30 pORF1, and its production by *in vitro* translation. From this, we give a detailed update on the  
31 breakdown into domains of HEV pORF1. We also provide evidence that pORF1's N-terminal domain is  
32 likely to oligomerize to form a dodecameric pore, homologously to what has been described for  
33 Chikungunya virus. Beyond providing accurate folds for its five domains, our work highlights that there  
34 is no canonical protease encoded in pORF1 and that flexibility in several functionally important regions  
35 rather than proteolytic processing may serve to regulate HEV RNA synthesis.

36

37

38

39

40

41 **KEYWORDS**

42 Replication polyprotein pORF1

43 Single-stranded, positive-sense RNA virus

44 In vitro translation

45 AlphaFold2

46 Hepatitis E virus

47

48

49

50

51

52 **1. INTRODUCTION**

53 With 20 million cases per year, hepatitis E virus (HEV) is the major cause of human acute viral  
54 hepatitis worldwide. At least four genotypes (gt) affecting human have been identified <sup>1,2</sup>. Gt1 and gt2  
55 circulate in developing countries with poor levels of hygiene, especially in Asia. People are infected  
56 during epidemic waves through faecally contaminated drinking water. Gt3 and gt4 are found in  
57 developed countries where the virus is mainly transmitted through consumption of undercooked  
58 meat, essentially from pork products. HEV prevalence in industrialized countries is low and sporadic.  
59 However, constant identification of new genotypes and reservoir animals such as rat <sup>3</sup> may suggest a  
60 wider spread of the virus in the relatively short term, greatly increasing the number of people affected  
61 and potentially leading to new outbreaks.

62 HEV infection is usually asymptomatic or self-resolving for immunocompetent individuals.  
63 However, some infected people can occasionally develop a fatal fulminant hepatitis leading to 70,000  
64 deaths per year. The mortality rate is particularly high for pregnant women infected by gt1 in certain  
65 geographical areas in India <sup>4</sup>. Finally, extra-hepatic manifestations are occasionally reported <sup>5,6</sup>. One  
66 vaccine has been approved in China <sup>7</sup> and there is no specific treatment to cure HEV infection. Ribavirin,  
67 a broad-range non-specific antiviral drug, can be administered to patients with chronic hepatitis but  
68 not to pregnant women because of its teratogen effects <sup>8</sup>. According to World Health Organization,  
69 HEV is therefore a significant public health problem and a better characterization of its viral life cycle  
70 is needed to develop better vaccines or specific antiviral drugs. This is particularly true at a structural  
71 level because the knowledge of the three-dimensional structure of viral proteins both sheds light on  
72 the viral cycle and helps in the development of specific antiviral drugs, as witness the spectacular  
73 successes in hepatitis C virus and Coronavirus research <sup>9,10</sup>.

74 HEV is a single-stranded, positive-sense RNA (+RNA) virus belonging to the Orthohepevirus  
75 genus in Hepeviridae family <sup>1</sup>. Like most +RNA viruses, it encodes a multi-domain replication  
76 polyprotein in its first open reading frame named pORF1. In the case of HEV, pORF1 is composed of  
77 ~1,700 amino acid residues <sup>11</sup>. The major function of pORF1 is the synthesis of new viral genomes that

78 will be encapsidated before propagation to other cells. Indeed, pORF1 replication polyprotein is  
79 essential for HEV life cycle and is therefore an attractive and innovative therapeutic target.

80         30 years ago, Koonin *et al.* took advantage of phylogenetic relationships between HEV,  
81 Alphaviruses and other Alphavirus-like viruses such as Rubella virus (RUBV) <sup>12</sup> to tentatively propose  
82 by sequence comparisons the nature of non-structural proteins embedded in HEV pORF1 and the  
83 location of their core segments. Local sequence similarities led them to assign six putative functional  
84 domains: from N- to C-terminus, a methyltransferase (Met) domain, a Y domain of unknown function,  
85 a papain-like cysteine protease (PCP) domain, an X domain of unknown function, an RNA helicase (HEL)  
86 domain and an RNA-dependent RNA polymerase (RdRp) domain. In addition, they identified a  
87 hypervariable region (HVR) that contains a proline-rich region (PRR) between the putative PCP and the  
88 X domain <sup>13</sup>. Although similarities were detected with a limited number of sequences, this description  
89 of HEV pORF1 organization has remained unchanged since this seminal work. Yet, while Met, HEL and  
90 RdRp enzymatic functions have been experimentally confirmed and functions have been proposed for  
91 Y and X domains (reviewed in references <sup>14-17</sup>), the existence of the PCP domain was never established.  
92 Indeed, Koonin *et al.* already pointed out in 1992 that the sequence similarities in core segments were  
93 significant for Met, Y, X, HEL and RdRp, but not for PCP. They assigned a protease domain in HEV pORF1  
94 mainly because other animal Alphavirus-like viruses always encode a protease and despite a very low  
95 similarity, even in the putative catalytic region, with other viral or cellular proteases of any kind <sup>13</sup>.  
96 Experimentally, it could never be clearly established that HEV pORF1 is proteolytically processed in  
97 discrete functional units as it is the case for most replication polyproteins encoded by +RNA viruses <sup>15</sup>.  
98 Indeed, several studies concluded that HEV pORF1 is processed while others concluded it is not <sup>18-28</sup>.  
99 Of note, in all the aforementioned works, any pORF1 processing was at best minor, *i.e.* full-length  
100 pORF1 remained by far the most abundant species. However, in these previous works pORF1 was  
101 expressed in heterologous systems, not in the context of infection or at least with the complete HEV  
102 genome.. Finally, a recent study described the crystal structure of a fragment of a gt1 HEV pORF1  
103 (residues 510-691) <sup>29</sup> that overlaps the C-terminal sequence of the putative PCP domain (residues 434-

104 592 according to Koonin *et al.* <sup>13</sup>) and the N-terminal part of the hypervariable region (residues 593-  
105 783 <sup>13</sup>). This structure revealed that this portion, formerly assigned to two distinct parts of HEV pORF1,  
106 actually is a single globular domain that is not a protease but folds as a fatty acid binding domain (FABD)  
107 of unknown function <sup>29</sup>. Thus, there are discrepancies in the literature regarding the composition of  
108 the HEV replication polyprotein pORF1 and it is of high importance to collect structural information  
109 about it.

110 In this work, using the new AlphaFold2 *de novo* modelling tool, we have modeled the three-  
111 dimensional (3D) structure of the full-length HEV pORF1, for both gt1 and gt3 sequences. In addition,  
112 in order to study its structure by experimental methods, we have developed a protocol to express and  
113 purify the full-length HEV pORF1. Collectively, our results allow us to update the architecture of the  
114 HEV replication polyprotein with sufficient precision to rule out the presence of a canonical protease  
115 domain and propose that flexibility is a major form of functional regulation.

116

## 117 **2. MATERIALS AND METHODS**

### 118 **2.1 Molecular modeling using AlphaFold2**

119 We used inhouse ColabFold implementations <sup>30</sup> of AlphaFold2 <sup>31</sup> to generate models of pORF1  
120 from gt1 (Sar55 strain, GenBank accession no. AF444002, 1693 residues) or gt3 (Kernow C1-p6 strain,  
121 GenBank accession no. JQ679013, 1765 residues) sequences. The amino acid sequences were used to  
122 query the UniRef30 database (March or June 2021 release) with either the HHBlits web service <sup>32-34</sup> ,  
123 either with default parameters or with the maximum allowed number of iterations (8, called HHBlits  
124 MAX thereafter) or MMseqs2 <sup>35</sup>. These alignments were used as input to AlphaFold2 and five models  
125 generated for each. A total of 25 full-length pORF1 models were generated. The limits of domains were  
126 thus defined and the structures of individual domains were used as input for a final modelling of gt1  
127 HEV pORF1. Reference MetY and RdRp domains were shortened to reduce position constraint in the  
128 modelling of the domains with respect to each other (Table 1). Sterical clashes were then removed  
129 using the ISOLDE plugin <sup>36</sup> in ChimeraX <sup>37,38</sup>.

130

131 **Table 1: Reference structures and models for the final modelling of the gt1 HEV pORF1.** Models  
132 shortened with respect to domain boundaries (Figure 7) are marked with \*.

Domain	Reference structure	Reference range
MetY	MetY HHBlist MAX	1-447*
FABD-like	PDB 6NU9	7-182
X	X HHBlist MAX	778-926
HEL	HEL HHBLIST MAX	930-1207
RdRp	RdRp HHBLIST MAX	1238*-1693

133

134 For oligomer modelling, thirty 2-mers of the N-terminus of gt1 (residues 1-490 or 1-700) were  
135 modelled in six runs with different protocols (either "AlphaFold2-ptm" or "AlphaFold2-multimer" (v1))  
136 and different random seeds. Resulting models were curated by visual inspection and predicted  
137 alignment error (PAE) plots. The top-ranked 2-mer has a rotation angle of 32.89° and a translation  
138 along its transversal axis of -5.12Å (cf. Fig.S3B), resulting in a semi-helical oligomer with a pitch of  
139 55.99Å and 10.95 monomers per turn. Following the procedure depicted in <sup>39,40</sup>, Monte-Carlo  
140 simulation steps were used to optimize the rotation angle to 30° and the translation to 0° to create a  
141 cyclic dodecamer with a pitch of 0Å and 12 monomers per turn (Fig.2 and Fig.S3C). Alignments  
142 between several structures were made using TM-align <sup>41</sup> with its dedicated PyMoL Addon. All figures  
143 were rendered with Protein Imager <sup>42</sup> or ChimeraX <sup>37,38</sup>.

## 144 2.2 Cloning and plasmids preparation

145 The coding region of full-length HEV pORF1 was PCR-amplified and inserted via EcoRV and NotI  
146 restriction sites into the pEU-E01-MCS vector that contains a SP6 promotor (CellFree Sciences). We  
147 used pSK-HEV-2 (GenBank accession no. AF444002) <sup>43</sup> and pSK-p6-HEV (GenBank accession no.  
148 JQ679013) <sup>44</sup> plasmids as PCR templates, corresponding to the complete genome of ORF1 from Sar55  
149 strain of HEV genotype 1 and Kernow C1-p6 strain of HEV genotype 3, respectively. A Strep-tag II  
150 (WSHPQFEK) flanked by two small linkers (AS and TG in N- and C-terminal sides, respectively) was

151 inserted at the N-terminal extremity of both proteins (named s-ORF1). In a second construct, an 8His-  
152 tag was also fused to the C-terminal extremity of genotype 1 HEV s-ORF1, resulting in a double-tagged  
153 protein (named s-ORF1-h). The plasmids were amplified in NEB Turbo thermo-competent cells (New  
154 England Biolabs) and purified using a NucleoBond Xtra Maxi kit (Macherey-Nagel). In order to remove  
155 any RNase traces, plasmids were further purified by phenol/chloroform extraction in RNase-free  
156 conditions and resuspended in nuclease-free water.

157 The synthetic genes (Invitrogen, Thermo Fisher Scientific) coding for gt1 or gt3 HEV FABD-like  
158 domain (residues 510-691) flanked by an 8His-tag at their C-terminal extremity were inserted into  
159 pETM40 vector between NcoI and NotI restriction sites. Since this vector contains an N-terminal MBP  
160 coding sequence followed by a TEV cleavage site, the resulting plasmids encode TEV cleavable MBP –  
161 FABD-like domain fusion proteins. The plasmids were amplified in NEB Turbo thermo-competent cells  
162 (New England Biolabs) and purified using a NucleoSpin Plasmid kit (Macherey-Nagel).

163 All plasmids were verified by sequencing (GATC Biotech, Constance, Germany).

### 164 **2.3 Small-scale pORF1 expression by a wheat-germ cell-free expression system**

165 Genotype 1 and 3 HEV s-ORF1 were expressed in a wheat-germ cell-free expression system,  
166 with uncoupled transcription and translation steps as previously described <sup>45</sup> and adapted for viral  
167 replication polyproteins <sup>46</sup>. We used a home-made wheat-germ extract prepared from non-treated  
168 durum wheat seeds as previously described <sup>47</sup>. *In vitro* transcription and translation steps were  
169 performed in RNase-free conditions.

#### 170 **2.3.1 *In vitro* transcription**

171 A transcription mix composed of 100 ng/μL plasmid encoding the Strep-tagged protein (s-  
172 ORF1), 2.5mM rNTPs mix (Promega), 1 U/μL RNAsin (CellFree Sciences), 1 U/μL SP6 RNA polymerase  
173 (CellFree Sciences) in transcription buffer (80 mM HEPES-KOH pH 7.6, 16 mM magnesium acetate, 10  
174 mM DTT, and 2 mM spermidine; CellFree Sciences) was incubated 7 h at 37°C.

#### 175 **2.3.2 *In vitro* translation**

176 *In vitro* translation was performed using the bilayer method <sup>45</sup>. Optimization of protein  
177 expression was realized at small-scale in 96-well plates, using two wells per sample. The 200  $\mu$ L upper  
178 layer, composed of feeding buffer (30 mM Hepes-KOH, pH 7.6, 100 mM potassium acetate, 2.7 mM  
179 magnesium acetate, 16 mM creatine phosphate, 0.4 mM spermidine, 1.2 mM ATP, 0.25 mM GTP, 4  
180 mM DTT; CellFree Sciences) supplemented by 0.3 mM of amino acids mix (Cambridge Isotopes  
181 Laboratories), was deposited on each well. Then, the 20  $\mu$ L translation mixture, composed of 10  $\mu$ L of  
182 transcript and 10  $\mu$ L of the wheat-germ extract supplemented by 0.3 mM of amino acids mix  
183 (Cambridge Isotopes Laboratories) and 40  $\mu$ g/mL creatine kinase (Roche), was deposited under the  
184 upper layer to form the bottom layer. When translation was performed in the presence of detergent  
185 (Anatrace), it was added in both layers. The plate was covered with an adhesive film and incubated at  
186 22°C for 16 h without agitation.

### 187 **2.3.3 Sample treatment and analysis**

188 For each sample, the content of two wells called cell-free sample (CFS) was pooled and  
189 supplemented by home-made benzonase. After incubation at room temperature on a rotary wheel  
190 during 30 min, 350  $\mu$ L of CFS were centrifuged 30 min at 20,000 g at 4°C. The pellet (P) was resuspended  
191 in 350  $\mu$ L of washing buffer (50 mM Tris-HCl, 150 mM NaCl, pH 8.0). Samples kept at each step of the  
192 procedure and supplemented with denaturing loading buffer were separated by SDS-PAGE on 10 %  
193 acrylamide gels. Western-blot analysis was then performed by protein transfer on a nitrocellulose  
194 membrane. Membranes were incubated overnight at 4°C with Strep-MAB classics primary antibodies  
195 (IBA Lifesciences) diluted 5,000 times and then 3 h at room temperature with anti-mouse IgG HRP  
196 conjugate secondary antibodies (Promega) diluted 4,000 times. Membranes were analyzed using the  
197 enhanced chemiluminescence method using ECL Prime Western Blotting Detection Reagent (Cytivia)  
198 on a CCD camera (Gbox, Syngene).

### 199 **2.4 Large scale protein expression and purification of pORF1**

200 In order to purify the double-tagged protein (s-ORF1-h) from gt1 HEV, translation reaction was  
201 then up-scaled in four 6-well plates. In this case, we deposited a 5.5 mL upper layer and a 500  $\mu$ L

202 bottom layer on each well, using the same mixtures described for small-scale expression. Large-scale  
203 expression was done in presence of 0.1% GDN in both layers. The plates were covered and incubated  
204 at 22°C for 16 h without agitation.

205 The whole cell-free sample (144 mL) was supplemented with home-made benzonase and  
206 incubated 1 h at room temperature under agitation. The sample was then centrifuged 30 min at 20,000  
207 g at 4°C. The supernatant (144 mL) was incubated overnight with 3 mL of pre-equilibrated Strep-Tactin  
208 Superflow high capacity resin (IBA Lifesciences) at 4°C on a rolling wheel. The flow-through (FT) was  
209 collected and the resin was washed with 5 column volumes (CV) of buffer W (100 mM Tris-HCl, 150  
210 mM NaCl, 1 mM EDTA, pH 8.0) containing 0.1% GDN. s-ORF1-h was eluted in eight fractions, each  
211 consisting of 2 mL of buffer E (100 mM Tris-HCl, 150 mM NaCl, 1mM EDTA, 5 mM desthiobiotin, 0.1%  
212 GDN pH 8.0). Fractions containing s-ORF1-h were pooled and diluted two times in binding buffer (100  
213 mM Tris-HCl, 150 mM NaCl, 0.02% GDN, pH 8.0). Sample was then loaded on a pre-equilibrated 1 mL  
214 HisTrap HP column (Cytivia) and protein was eluted with elution buffer (100 mM Tris-HCl, 150 mM  
215 NaCl, 0.02% GDN, 500 mM imidazole, pH 8.0). Finally, the fractions containing s-ORF1-h were pooled  
216 and incubated with 200 µL of pre-equilibrated Strep-Tactin Superflow high capacity resin (IBA  
217 Lifesciences) during 4 h at 4°C on a rolling wheel. The flow-through (FT) was collected and the resin  
218 was washed with 5 CV of binding buffer W containing 0.01% GDN. s-ORF1-h was eluted in ten fractions,  
219 each consisting of 100 µL of buffer E (0.01% GDN instead of 0.1%). Purification process was monitored  
220 by SDS-PAGE on 10% acrylamides gels.

## 221 **2.5 Expression and purification of FABD-like domain**

222 BL21(DE3) cells transformed with plasmid encoding MBP – FABD-like domain fusion protein  
223 were grown at 30°C in LB medium supplemented with kanamycin. Protein expression was induced by  
224 1mM IPTG overnight at 18°C. Cells were harvested by centrifugation, resuspended in lysis buffer (50  
225 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mg/mL lysozyme, 5% glycerol supplemented by homemade  
226 benzonase and EDTA-free protease inhibitor (Roche)) and lysed by sonication. After centrifugation at  
227 20,000 g during 60 min at 4°C, the supernatant was filtrated on a 0.22 µm filter. The lysate was then

228 incubated under agitation during 2 h at 4°C with 10 mL of pre-equilibrated Dextrin Sepharose HP resin  
229 (Cytivia). The flow-through was collected and the resin was washed with 2.5 CV of washing buffer (50  
230 mM Tris-HCl, 200 mM NaCl, pH 8.0). Protein was eluted with 5 CV of elution buffer (50 mM Tris-HCl,  
231 200 mM NaCl, 10 mM maltose, pH 8.0). 2 mg of TEV was added to elution fractions containing the MBP  
232 – FABD-like domain fusion protein. After incubation 4 h at room temperature on a rotary wheel, the  
233 sample was centrifuged during 30 min at 20,000 and 4°C. The supernatant was incubated under  
234 agitation during 1 h at 4°C with 20 mL of pre-equilibrated Ni Sepharose HP resin (Cytivia). The flow-  
235 through was collected and the resin was washed with binding buffer (50 mM Tris-HCl, 200 mM NaCl,  
236 30 mM imidazole, pH 8.0). Elution was performed with 8 CV of elution buffer (50 mM Tris-HCl, 200 mM  
237 NaCl, 500 mM imidazole, pH 8.0). Finally, elution fractions containing the cleaved FABD-like domain  
238 were concentrated and loaded onto a pre-equilibrated HiLoad 16/600 Superdex75 gel filtration column  
239 (Cytivia). Elution was performed with 50 mM Tris-HCl, 150 mM NaCl, 1 mM DTT, pH 8.0. The pure  
240 protein was concentrated and stored at -80°C. Purification process was monitored by SDS-PAGE on  
241 12% acrylamides gels.

242 The polyclonal antiserums against the FABD-like domain of HEV gt1 or gt3 were obtained by  
243 injecting rabbits with the recombinant proteins. Immunization and antibodies purification were  
244 performed by Covalab according to standard protocols. Immunoblot analysis were performed as  
245 described above with a 100,000-fold dilution of the anti-FABD-like domain antibodies.

## 246 **2.6 Mass spectrometry analyses**

### 247 **2.6.1 Tryptic digestion**

248 Tryptic digestion of s-ORF1-h from gt1 HEV and peptide analysis were performed as previously  
249 described <sup>48</sup>.

### 250 **2.6.2 Denaturing and native mass spectrometry analysis**

251 Samples of FABD-like domain from gt3 HEV were first buffer exchanged in 200 mM ammonium  
252 acetate (AcONH<sub>4</sub>) at pH 8 using 0.5 mL Zeba™ Spin desalting Columns with a 7 kDa molecular weight  
253 cutoff (Thermo Fisher Scientific). Then, concentrations were determined by UV-Vis using a Nanodrop

254 2000 Spectrophotometer (Thermo Fisher Scientific). Finally, analyses were performed on an  
255 electrospray time-of-flight mass spectrometer (LCT, Waters) equipped with an automated chip-based  
256 nanoESI source (Triversa Nanomate, Advion).

257 For analysis under denaturing conditions, the instrument was calibrated using multiply  
258 charged ions of a 2  $\mu$ M horse heart myoglobin solution. Samples of FABD-like domain from gt3 HEV  
259 were diluted to 5  $\mu$ M for infusion in a 1/1 (v/v) water/acetonitrile mixture acidified with 1% formic acid  
260 and interface parameters were fixed to 40 V for accelerating voltage (Vc) and 2 mbar for backing  
261 pressure (bP) to obtain the best mass accuracy. Analyses under native conditions were performed after  
262 careful optimization of instrumental settings to obtain sensitive detection of the protein. The mass  
263 spectrometer was calibrated using singly charged ions produced by a 2 g/L solution of cesium iodide  
264 (Acros organics, Thermo Fisher Scientific) in 2-propanol/water (50/50 v/v). Vc and bP were fixed to 120  
265 V and 6 mbar, respectively. Protein was diluted to 25  $\mu$ M in AcONH<sub>4</sub> (pH 8). Acquisitions were  
266 performed in 1,000–10,000 m/z range with a 4 sec scan time in positive mode. Data analysis was  
267 performed with MassLynx V4.1 (Waters).

268

### 269 3. RESULTS AND DISCUSSION

#### 270 3.1 Full-length HEV pORF1 modelling: An organization in two modules and five domains

271 Until very recently, it was almost impossible to produce accurate 3D atomic models by  
272 homology modelling for proteins that share sequence identity with homologues lower than 30%. In  
273 2018, AlphaFold came and revolutionized the field of molecular modelling. Two years later, a novel  
274 breakthrough allowed AlphaFold2 to produce models with high accuracy and very good estimation of  
275 the error in the coordinates<sup>31,49</sup>. Indeed, several groups reported the resolution of experimental  
276 structures using 3D models generated by AlphaFold2, or showed that AlphaFold2 models are highly  
277 similar to experimental structures<sup>50,51</sup>.

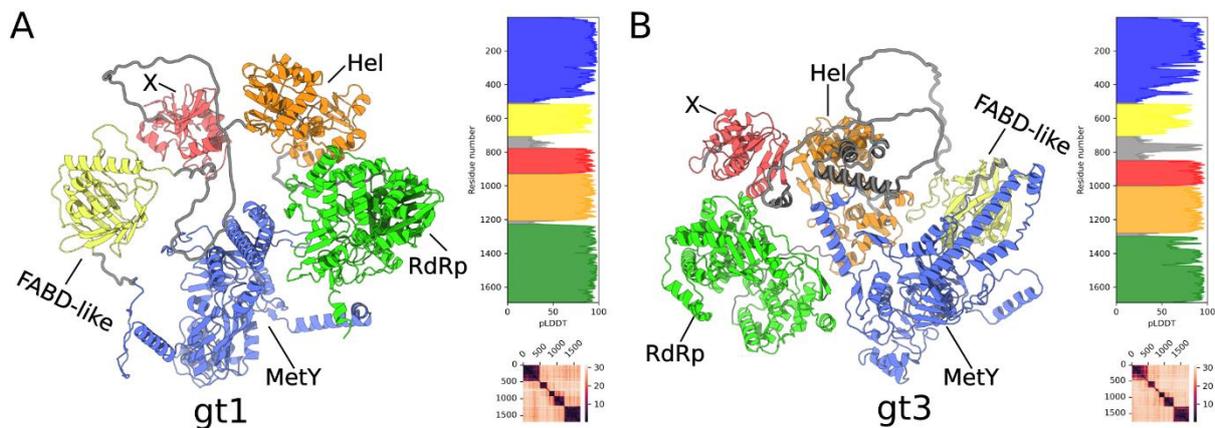
278 Thus, in order to gain insight about the structural properties of HEV replication polyprotein,  
279 we produced 25 AlphaFold2 3D models for the full-length pORF1, for both genotypes 1 and 3 (Fig.1

280 and Fig.S1). To improve our gt1 model, we used the crystallographic structure and other generated  
281 domain models to produce a meta-model of the full-length pORF1. The average pLDDT scores for the  
282 C $\alpha$  of the predicted structural models are higher than 70 (76 and 78 for gt1 and gt3, respectively),  
283 which indicates their overall quality <sup>49</sup>. However, analysis of C $\alpha$  pLDDT scores along the sequence  
284 reveals some parts of the models of very low confidence (pLDDT values < 50; Fig.1 and Fig.S1), which  
285 is usually indicative of disordered regions <sup>49</sup>, possibly linkers between distinct domains. Indeed,  
286 combining this analysis to the observation of the structural models and the predicted alignment error  
287 (PAE) plots, we found that HEV pORF1, whatever the genotype 1 or 3, is composed of five main  
288 domains. This finding is similar to a recent work that has been published when we were writing this  
289 paper <sup>52</sup>. Indeed, our gt1 and gt3 HEV pORF1 models and the gt3 HEV pORF1 model by Goulet *et al.*  
290 exhibit the same five domains (Fig.S2), albeit the relative orientations of all domains are different.  
291 Actually, both the high PAEs between domains and their varying relative positions in different models  
292 indicate that generally no favored position of one domain with respect to another is detected by  
293 AlphaFold2, in accordance with the report of Goulet *et al.* who modelled pORF1 in two overlapping  
294 fragments <sup>52</sup>. However, in a few of our 25 full-length pORF1 models, PAEs are distinctly lower between  
295 the fifth and first domains (*e.g.*, gt1-hhblits-rank4) or between the fifth and fourth (*e.g.*, gt1-  
296 mmseqs\_rank2) (Fig.S1). This finding may reveal transient interactions between enzymes in HEV  
297 pORF1 regulating different steps of the viral cycle (see below, section 3.4).

298         The five domains are organized into two modules separated by a long (> 70 residues)  
299 disordered region (Fig.1). The nature of this region is consistent with what has been previously  
300 proposed for the hypervariable region HVR, for both its position and unstructured properties <sup>13,53,54</sup>. In  
301 contrast to the work by Goulet *et al.*, we modeled the structure of a gt3 HEV pORF1 that contains an  
302 insertion within the HVR (*i.e.*, Kernow C1-p6 strain), corresponding to a fragment of cellular gene that  
303 has been described for some gt3 strains <sup>44,55-57</sup>. Indeed, we found a 50-residue structured region within  
304 the HVR in gt3 HEV (Fig.1).

305 Since it was commonly accepted up to now that HEV pORF1 contains six domains <sup>13</sup> and not  
 306 five as suggested by the AlphaFold2 models, we carefully analyzed each domain. For that purpose, we  
 307 searched for structural homologs for each domain. Of note, we used DALI server <sup>58,59</sup> which is  
 308 recognized as the most sensitive tool. DALI sorts the structural homologs with a Z-score, value above  
 309 10 meaning the two structures have the same fold <sup>60</sup>. Analysis of these results are presented below for  
 310 gt1 HEV only, with corresponding amino acid residues numbering.

311



312

313 **Fig. 1. Selected AlphaFold2 models for full-length HEV pORF1 gt1 and gt3.**  
 314 Global AlphaFold2 model for gt1 (A) and gt3 (B) HEV pORF1 with respective pLDDT plot  
 315 and PAE matrix. Models and pLDDT graphs are colored according to their domain  
 316 decompositions: MetY in blue, FABD-like in yellow, X-macrodomain in red, helicase in  
 317 orange and polymerase in green. Linkers and the hypervariable regions are colored in  
 318 gray. pLDDT values for gt1 FABD-like are overestimated since the reference used was a  
 319 crystallographic structure (see Table 1 in Methods).

320

### 321 3.2 The N-terminal module: reassessment of the boundaries and nature of domains

#### 322 3.2.1 Structural analysis of the N-terminal domain: a single MetY domain

323 In their first analysis, Koonin *et al.* mapped the fragment 56-433 of pORF1 as two different Met  
 324 and Y domains that partially overlap (residues 56-240 and 219-433, respectively) <sup>13</sup>. Later, sequence  
 325 analyses have suggested that the Met core domain would span from residue 33 to residue 353 <sup>43</sup> and  
 326 that the Y domain would be an extension of the Met core domain <sup>61</sup>, essential for viral life cycle <sup>62</sup>.  
 327 Indeed, the AlphaFold2 models suggest only one domain, ranging from residue 1 to residue 506. This  
 328 domain is composed of both  $\alpha$ -helices and  $\beta$ -strands (Fig.2A) as previously suggested <sup>61-63</sup>. Some of the

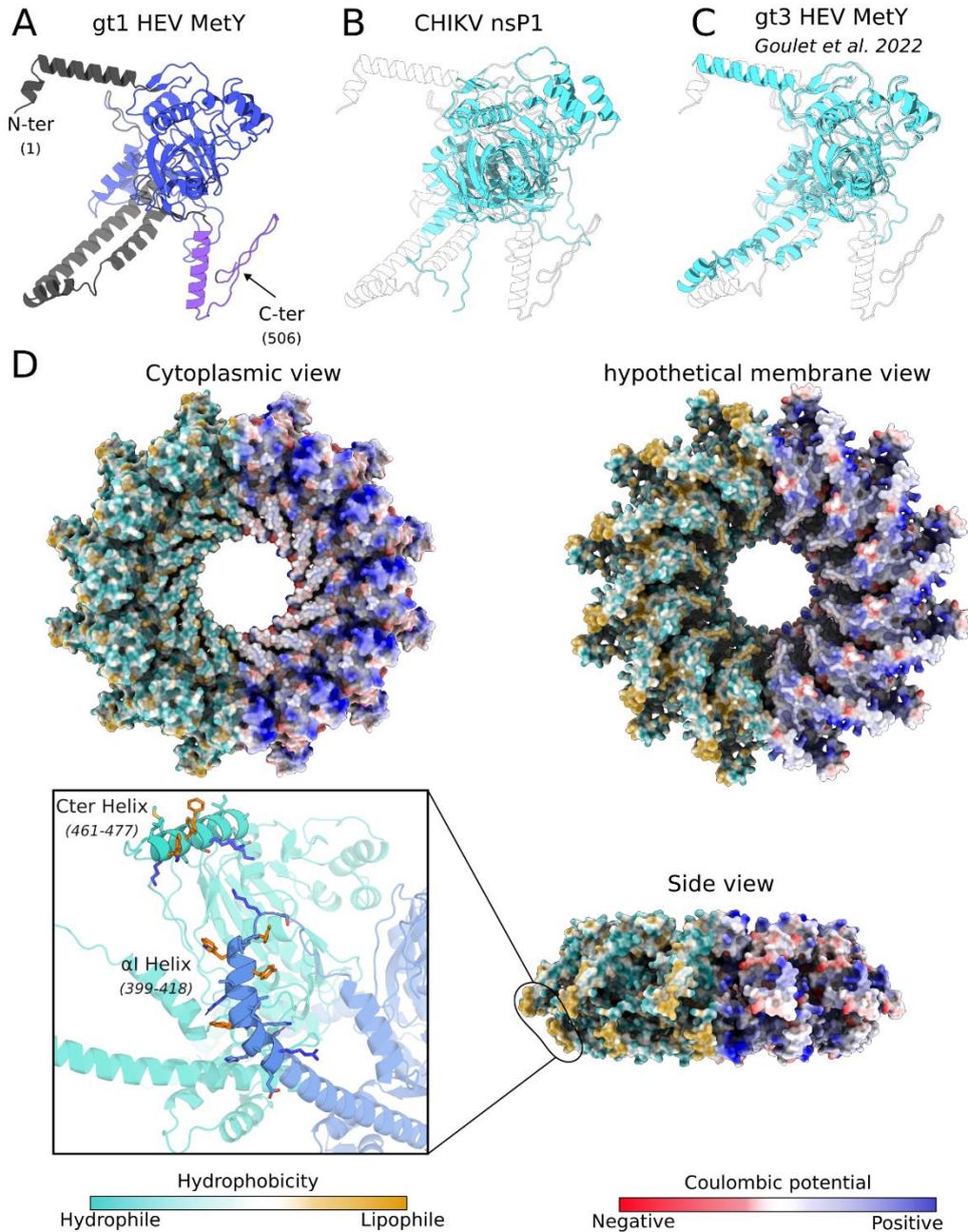
329 peripheral helices may have slightly different orientations and/or lengths in our different models, as  
330 well as lower pLDDT (Fig.S1). However, the folds remain clearly identical.

331 In order to not influence our search for structural homologs, we used as templates either the  
332 9-270 and 285-495 fragments separately that would possibly correspond to the independent Met and  
333 Y domains, or the whole 1-506 fragment. All three searches identified the same structural homologs,  
334 corresponding mostly to viral capping enzymes that are a hallmark of Alphavirus-like viruses <sup>64</sup>. This  
335 result is consistent with the Met function of this HEV pORF1 region (*i.e.*, adding a 5'-methylguanine  
336 (m7G) cap to the viral RNA progeny) <sup>13,63,65</sup> and thus indicative of the AlphaFold2 model quality. The  
337 DALI search identified the nsP1 capping enzyme from the Alphavirus Chikungunya virus (CHIKV) (PDB  
338 codes 6Z0V <sup>66</sup> and 7DOP <sup>67</sup>) as the closest structural homolog (Z-score > 21 for the search using the  
339 whole 1-506 fragment). Comparison of the HEV 1-506 domain (Fig.2A) with the CHIKV nsP1 monomer  
340 (Fig.2B) reveals a common core, corresponding to residues 33-353 in HEV. Incidentally, this HEV  
341 fragment has been shown to be the active catalytic core of the HEV Met domain <sup>43,63</sup>. In addition, the  
342 superimposition reveals several notable  $\alpha$ -helices and loops that are specific to the HEV domain and  
343 not present in the CHIKV nsP1 structure. First, HEV 1-506 domain presents an extra  $\alpha$ -helix in its N-  
344 terminal extremity (Fig.2AB), also identified by the 3D model published by Goulet *et al.* <sup>52</sup>. This helix is  
345 preceded by nine residues that tend to form one turn of an  $\alpha$ -helix in most of our models (Fig.2A and  
346 Fig.S1). We suggest that the alternate unfolded/folded states of these residues would regulate some  
347 functions of HEV pORF1. Second, a set of four  $\alpha$ -helices and several loops is inserted in various places  
348 of the C-terminal region in HEV (Fig.2AB). Interestingly, these C-ter structural elements match to  
349 several portions of the Y domain that were proposed to be actually an extension of the Met core  
350 domain, known as the Iceberg region <sup>61</sup>. Moreover, this extension was already proposed to fold as  $\alpha$ -  
351 helices <sup>61,62</sup> including its L<sub>410</sub>YSWLFE<sub>416</sub> motif in HEV ORF1 that does form a small  $\alpha$ -helix in all AlphaFold  
352 models as predicted by Ahola and Karlin in 2015 <sup>61</sup> (their helix  $\alpha$ I) . It has been experimentally shown  
353 that helix  $\alpha$ I of the Iceberg region is involved in targeting the replication complex of several alphavirus-  
354 like plant viruses to the correct cellular endomembrane <sup>68,69</sup>. Indeed, guided by reference <sup>61</sup>,

355 Sathanantham *et al.* have shown that helix  $\alpha$ 1 is sufficient to target soluble proteins to membrane  
356 compartments not only for several plant viruses but also for HEV. This region ends by looping back to  
357 the Met part, donating an extra  $\beta$ -strand to one of the core Met sheets. Thus, we conclude that the  
358 two regions formerly annotated as Met and Y are linked in a single domain we term MetY.

359 Finally, in comparison to the AlphaFold2 model proposed by Goulet *et al.* (Fig.2C), we found a  
360 longer MetY domain, with an additional  $\alpha$ -helix 461-477 at the C-terminal extremity extended by a  
361 disordered loop (Fig.2A). Thus, we set the domain boundary at residue 506 and not 459. The fact that  
362 this C-terminal helix is in different orientations in our 25 models and sometimes incomplete suggests  
363 that it could alternate between structured and unstructured states, that could also be an important  
364 regulation of the viral life cycle (see below).

365



366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

**Fig. 2. MetY domain modelling and its hypothetical dodecamerization.**

(A) Our gt1 HEV MetY domain model (residues 1-506), where residues 33 to 353 are colored in blue while other residues are in black except for the C-terminal helix (460-506) in purple. The same model in transparent white is superimposed on (B) the cryo-EM structure of CHIKV nsP1 (PDBID: 7DOP<sup>67</sup>) and (C) the HEV MetY model from Goulet *et al.*<sup>52</sup>, both in cyan. (D) Dodecamer assembly of the MetY Domain based on the best 2-mer model repeated 6 times and adjusted to a closed ring (see Methods and Fig.S3). Inset, focus on the two-part aromatic patch made of helix  $\alpha$  of reference<sup>61</sup>(399-418, blue) and the C-terminal helix of a second subunit (461-477, cyan). Colored per hydrophobic scale (left side of each image) and electrostatics (right side of the image). Colors range from cyan for hydrophilic region to gold for lipophilic area and from red for negative to blue for positive area. In the atomic representation aromatics are colored in orange and basics in blue. Other amino acids are colored per sub-unit.

### 381 **3.2.2 Putative oligomerization of MetY**

382 The cryo-electron microscopy (cryoEM) structures of CHIKV nsP1 revealed that the protein is  
383 assembled in a dodecameric ring that likely constitutes a membrane pore<sup>66,67</sup>. The authors identified  
384 two membrane-interacting loops that protrude from the core protein. Interestingly, these loops are  
385 not present in HEV MetY. Rather similar positions are occupied by helices, including the  
386 aforementioned helices  $\alpha$ I and 461-477 (Fig.2AB). Thus, we wondered if and how MetY could  
387 oligomerize. For that purpose, we used again AlphaFold2 to model 2-mers or 3-mers of MetY and  
388 examined their organizations. Although some came out as C2 dimers or C3 trimers, several were side-  
389 by-side 2- or 3-molecule open assemblies that were possible building blocks of a higher order oligomer.  
390 Indeed, for the 2-mer with the best inter-molecule PAE scores, simply applying the transformation  
391 between the two molecules twelve times successively yielded a dodecamer that is almost a closed ring  
392 (Fig.2D and Fig.S3), highly similar to the CHIKV nsP1 dodecamer, which we simply straightened out (see  
393 Methods). Compared to CHIKV nsP1, the extra MetY N-terminal helix seems to be an important  
394 oligomerization factor. The C-terminal helix 461-477 contributes to an aromatic patch, the other side  
395 of which is the hydrophobic face of amphipathic helix  $\alpha$ I of the next MetY in the oligomer. We propose  
396 that this hydrophobic patch, that would form only in the assembled form of MetY, may be a central  
397 membrane interaction surface (Fig.2D).

398 We conclude from these structural AlphaFold2 models that the HEV pORF1 MetY domain folds  
399 as a single domain that contains the methyltransferase capping enzyme activity and that is actually the  
400 counterpart of nsP1 of Alphaviruses. The domain consists of a catalytic core domain with a C-terminal  
401 extension, previously known as the Y domain and Iceberg regions, that is likely to both interact with  
402 cellular endomembranes during infection and trigger the domain oligomerization . We propose that  
403 this MetY domain may oligomerize to form a capping pore similar to the Alphavirus nsP1, but with  
404 distinctly different membrane interaction surfaces. This putative oligomerization has also been  
405 suggested by Goulet *et al.*<sup>52</sup>.

### 406 **3.2.3 Structural analysis of the second domain: a fatty acid binding protein fold**

407           Following the MetY domain, the second pORF1 domain spans from residue 515 to residue 707.  
408   Almost all the AlphaFold2 models are identical to the crystal structure of the gt1 HEV 510-691 fragment  
409   <sup>29</sup> which returned to be the best DALI hit (Z-score = 32.1, rmsd = 0.9 Å over 167 Cα) (Fig.3A). In addition,  
410   as previously reported <sup>29</sup>, we found that this domain is structurally related to many fatty acid binding  
411   proteins (FABPs) (Z-scores in the range 10-11). FABPs are chaperones that bind and transport  
412   hydrophobic molecules in animal cells <sup>70</sup>. FABPs structure consists in two five-stranded β-sheets  
413   forming a β-barrel, the cavity of which corresponds to the ligand binding cavity <sup>70,71</sup> (Fig.3B). In addition,  
414   the β-barrel is partially capped by an N-terminal helix-turn-helix motif that is proposed to act as a  
415   regulatory portal for ligand binding <sup>70,71</sup>. Since no function has been assigned yet to this HEV domain  
416   that folds as a fatty acid binding domain (FABD), we will refer to it as a FABD-like domain.

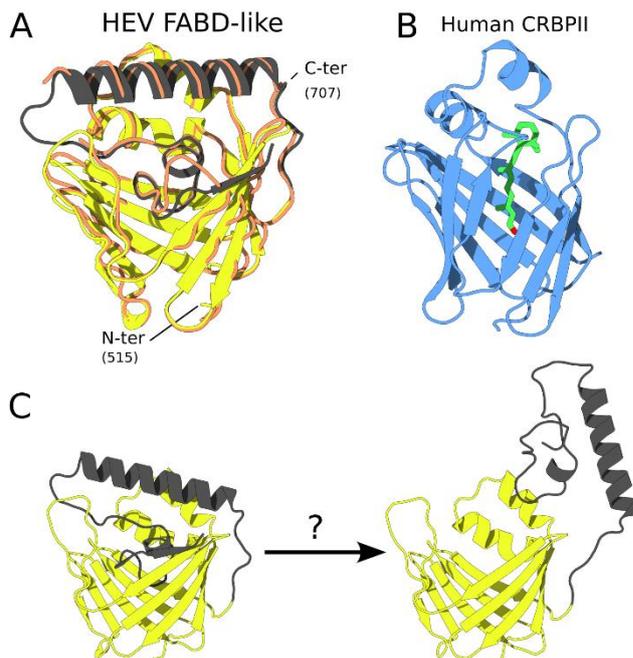
417           Importantly, compared to FABPs, HEV FABD-like domain presents a long extension in its C-  
418   terminal extremity (residues 635-690 highlighted in grey in Fig.3A), essential for protein folding and  
419   possibly also for its function <sup>29</sup>. This extension consists in a long α-helix that caps the β-barrel next to  
420   the FABP helix-turn-helix, followed by a loop with a helical turn in its middle (Fig.3A) that occupies the  
421   β-barrel cavity where ligands bind in FABPs (Fig.3B). In some of our models, this C-terminal loop is  
422   extended to residue 707 and folds over the domain, forming an additional cover (Fig.S4A). Moreover,  
423   in other models (Fig.S4B), this loop is actually out of the cavity, suggesting that this block may be  
424   mobile. Indeed, it is tempting to hypothesize that the loop that connects the C-terminal extension to  
425   the FABD-like domain core would be a hinge that would allow a conformational change of this  
426   extension. Thus we propose that this C-terminal extension could open, in a large movement starting  
427   from residue 635, resulting in an empty β-barrel cavity (Fig.3C). Whether this movement would then  
428   allow the binding of ligands remains to be established.

429           Authors of the crystal structure of the gt1 FABD-like domain located a metal ion coordinated  
430   by residues His671, Glu673 and possibly His686 and they suggested that it could have a catalytic role  
431   <sup>29</sup>. However, in our hands, the purified recombinant gt3 FABD-like domain is free from any metal  
432   (Fig.S4C). Thus, it is not clear if a metal ion binds this domain and, if so, what its role is. Goulet *et al.*

433 also noted that the putative metal binding site was the most structurally variable part when comparing  
434 gt3 and gt1 FABD-like domains, and suggested it might be something else such as an iron-sulfur binding  
435 site<sup>52</sup>.

436 Since (+)RNA viruses have a compact genome that usually contains only what is strictly  
437 necessary for viral replication, the conservation of an FABD-like domain sequence in all HEV genotypes  
438 means that this domain is under a strong positive selection pressure. Clearly, the function of the HEV  
439 FABD-like domain thus deserves to be investigated, especially the role of its C-terminal extension. The  
440 simplest hypothesis is that it binds fatty acids or similar molecules as all other known members of the  
441 family<sup>70,71</sup>, but we failed to identify ligands for the recombinant FABD-like domain (data not shown).  
442 Possibly an external signal triggering the putative cap opening is required.

443



**Fig. 3. FABD-like domain modelization.** (A) Comparison between our gt1 HEV FABD-like model (yellow and black) and the crystal structure (PDBID: 6NU9<sup>72</sup>, orange ribbon). (B) Structure of the human CRBP II with its retinal ligand bound in its binding pocket (PDBID: 6QYP<sup>73</sup>). (C) Side view of our model in closed and hypothetical open conformation. In our AlphaFold2 models, the FABD-like core domain is yellow (505-634) and its C-terminal extension (635-690) is in gray.

444

### 445 3.3 The C-terminal module: a X-macrodomain, a HEL domain and an RNA polymerase

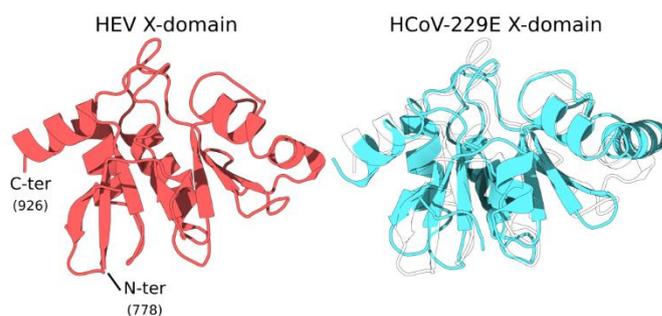
#### 446 3.3.1 Structural analysis of the X-macrodomain

447 The first domain found after the HVR spans from residue 778 to 926. The DALI search for  
448 structural homologs returned many hits, all corresponding to viral X-macrodomains. The 100 first hits  
449 were obtained with Z-scores higher than 19, values that indicate a high structural homology. Indeed,

450 the comparison of the AlphaFold2 models for the 778-926 domain of gt1 HEV and the most closely  
451 related structural homolog, the X-macrodomain of the human coronavirus E229E (PDB code 3EJG <sup>74</sup>),  
452 reveals a common fold, composed of a seven-stranded  $\beta$ -sheet sandwiched in by four to six  $\alpha$ -helices  
453 (Fig.4). The HEV X-macrodomain does not present any insertion or deletion compared to other viral X-  
454 macrodomains (Fig.4).

455 Thus, both the position, the size and the fold of this domain perfectly match with previous  
456 works that report downstream the HVR a putative X-macrodomain which exhibits an ADP-ribose-1''-  
457 monophosphatase activity and binds ADP-ribose <sup>13,75,76</sup>. Although it has been suggested that the HEV  
458 X-macrodomain could interact with the HEV pORF1 MetY domain and pORF3 <sup>77,78</sup>, its exact roles in viral  
459 life cycle remain to be understood.

460



**Fig. 4. X-macrodomain modelization.**

Comparison between our gt1 HEV X-macrodomain model (residues 778-926) (left, red) and HCoV-229E X-macrodomain (PDBID: 3EJG <sup>79</sup>, right, cyan). Superposition of our HEV-X domain is represented by a transparent cartoon.

461

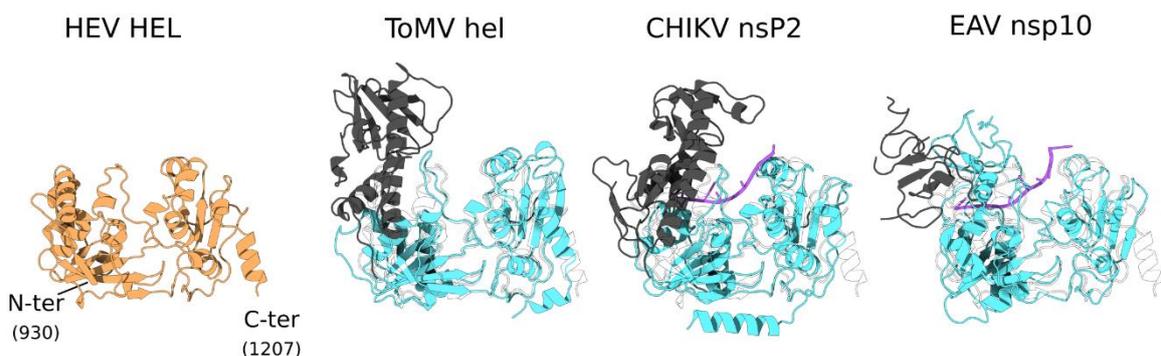
### 462 3.3.2 Structural analysis of the HEL domain

463 Immediately after the X-macrodomain, starting at residue 930 and finishing at residue 1207,  
464 the AlphaFold2 models exhibit a bi-lobe domain (Fig.5 left). All structural homologs identified by the  
465 DALI search are helicases, with Z-scores higher than 19 for the 10 first hits. Most of them are encoded  
466 by viruses belonging to the Alphavirus-like superfamily (Tomato mosaic virus (ToMV), CHIKV) or  
467 Nidovirales (equine arteritis virus (EAV), porcine reproductive and respiratory syndrome (PRRS), SARS-  
468 CoV-2). Because of sequence similarities and conservation of consensus motifs, all these enzymes are  
469 classified into the superfamily 1 (SF1) helicases <sup>80</sup>. This structural homology is consistent with previous

470 works that reported a putative HEV HEL domain to have conserved motifs typical of SF1 helicases<sup>11,13</sup>  
471 and a NTPase/RNA helicase activity<sup>81,82</sup>.

472 Helicases are enzymes that unwind DNA or RNA duplex substrates through a nucleic acid–  
473 dependent ATPase activity. They are expressed by all (+)RNA viruses whose genome is larger than 6-7  
474 kb<sup>83</sup>. The core helicase bi-lobe domain ensures the unwinding activity, with the polynucleotide  
475 substrate bound at the interface between the two lobes (see for example CHIKV nsP2 and EAV nsp10,  
476 Fig.5). Helicases often contain accessory domains that ensure additional functions such as modulation  
477 of the substrate specificity or helicase activity, or allow interaction with partners<sup>84</sup>. In contrast to its  
478 closest structural homologs, the HEV HEL enzyme presents only the bi-lobe core domain (Fig.5 and  
479 reference<sup>52</sup>), suggesting a low level of regulation. Nevertheless, several mutations of this domain that  
480 severely impair the helicase activity have been reported, some of them being associated to fulminant  
481 hepatic failures<sup>82,85,86</sup>. These results highlight the role of the 930-1207 HEL domain in HEV replication  
482 cycle and its interest to develop specific antiviral molecules. Our accurate AlphaFold2 models are likely  
483 to further this aim.

484



485

486 **Fig. 5. HEL domain (930-1207) modelization.**

487 Comparison between our gt1 HEV HEL domain model (residues 930-1207) (left, orange)  
488 with the ToMV HEL (PDBID: 3VKW<sup>87</sup>), the HEL domain from CHIKV nsP2 (PDBID: 6JIM<sup>88</sup>)  
489 and EAV nsp10 (PDBID: 4N0O<sup>89</sup>) structures (cyan). Superposition of our HEL domain is  
490 represented by a transparent cartoon.

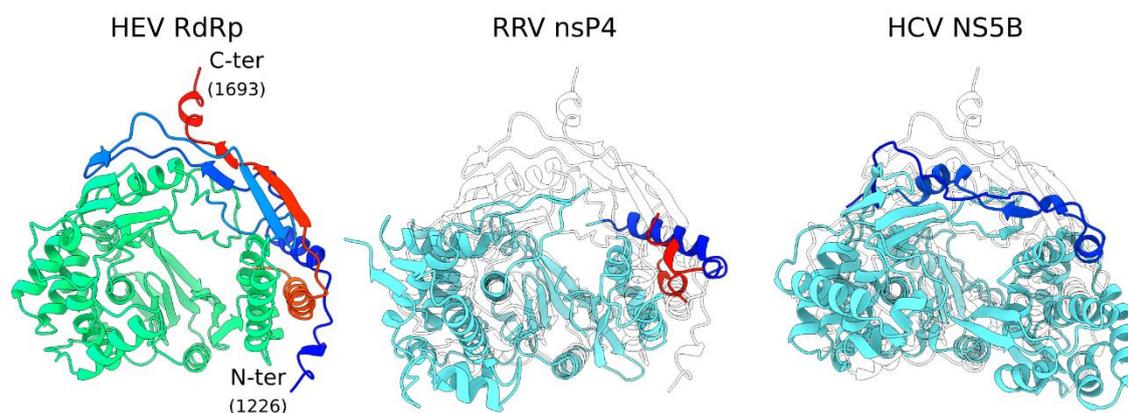
491

492 **3.3.3 Structural analysis of the RNA-dependent RNA polymerase domain**

493 According to the AlphaFold2 models, HEV pORF1 ends with a domain delimited by residues  
494 1226 and 1693 (Fig.6), which corresponds well to the region 1207-1693 that has been proposed to be  
495 the RdRp enzyme <sup>13</sup>. The search for structural homologs identified many viral RNA polymerases, a  
496 result that confirms the validity of the structural models. Using sequence alignments, three super-  
497 groups (I, II and III) of RdRp have been defined <sup>12</sup>. Due to its homology with Alphavirus-like viruses,  
498 especially RUBV and beet necrotic yellow vein virus (BNYVV), the HEV RdRp belongs to super-group III  
499 <sup>12,13</sup>. No experimental structures of super-group III RdRp had been released until recently, where partial  
500 structures of the nsP4 from the Alphaviruses Ross River virus (RRV) and Sindbis virus (SINV) were  
501 published <sup>90</sup>. Accordingly, our DALI search finds these nsP4 with high Z-scores (~15). However, the best  
502 hits (Z-score > 19) are RdRps encoded by bovine viral diarrhea virus (BVDV), classical swine fever virus  
503 (CSFV) and hepatitis C virus (HCV), *i.e.* super-group II RdRp from the *Flaviviridae* family. This is due to  
504 disorder in the Alphavirus nsP4 that leads to only some 400 residues (out of 502 in the RRV RdRp  
505 fragment that was crystallized and 611 in the full-length nsP4) included in the crystal structure <sup>90</sup>.  
506 Specifically, 'right hand' polymerases are described as composed of three subdomains: fingers, palm  
507 and thumb. In viral RdRps, a specific extension of the fingers termed 'fingertips' reaches out to the  
508 thumb <sup>91</sup>. The fingertips and their surroundings are completely disordered in the two available crystal  
509 structures of super-group III RdRp (RRV and SINV nsP4 <sup>90</sup>) and these missing parts account for their  
510 lower Z-score in searches with our HEV RdRp model. The fingertips usually harbor only two loops  
511 inserted in the fingers, as in HCV (Fig. 6, right) or in most super-group I RdRps. Our HEV RdRp model  
512 displays remarkable elaborations on this basic architecture: Two extra segments contributed by its N-  
513 terminus and C-terminus are actually incorporated in the fingertips (Fig. 6, left). Superposition shows  
514 that the counterpart segments are mostly disordered in the RRV nsP4 structure, but the few ordered  
515 residues are positioned identically in the two proteins (Fig. 6, middle). The dynamic nature of the  
516 folding of the fingertips of RRV nsP4 was further established by HDX-MS <sup>90</sup> and its general character in  
517 super-group III RdRp may explain why these enzymes resisted crystallization for so long. We note that  
518 the RdRp domain was one of the three HEV pORF1 domains, with the MetY and FABD-like, for which

519 some of our >30 AlphaFold2 models were in alternate conformations and/or states of folding. Notably  
 520 in the alternate RdRp models, either the C-terminal segment was not incorporated in the fingertips  
 521 (Fig.S5) or the fingertips were disordered altogether. It is tempting to speculate that this flexibility is a  
 522 real feature of the supergroup III RdRp fold and constitutes a layer of regulation of the RdRp activity,  
 523 the fingertips being a critical point for activating or inhibiting RNA synthesis<sup>92</sup>. In accordance with this,  
 524 the C-terminus of the HEV RdRp is not permissive to changes, as shown by the fact that adding a tag  
 525 at the C-terminus of pORF1 prevents viral replication<sup>93</sup>. In HEV, such a regulation might indeed be  
 526 necessary if pORF1 is not proteolytically processed, *e.g.* to restrict RNA polymerase activity to a small  
 527 subset of the 12 RdRp molecules tethered to the dodecameric MetY pore.

528



529

530 **Fig. 6. RdRp domain modelization.**

531 Comparison between our gt1 HEV RdRp domain model (residues 1226-1693) (left, gold)  
 532 with the RRV nsP4 (middle, PDBID: 7F0S<sup>90</sup>) and HCV NS5B (right, PDBID: 1GX5<sup>94</sup>)  
 533 structures (cyan). Middle and right, superposition of the HEV RdRp domain is represented  
 534 by a transparent cartoon. Alignment made with TM-align<sup>41</sup> with its dedicated PyMOL  
 535 Addon and image rendered with ChimeraX<sup>37,38</sup>. The first loop of HCV NS5B's 'fingertips'  
 536 (residues 1-56) is in blue, as is the larger N-terminus of the HEV RdRp (residues 1226-1314)  
 537 and the ordered part of the N-terminus of RRV nsP4 (residues 111-134). The C-terminus  
 538 of the HEV RdRp (residues 1650-1693) and the ordered part of the C-terminus of RRV nsP4  
 539 (residues 579-600) are in red. There is no counterpart for this in HCV NS5B.

540

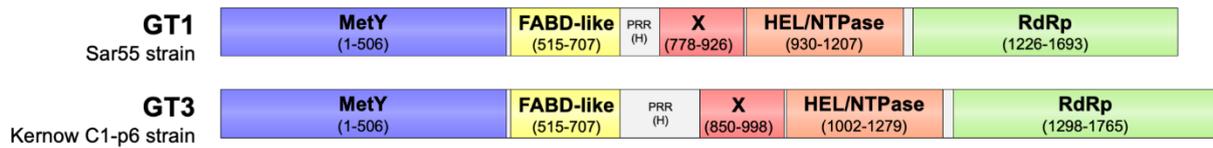
541 **3.4 Update of HEV pORF1 organization**

542 The reliability of our AlphaFold2 models, allows us to update the domain composition of the  
 543 HEV pORF1, especially for the N-terminal part. The new organization is applicable for both genotypes  
 544 1 and 3. Thus, HEV replication polyprotein contains five independent domains, organized into two

545 modules separated by a long disordered loop corresponding to the proline-rich hypervariable region  
546 (HVR) (see Fig.7 for domain boundaries). From the N- to the C-terminal extremities, we found: 1) a  
547 unique MetY domain that combines the two Met and Y domains previously described, 2) a FABD-like  
548 domain, 3) a X-macrodomein, 4) a HEL domain and 5) a RdRp domain. The MetY domain is likely to  
549 oligomerize and to bind cellular membranes. The function of the FABD-like domain has to be  
550 investigated to understand its role in HEV replication and/or pathogenicity. Downstream the HVR, the  
551 three C-terminal domains are typical X-macrodomein, helicase HEL and viral RNA-dependent RNA  
552 polymerase RdRp with some noteworthy specific features as described above. It is important to note  
553 that this breakdown establishes that there is nothing with a canonical protease fold encoded in the  
554 pORF1 N-terminal module. It is not uncommon in the evolution of RNA viruses to acquire new folds for  
555 canonical enzymatic activities. However, as all HEV pORF1 five domains belong to families with well  
556 characterized functions, none of which include a reported protease activity, the most likely conclusion  
557 is that unlike other Alphavirus-like animal viruses, but similarly to some Alphavirus-like plant viruses,  
558 HEV pORF1 does not encode its own self-processing activity.

559         In contrast to Goulet *et al.*<sup>52</sup>, we do find hints of preferential contacts between pairs of  
560 domains in some of our models by analyzing PAE plots (see above, section 3.1). While they are not on  
561 the same level as those found in a stable oligomer (*e.g.*, the MetY 2-molecule models hinting at a  
562 dodecameric organization), it is noteworthy that they occur between RdRp and methyltransferase on  
563 the one hand and between RdRp and helicase on the other hand. Moreover, the connection between  
564 domains X and HEL is very short, with less than five residues between the two folded domains. As a  
565 result, X and HEL remain very close together and it is unlikely that the linker between them could be  
566 cleaved if the HEV pORF1 is processed by a (cellular) protease. This is in sharp contrast to Alphaviruses,  
567 where X and HEL although synthesized in the same replication polyprotein reside in different mature  
568 products (nsP3 and nsP2, respectively) and where a late cleavage by the viral protease between nsP2  
569 and nsP3 actually regulates RNA synthesis<sup>95</sup>. The linker between HEL and RdRp is much longer (*ca* 20  
570 residues between folded domains), long enough in fact to allow considerable relocation of HEL around

571 RdRp even in the absence of cleavage between the domains (Fig.S6). This, together with a transient  
 572 interaction between HEL and RdRp, could be involved in RNA synthesis regulation.



573

574 **Fig. 7. Updated organization of HEV pORF1.**

575 Schematic representation of the HEV pORF1 replication polyprotein, containing a  
 576 methyltransferase (MetY) domain, a FABD-like domain, a macrodomain-X, a helicase  
 577 (HEL/NTPase) domain and a RNA-dependent RNA polymerase (RdRp) domain. A proline-  
 578 rich (PRR) hypervariable region (H) is found between the N-terminal module (MetY-FABD)  
 579 and the C-terminal module (X-HEL-RdRp).  
 580

581 **3.5 Wheat-germ cell-free expression and purification of HEV pORF1**

582 Although models generated by the AlphaFold2 tool (this work and reference <sup>52</sup>) are very  
 583 accurate, they do not tell us about the putative oligomerization of pORF1 (for example as for the MetY  
 584 domain as we suspect), nor about the interaction between domains and thus about the  
 585 communication between domains in order to drive and regulate viral RNA synthesis. It is therefore  
 586 necessary to obtain structural data by experimental approaches. Thus, we developed a protocol to  
 587 express and purify the full-length HEV pORF1. We expressed the polyprotein from gt1 and gt3 HEV by  
 588 the wheat-germ cell-free expression system, as we did for other viral replication polyproteins before  
 589 <sup>46</sup>. The Western-blot analyses reveal the presence of a major band in comparison to the negative  
 590 control, , with a molecular weight matching the expected size of the Strep-tagged full-length pORF1,  
 591 *i.e.* ~187.5 and 196 kDa for genotypes 1 and 3 respectively (Fig.8A, lanes CFS). In parallel, in order to  
 592 specifically detect HEV pORF1, we produced polyclonal antibodies against the FABD-like domain from  
 593 HEV gt1 and 3. These antibodies specifically recognize the recombinant protein expressed in *E. coli*  
 594 (Fig.S7A) and also allow the detection of full-length pORF1 for both gt1 and gt3 (Fig.S7B). Although the  
 595 band corresponding to the full-length pORF1 appears as the main band on Western-blots, we also  
 596 observe additional minor bands of lower molecular weights: one band of lower intensity around 70

597 kDa that is present for both genotypes, and two bands around 80 and 37 kDa for gt3 pORF1 only  
598 (Fig.8A). The 70 kDa band is always seen, sometimes with a very low intensity. The two bands that are  
599 specific to gt3 HEV are not systematically revealed in Western-blot (Fig.S7). Since pORF1 does not  
600 contain any canonical protease domain as described above and because the wheat-germ extract has  
601 no significant protease activity<sup>47,96</sup>, these results suggest that these fragments do not correspond to a  
602 specific maturation process occurring in the *in vitro* translation mixture but are likely to be degradation  
603 products, with spontaneous cleavage at defined sites prone to cleavage. This is consistent with several  
604 works that studied the maturation of recombinantly produced HEV pORF1 and who concluded that the  
605 polyprotein is not proteolytically processed<sup>18–23</sup> despite the presence of low molecular weight  
606 fragments obtained after expression of the full-length pORF1. Indeed, these fragments were obtained  
607 after prolonged incubation times, or even when the putative protease domain was catalytically  
608 inactivated by selected mutants, or even in absence of the N-terminal pORF1 region that putatively  
609 contains the hypothetical protease domain<sup>20,21,24,25</sup>. In all cases, including our present results, the  
610 fragments reported are at most a small fraction of the apparently full-length pORF1. Most of the time  
611 they were identified as faint bands on autoradiograms or Western-blot, showing that they are  
612 released in very low amounts. Collectively, these results suggest that the lower molecular weight  
613 fragments observed when HEV pORF1 is recombinantly expressed in different heterologous systems  
614 could correspond to artifacts, with spontaneous cleavage of the polyprotein in flexible loops.  
615 Interestingly, the 78-80 and 107 kDa fragments identified in several works<sup>21,26,65</sup> roughly correspond  
616 to the N- and C-terminal portions on the two sides of the long disordered HVR loop, *i.e.*, MetY – FABD-  
617 like and X – HEL – RdRp fragments, respectively.

618

619 Finally, Metzger et al.<sup>97</sup> reported that such bands can also be detected in an infection context,  
620 but still remain very minor and almost undetectable compared to the full-size pORF1. Such minor  
621 processing could be achieved by recruitment of a cellular protease acting on a minority of pORF1  
622 molecules, for example thrombin or factor Xa as suggested previously<sup>28,98</sup>. In this case, only the RdRp

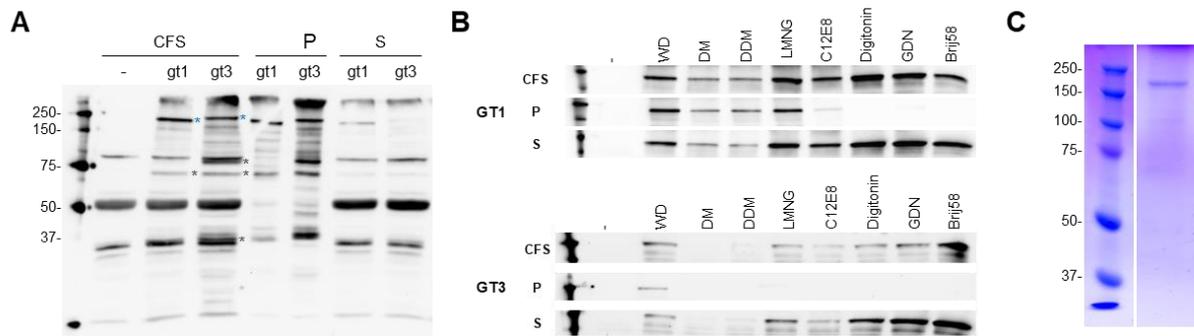
623 would be released, in very few amounts virtually undetectable. This would regulate the activity of the  
624 RNA polymerase, first to synthesize the negative sense RNA from the viral genome, then the switch to  
625 synthesize the positive sense progeny RNA. This cleavage would lead to two fragments of 130 and 57  
626 kDa that have been reported in one study <sup>20</sup>.

627         After centrifugation of the cell-free sample, the HEV pORF1 is found mainly in the insoluble  
628 fraction (Fig.8A, lanes P). The solubility of the protein can be improved by the addition of detergents  
629 within the translation mixture <sup>46,99</sup>. Therefore, we screened different detergents, and we found that  
630 the pORF1 solubility is strongly increased in presence of Digitonin, GDN or Brij58 for both genotypes 1  
631 and 3 (Fig.8B).

632         The GDN-solubilized gt1 HEV pORF1, tagged with both a Strep-tag and a His-tag, was then  
633 expressed in a large scale and purified through successive affinity chromatography steps, leading to a  
634 pure protein with an electrophoretic mobility compatible with the 187.5 kDa full-length polyprotein  
635 (Fig.8C). The integrity of the protein was assessed by tryptic digestion followed by a mass spectrometry  
636 analysis. Tryptic peptides cover 47% of the gt1 HEV pORF1 total sequence, from the first residue of the  
637 N-terminal Strep-tag to the residue 1679 (data not shown). Thus, we purified the full-length pORF1  
638 that represents the major form of the protein, the fragments observed by Western-blot being  
639 produced in negligible amounts. This observation contrasts to what we observed with other viral  
640 replication polyproteins that are spontaneously and quantitatively processed through their active  
641 protease <sup>46</sup>.

642         Although the quantity and the concentration of the pure protein are still too low to start the  
643 structural characterization, these results represent a major breakthrough to obtain structural  
644 information.

645



646

647 **Fig. 8. Expression and purification of HEV pORF1 replication polyprotein by wheat-germ**  
 648 **cell-free expression system.**

649 (A) Analysis of Strep-tagged gt1 and gt3 HEV pORF1 (s-ORF1) expression using wheat-germ  
 650 cell-free expression system. Protein samples were analyzed by SDS-PAGE followed by  
 651 Western-blot using anti Strep-tag antibodies. - is the negative control performed in  
 652 absence of transcript. CFS is the total translation reaction mixture, P and S are respectively  
 653 the insoluble and soluble proteins obtained after centrifugation during 30 min at 20,000  
 654 g of the CFS. S-B corresponds to the supernatant enriched through incubation with Strep-  
 655 Tactin magnetic beads. The stars indicate significant bands compared to the negative  
 656 control. Molecular weights markers are indicated on the left. (B) Analysis of s-ORF1  
 657 expression by wheat-germ cell-free expression system in absence (WD, without  
 658 detergent) or in presence of eight different detergents: DM, n-decyl-β-D-maltoside; DDM,  
 659 n-dodecyl-β-D-maltoside; LMNG, lauryl maltose neopentyl glycol; C12E8, dodecyl  
 660 octaethylene glycol ether; Digitonin; Brij58; GDN, glyco-diosgenin; Brij-58. Protein  
 661 samples were analyzed by Western-blot using anti Strep-tag antibodies. (C) Pure Strep-  
 662 and His-tagged gt1 HEV pORF1 (s-ORF1-h) obtained after expression and purification in  
 663 presence of GDN.  
 664

665 **4. CONCLUSION**

666 In this work, we have produced accurate *in silico* models defining the boundaries and 3D  
 667 structures of the five domains forming two modules separated by the HVR in the HEV replication  
 668 polyprotein pORF1. Analysis of these models partly confirmed what has been already published, except  
 669 for the N-terminal region. Indeed, the models reveal that the two previously reported Met and Y  
 670 domains are better considered a single MetY domain matching the alphavirus nsP1 protein, probably  
 671 up to its oligomerization as a membrane pore, although the membrane interactions would be distinctly  
 672 different and closer to some alphavirus-like plant viruses. We also report that there is no protease  
 673 domain encoded in HEV pORF1. These results allowed us to update the composition of HEV pORF1, in  
 674 terms of domains and boundaries. Nevertheless, we have limited our analysis to a molecular level and  
 675 the experimental structure is still required to allow the analysis at an atomic level. Moreover, we will

676 be able to understand the interactions between the different domains, as well as the dynamics of  
677 pORF1. To this goal, we have developed a protocol to express and purify the full-length HEV pORF1, a  
678 prerequisite step for structural study. . Our results confirm that the recombinant HEV pORF1 is not  
679 significantly self-processed, similarly to many unprocessed plant replication polyproteins<sup>100,101</sup> or the  
680 Alphavirus-like flock house virus (FHV) protein A<sup>102</sup>. Although we cannot rule out that during infection  
681 pORF1 may be partially processed by a cellular protease, we suggest that its functions may be  
682 regulated rather by structural flexibility.

683

## 684 **5. ACKNOWLEDGMENTS**

685 We thank Laurence Cocquerel and Cécile-Marie Aliouat for the gift of plasmids from which  
686 plasmids used in this work were constructed. Plasmid encoding benzonase was a generous gift from  
687 Wolfgang Wende. We thank Raphaël Guérois for sharing a standalone ColabFold implementation of  
688 AlphaFold2 and the I2BC integrative bioinformatics core facility BIOI2 for assistance with the high  
689 performance computing infrastructure. We thank Chantal Prévost for her help using heligeom and for  
690 the use of her flattening script on the pORF1 METY dodecamer. We thank David Karlin for helpful  
691 discussions and suggestions on the manuscript. This work has benefited from the facilities and  
692 expertise of the I2BC proteomic platform (Proteomic-Gif, SICaPS) supported by IBISA, Ile de France  
693 Region, Plan Cancer, CNRS and Paris-Saclay University. Regarding mass spectrometry experiments, this  
694 work was supported by the CNRS, the University of Strasbourg, the “Agence Nationale de la  
695 Recherche” and the French Proteomic Infrastructure (ProFI; ANR-10-INBS-08-03).

696

## 697 **6. DATA AVAILABILITY**

698 All models have been made available on Zenodo (DOI: <https://doi.org/10.5281/zenodo.7382752>). Our  
699 final pORF1 model and MetY dodecameric model are available as Supplementary Files 1 and 2.

700

## 701 **7. REFERENCES**

- 702 1. Smith, D.B., and Simmonds, P. (2018). Classification and Genomic Diversity of Enterically  
703 Transmitted Hepatitis Viruses. *Cold Spring Harb Perspect Med* 8, a031880.  
704 10.1101/cshperspect.a031880.
- 705 2. Smith, D.B., Izopet, J., Nicot, F., Simmonds, P., Jameel, S., Meng, X.-J., Norder, H., Okamoto,  
706 H., van der Poel, W.H.M., Reuter, G., et al. (2020). Update: proposed reference sequences for  
707 subtypes of hepatitis E virus (species Orthohepevirus A). *J. Gen. Virol.* 10.1099/jgv.0.001435.
- 708 3. Reuter, G., Boros, Á., and Pankovics, P. (2020). Review of Hepatitis E Virus in Rats: Evident  
709 Risk of Species Orthohepevirus C to Human Zoonotic Infection and Disease. *Viruses* 12, E1148.  
710 10.3390/v12101148.
- 711 4. Navaneethan, U., Al Mohajer, M., and Shata, M.T. (2008). Hepatitis E and pregnancy:  
712 understanding the pathogenesis. *Liver International* 28, 1190–1199. 10.1111/j.1478-  
713 3231.2008.01840.x.
- 714 5. Dalton, H.R., Kamar, N., van Eijk, J.J.J., Mclean, B.N., Cintas, P., Bendall, R.P., and Jacobs, B.C.  
715 (2016). Hepatitis E virus and neurological injury. *Nat Rev Neurol* 12, 77–85.  
716 10.1038/nrneurol.2015.234.
- 717 6. Kamar, N., Marion, O., Abravanel, F., Izopet, J., and Dalton, H.R. (2016). Extrahepatic  
718 manifestations of hepatitis E virus. *Liver International* 36, 467–472. 10.1111/liv.13037.
- 719 7. Zhu, F.-C., Zhang, J., Zhang, X.-F., Zhou, C., Wang, Z.-Z., Huang, S.-J., Wang, H., Yang, C.-L.,  
720 Jiang, H.-M., Cai, J.-P., et al. (2010). Efficacy and safety of a recombinant hepatitis E vaccine in healthy  
721 adults: a large-scale, randomised, double-blind placebo-controlled, phase 3 trial. *Lancet* 376, 895–  
722 902. 10.1016/S0140-6736(10)61030-6.
- 723 8. Debing, Y., Moradpour, D., Neyts, J., and Gouttenoire, J. (2016). Update on hepatitis E  
724 virology: Implications for clinical practice. *J. Hepatol.* 65, 200–212. 10.1016/j.jhep.2016.02.045.
- 725 9. Owen, D.R., Allerton, C.M.N., Anderson, A.S., Aschenbrenner, L., Avery, M., Berritt, S., Boras,  
726 B., Cardin, R.D., Carlo, A., Coffman, K.J., et al. (2021). An oral SARS-CoV-2 Mpro inhibitor clinical  
727 candidate for the treatment of COVID-19. *Science*. 10.1126/science.abl4784.
- 728 10. Rice, C.M., and Saeed, M. (2014). Hepatitis C: Treatment triumphs. *Nature* 510, 43–44.  
729 10.1038/510043a.
- 730 11. Tam, A.W., Smith, M.M., Guerra, M.E., Huang, C.-C., Bradley, D.W., Fry, K.E., and Reyes, G.R.  
731 (1991). Hepatitis E virus (HEV): Molecular cloning and sequencing of the full-length viral genome.  
732 *Virology* 185, 120–131. 10.1016/0042-6822(91)90760-9.
- 733 12. Koonin, E.V. (1991). The phylogeny of RNA-dependent RNA polymerases of positive-strand  
734 RNA viruses. *J. Gen. Virol.* 72 ( Pt 9), 2197–2206. 10.1099/0022-1317-72-9-2197.
- 735 13. Koonin, E.V., Gorbalenya, A.E., Purdy, M.A., Rozanov, M.N., Reyes, G.R., and Bradley, D.W.  
736 (1992). Computer-assisted assignment of functional domains in the nonstructural polyprotein of  
737 hepatitis E virus: delineation of an additional group of positive-strand RNA plant and animal viruses.  
738 *PNAS* 89, 8259–8263. 10.1073/pnas.89.17.8259.
- 739 14. Kenney, S.P., and Meng, X.-J. (2018). Hepatitis E Virus Genome Structure and Replication  
740 Strategy. *Cold Spring Harb Perspect Med*. 10.1101/cshperspect.a031724.

- 741 15. LeDesma, R., Nimgaonkar, I., and Ploss, A. (2019). Hepatitis E Virus Replication. *Viruses* *11*,  
742 719. 10.3390/v11080719.
- 743 16. Parvez, M.K. (2017). The hepatitis E virus nonstructural polyprotein. *Future Microbiol* *12*,  
744 915–924. 10.2217/fmb-2017-0016.
- 745 17. Wang, B., and Meng, X.-J. (2021). Structural and molecular biology of hepatitis E virus.  
746 *Comput Struct Biotechnol J* *19*, 1907–1916. 10.1016/j.csbj.2021.03.038.
- 747 18. Ansari, I.H., Nanda, S.K., Durgapal, H., Agrawal, S., Mohanty, S.K., Gupta, D., Jameel, S., and  
748 Panda, S.K. (2000). Cloning, sequencing, and expression of the hepatitis E virus (HEV) nonstructural  
749 open reading frame 1 (ORF1). *J. Med. Virol.* *60*, 275–283.
- 750 19. Ju, X., Xiang, G., Gong, M., Yang, R., Qin, J., Li, Y., Nan, Y., Yang, Y., Zhang, Q.C., and Ding, Q.  
751 (2020). Identification of functional cis-acting RNA elements in the hepatitis E virus genome required  
752 for viral replication. *PLoS Pathog* *16*, e1008488. 10.1371/journal.ppat.1008488.
- 753 20. Perttilä, J., Spuul, P., and Ahola, T. (2013). Early secretory pathway localization and lack of  
754 processing for hepatitis E virus replication protein pORF1. *Journal of General Virology* *94*, 807–816.  
755 10.1099/vir.0.049577-0.
- 756 21. Ropp, S.L., Tam, A.W., Beames, B., Purdy, M., and Frey, T.K. (2000). Expression of the  
757 hepatitis E virus ORF1. *Arch. Virol.* *145*, 1321–1337.
- 758 22. Suppiah, S., Zhou, Y., and Frey, T.K. (2011). Lack of processing of the expressed ORF1 gene  
759 product of hepatitis E virus. *Virol J* *8*, 245. 10.1186/1743-422X-8-245.
- 760 23. Szkolnicka, D., Pollán, A., Da Silva, N., Oechslin, N., Gouttenoire, J., and Moradpour, D. (2019).  
761 Recombinant Hepatitis E Viruses Harboring Tags in the ORF1 Protein. *J. Virol.* *93*. 10.1128/JVI.00459-  
762 19.
- 763 24. Panda, S.K., Ansari, I.H., Durgapal, H., Agrawal, S., and Jameel, S. (2000). The in vitro-  
764 synthesized RNA from a cDNA clone of hepatitis E virus is infectious. *Journal of Virology* *74*, 2430–  
765 2437. 10.1128/JVI.74.5.2430-2437.2000.
- 766 25. Sehgal, D., Thomas, S., Chakraborty, M., and Jameel, S. (2006). Expression and processing of  
767 the Hepatitis E virus ORF1 nonstructural polyprotein. *Virology Journal* *3*, 38. 10.1186/1743-422X-3-  
768 38.
- 769 26. Parvez, M.K. (2013). Molecular characterization of hepatitis E virus ORF1 gene supports a  
770 papain-like cysteine protease (PCP)-domain activity. *Virus Res.* *178*, 553–556.  
771 10.1016/j.virusres.2013.07.020.
- 772 27. Kumar, M., Hooda, P., Khanna, M., Patel, U., and Sehgal, D. (2020). Development of BacMam  
773 Induced Hepatitis E Virus Replication Model in Hepatoma Cells to Study the Polyprotein Processing.  
774 *Front Microbiol* *11*, 1347. 10.3389/fmicb.2020.01347.
- 775 28. Paliwal, D., Panda, S.K., Kapur, N., Varma, S.P.K., and Durgapal, H. (2014). Hepatitis E virus  
776 (HEV) protease: a chymotrypsin-like enzyme that processes both non-structural (pORF1) and capsid  
777 (pORF2) protein. *J Gen Virol* *95*, 1689–1700. 10.1099/vir.0.066142-0.

- 778 29. Proudfoot, A., Hyrina, A., Holdorf, M., Frank, A.O., and Bussiere, D. (2019). First Crystal  
779 Structure of a Nonstructural Hepatitis E Viral Protein Identifies a Putative Novel Zinc-Binding Protein.  
780 *J. Virol.* *93*. 10.1128/JVI.00170-19.
- 781 30. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022).  
782 ColabFold: making protein folding accessible to all. *Nat Methods* *19*, 679–682. 10.1038/s41592-022-  
783 01488-1.
- 784 31. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool,  
785 K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with  
786 AlphaFold. *Nature* *596*, 583–589. 10.1038/s41586-021-03819-2.
- 787 32. Remmert, M., Biegert, A., Hauser, A., and Söding, J. (2012). HHblits: lightning-fast iterative  
788 protein sequence searching by HMM-HMM alignment. *Nat Methods* *9*, 173–175.  
789 10.1038/nmeth.1818.
- 790 33. Gabler, F., Nam, S.-Z., Till, S., Mirdita, M., Steinegger, M., Söding, J., Lupas, A.N., and Alva, V.  
791 (2020). Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Curr Protoc Bioinformatics*  
792 *72*, e108. 10.1002/cpbi.108.
- 793 34. Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J.,  
794 Lupas, A.N., and Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New  
795 HHpred Server at its Core. *Journal of Molecular Biology* *430*, 2237–2243. 10.1016/j.jmb.2017.12.007.
- 796 35. Steinegger, M., and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching  
797 for the analysis of massive data sets. *Nat Biotechnol* *35*, 1026–1028. 10.1038/nbt.3988.
- 798 36. Croll, T.I. (2018). ISOLDE: a physically realistic environment for model building into low-  
799 resolution electron-density maps. *Acta Cryst D* *74*, 519–530. 10.1107/S2059798318002425.
- 800 37. Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin,  
801 T.E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science*  
802 *27*, 14–25. 10.1002/pro.3235.
- 803 38. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H.,  
804 and Ferrin, T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and  
805 developers. *Protein Science* *30*, 70–82. 10.1002/pro.3943.
- 806 39. Boyer, B., Ezelin, J., Poulain, P., Saladin, A., Zacharias, M., Robert, C.H., and Prévost, C. (2015).  
807 An Integrative Approach to the Study of Filamentous Oligomeric Assemblies, with Application to  
808 RecA. *PLOS ONE* *10*, e0116414. 10.1371/journal.pone.0116414.
- 809 40. Tran, L., Basdevant, N., Prévost, C., and Ha-Duong, T. (2016). Structure of ring-shaped Aβ42  
810 oligomers determined by conformational selection. *Sci Rep* *6*, 21429. 10.1038/srep21429.
- 811 41. Zhang, Y., and Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on  
812 the TM-score. *Nucleic Acids Research* *33*, 2302–2309. 10.1093/nar/gki524.
- 813 42. Tomasello, G., Armenia, I., and Molla, G. (2020). The Protein Imager: a full-featured online  
814 molecular viewer interface with server-side HQ-rendering capabilities. *Bioinformatics* *36*, 2909–2911.  
815 10.1093/bioinformatics/btaa009.

- 816 43. Emerson, S.U., Zhang, M., Meng, X.J., Nguyen, H., St Claire, M., Govindarajan, S., Huang, Y.K.,  
817 and Purcell, R.H. (2001). Recombinant hepatitis E virus genomes infectious for primates: importance  
818 of capping and discovery of a cis-reactive element. *Proc Natl Acad Sci U S A* 98, 15270–15275.  
819 10.1073/pnas.251555098.
- 820 44. Shukla, P., Nguyen, H.T., Faulk, K., Mather, K., Torian, U., Engle, R.E., and Emerson, S.U.  
821 (2012). Adaptation of a genotype 3 hepatitis E virus to efficient growth in cell culture depends on an  
822 inserted human gene segment acquired by recombination. *J Virol* 86, 5697–5707. 10.1128/JVI.00146-  
823 12.
- 824 45. Takai, K., Sawasaki, T., and Endo, Y. (2010). Practical cell-free protein synthesis system using  
825 purified wheat embryos. *Nat Protoc* 5, 227–238. 10.1038/nprot.2009.207.
- 826 46. Habersetzer, J., Debbah, M., Fogeron, M.-L., Böckmann, A., Bressanelli, S., and Fieulaine, S.  
827 (2020). In vitro translation of virally-encoded replication polyproteins to recapitulate polyprotein  
828 maturation processes. *Protein Expr. Purif.* 175, 105694. 10.1016/j.pep.2020.105694.
- 829 47. Fogeron, M.-L., Badillo, A., Penin, F., and Böckmann, A. (2017). Wheat Germ Cell-Free  
830 Overexpression for the Production of Membrane Proteins. *Methods Mol. Biol.* 1635, 91–108.  
831 10.1007/978-1-4939-7151-0\_5.
- 832 48. Herbert, C.J., Labarre-Mariotte, S., Cornu, D., Sophie, C., Panozzo, C., Michel, T., Dujardin, G.,  
833 and Bonnefoy, N. (2021). Translational activators and mitoribosomal isoforms cooperate to mediate  
834 mRNA-specific translation in *Schizosaccharomyces pombe* mitochondria. *Nucleic Acids Res* 49,  
835 11145–11166. 10.1093/nar/gkab789.
- 836 49. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie,  
837 A., Meyer, C., Laydon, A., et al. (2021). Highly accurate protein structure prediction for the human  
838 proteome. *Nature* 596, 590–596. 10.1038/s41586-021-03828-1.
- 839 50. Kryshchak, A., Moutl, J., Albrecht, R., Chang, G.A., Chao, K., Fraser, A., Greenfield, J.,  
840 Hartmann, M.D., Herzberg, O., Josts, I., et al. (2021). Computational models in the service of X-ray  
841 and cryo-electron microscopy structure determination. *Proteins* 89, 1633–1646. 10.1002/prot.26223.
- 842 51. Wang, H., Xiao, Y., Chen, X., Zhang, M., Sun, G., Wang, F., Wang, L., Zhang, H., Zhang, X.,  
843 Yang, X., et al. (2022). Crystal Structures of *Wolbachia* CidA and CidB Reveal Determinants of  
844 Bacteria-induced Cytoplasmic Incompatibility and Rescue. *Nat Commun* 13, 1608. 10.1038/s41467-  
845 022-29273-w.
- 846 52. Goulet, A., Cambillau, C., Roussel, A., and Imbert, I. (2022). Structure Prediction and Analysis  
847 of Hepatitis E Virus Non-Structural Proteins from the Replication and Transcription Machinery by  
848 AlphaFold2. *Viruses* 14, 1537. 10.3390/v14071537.
- 849 53. Purdy, M.A. (2012). Evolution of the hepatitis E virus polyproline region: order from disorder.  
850 *J. Virol.* 86, 10186–10193. 10.1128/JVI.01374-12.
- 851 54. Purdy, M.A., Lara, J., and Khudyakov, Y.E. (2012). The Hepatitis E Virus Polyproline Region Is  
852 Involved in Viral Adaptation. *PLOS ONE* 7, e35974. 10.1371/journal.pone.0035974.
- 853 55. Johne, R., Reetz, J., Ulrich, R.G., Machnowska, P., Sachsenröder, J., Nickel, P., and Hofmann, J.  
854 (2014). An ORF1-rearranged hepatitis E virus derived from a chronically infected patient efficiently  
855 replicates in cell culture. *J Viral Hepat* 21, 447–456. 10.1111/jvh.12157.

- 856 56. Muñoz-Chimeno, M., Cenalmor, A., Garcia-Lugo, M.A., Hernandez, M., Rodriguez-Lazaro, D.,  
857 and Avellon, A. (2020). Proline-Rich Hypervariable Region of Hepatitis E Virus: Arranging the Disorder.  
858 *Microorganisms* 8, 1417. 10.3390/microorganisms8091417.
- 859 57. Nguyen, H.T., Torian, U., Faulk, K., Mather, K., Engle, R.E., Thompson, E., Bonkovsky, H.L., and  
860 Emerson, S.U. (2012). A naturally occurring human/hepatitis E recombinant virus predominates in  
861 serum but not in faeces of a chronic hepatitis E patient and has a growth advantage in cell culture. *J*  
862 *Gen Virol* 93, 526–530. 10.1099/vir.0.037259-0.
- 863 58. Holm, L. (2022). Dali server: structural unification of protein families. *Nucleic Acids Res*,  
864 *gkac387*. 10.1093/nar/gkac387.
- 865 59. Holm, L., and Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends*  
866 *Biochem Sci* 20, 478–480. 10.1016/s0968-0004(00)89105-7.
- 867 60. Holm, L. (2020). Using Dali for Protein Structure Comparison. *Methods Mol Biol* 2112, 29–42.  
868 10.1007/978-1-0716-0270-6\_3.
- 869 61. Ahola, T., and Karlin, D.G. (2015). Sequence analysis reveals a conserved extension in the  
870 capping enzyme of the alphavirus supergroup, and a homologous domain in nodaviruses. *Biol. Direct*  
871 10, 16. 10.1186/s13062-015-0050-0.
- 872 62. Parvez, M.K. (2017). Mutational analysis of hepatitis E virus ORF1 “Y-domain”: Effects on RNA  
873 replication and virion infectivity. *World J Gastroenterol* 23, 590–602. 10.3748/wjg.v23.i4.590.
- 874 63. Hooda, P., Ishtikhar, M., Saraswat, S., Bhatia, P., Mishra, D., Trivedi, A., Kulandaisamy, R.,  
875 Aggarwal, S., Munde, M., Ali, N., et al. (2022). Biochemical and Biophysical Characterisation of the  
876 Hepatitis E Virus Guanine-7-Methyltransferase. *Molecules* 27, 1505. 10.3390/molecules27051505.
- 877 64. Rozanov, M.N., Koonin, E.V., and Gorbalenya, A.E. (1992). Conservation of the putative  
878 methyltransferase domain: a hallmark of the “Sindbis-like” supergroup of positive-strand RNA  
879 viruses. *J. Gen. Virol.* 73 ( Pt 8), 2129–2134. 10.1099/0022-1317-73-8-2129.
- 880 65. Magden, J., Takeda, N., Li, T., Auvinen, P., Ahola, T., Miyamura, T., Merits, A., and Kääriäinen,  
881 L. (2001). Virus-specific mRNA capping enzyme encoded by hepatitis E virus. *J Virol* 75, 6249–6255.  
882 10.1128/JVI.75.14.6249-6255.2001.
- 883 66. Jones, R., Bragagnolo, G., Arranz, R., and Reguera, J. (2021). Capping pores of alphavirus nsP1  
884 gate membranous viral replication factories. *Nature* 589, 615–619. 10.1038/s41586-020-3036-8.
- 885 67. Zhang, K., Law, Y.-S., Law, M.C.Y., Tan, Y.B., Wirawan, M., and Luo, D. (2021). Structural  
886 insights into viral RNA capping and plasma membrane targeting by Chikungunya virus nonstructural  
887 protein 1. *Cell Host & Microbe* 29, 757-764.e3. 10.1016/j.chom.2021.02.018.
- 888 68. Moriceau, L., Jomat, L., Bressanelli, S., Alcaide-Loridan, C., and Jupin, I. (2017). Identification  
889 and Molecular Characterization of the Chloroplast Targeting Domain of Turnip yellow mosaic virus  
890 Replication Proteins. *Front Plant Sci* 8, 2138. 10.3389/fpls.2017.02138.
- 891 69. Sathanantham, P., Zhao, W., He, G., Murray, A., Fenech, E., Diaz, A., Schuldiner, M., and  
892 Wang, X. (2022). A conserved viral amphipathic helix governs the replication site-specific membrane  
893 association. *PLOS Pathogens* 18, e1010752. 10.1371/journal.ppat.1010752.

- 894 70. Smathers, R.L., and Petersen, D.R. (2011). The human fatty acid-binding protein family:  
895 Evolutionary divergences and functions. *Human Genomics* 5, 170. 10.1186/1479-7364-5-3-170.
- 896 71. Marcelino, A.M.C., Smock, R.G., and Gierasch, L.M. (2006). Evolutionary coupling of structural  
897 and functional sequence information in the intracellular lipid-binding protein family. *Proteins* 63,  
898 373–384. 10.1002/prot.20860.
- 899 72. Proudfoot, A., Hyrina, A., Holdorf, M., Frank, A.O., and Bussiere, D. (2019). First Crystal  
900 Structure of a Nonstructural Hepatitis E Viral Protein Identifies a Putative Novel Zinc-Binding Protein.  
901 *Journal of Virology* 93, e00170-19. 10.1128/JVI.00170-19.
- 902 73. Szlávik, Z., Ondi, L., Csékei, M., Paczal, A., Szabó, Z.B., Radics, G., Murray, J., Davidson, J.,  
903 Chen, I., Davis, B., et al. (2019). Structure-Guided Discovery of a Selective Mcl-1 Inhibitor with Cellular  
904 Activity. *J Med Chem* 62, 6913–6924. 10.1021/acs.jmedchem.9b00134.
- 905 74. Piotrowski, Y., Hansen, G., Boomaars-van der Zanden, A.L., Snijder, E.J., Gorbalenya, A.E., and  
906 Hilgenfeld, R. (2009). Crystal structures of the X-domains of a Group-1 and a Group-3 coronavirus  
907 reveal that ADP-ribose-binding may not be a conserved property. *Protein Science* 18, 6–16.  
908 10.1002/pro.15.
- 909 75. Parvez, M.K. (2015). The hepatitis E virus ORF1 ‘X-domain’ residues form a putative  
910 macrodomain protein/Appr-1”-pase catalytic-site, critical for viral RNA replication. *Gene* 566, 47–53.  
911 10.1016/j.gene.2015.04.026.
- 912 76. Vikram, T., and Kumar, P. (2018). Analysis of Hepatitis E virus (HEV) X-domain structural  
913 model. *Bioinformatics* 14, 398–403. 10.6026/97320630014398.
- 914 77. Anang, S., Subramani, C., Nair, V.P., Kaul, S., Kaushik, N., Sharma, C., Tiwari, A., Ranjith-  
915 Kumar, C.T., and Surjit, M. (2016). Identification of critical residues in Hepatitis E virus macro domain  
916 involved in its interaction with viral methyltransferase and ORF3 proteins. *Sci Rep* 6, 25133.  
917 10.1038/srep25133.
- 918 78. Huang, H., Zheng, Z.-Z., Zhao, M., Li, J.-X., Lai, W.-S., Miao, J., Zhang, J., and Xia, N.-S. (2011).  
919 [Localization of functional domains of HEV ORF1 in cells]. *Bing Du Xue Bao* 27, 195–201.
- 920 79. Piotrowski, Y., Hansen, G., Boomaars-van der Zanden, A.L., Snijder, E.J., Gorbalenya, A.E., and  
921 Hilgenfeld, R. (2009). Crystal structures of the X-domains of a Group-1 and a Group-3 coronavirus  
922 reveal that ADP-ribose-binding may not be a conserved property. *Protein Science* 18, 6–16.  
923 10.1002/pro.15.
- 924 80. Kadaré, G., and Haenni, A.L. (1997). Virus-encoded RNA helicases. *J. Virol.* 71, 2583–2590.
- 925 81. Karpe, Y.A., and Lole, K.S. (2010). NTPase and 5’ to 3’ RNA duplex-unwinding activities of the  
926 hepatitis E virus helicase domain. *J Virol* 84, 3595–3602. 10.1128/JVI.02130-09.
- 927 82. Mhaindarkar, V., Sharma, K., and Lole, K.S. (2014). Mutagenesis of hepatitis E virus helicase  
928 motifs: Effects on enzyme activity. *Virus Research* 179, 26–33. 10.1016/j.virusres.2013.11.022.
- 929 83. Gorbalenya, A.E., and Koonin, E.V. (1989). Viral proteins containing the purine NTP-binding  
930 sequence pattern. *Nucleic Acids Res.* 17, 8413–8440.

- 931 84. Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007). Structure and mechanism of  
932 helicases and nucleic acid translocases. *Annu. Rev. Biochem.* 76, 23–50.  
933 10.1146/annurev.biochem.76.052305.115300.
- 934 85. Cao, D., Ni, Y.-Y., and Meng, X.-J. (2018). Substitution of amino acid residue V1213 in the  
935 helicase domain of the genotype 3 hepatitis E virus reduces virus replication. *Viol J* 15, 32.  
936 10.1186/s12985-018-0943-5.
- 937 86. Devhare, P., Sharma, K., Mhaindarkar, V., Arankalle, V., and Lole, K. (2014). Analysis of  
938 helicase domain mutations in the hepatitis E virus derived from patients with fulminant hepatic  
939 failure: effects on enzymatic activities and virus replication. *Virus Res* 184, 103–110.  
940 10.1016/j.virusres.2014.02.018.
- 941 87. Nishikiori, M., Sugiyama, S., Xiang, H., Niiyama, M., Ishibashi, K., Inoue, T., Ishikawa, M.,  
942 Matsumura, H., and Katoh, E. (2012). Crystal Structure of the Superfamily 1 Helicase from Tomato  
943 Mosaic Virus. *Journal of Virology* 86, 7565–7576. 10.1128/JVI.00118-12.
- 944 88. Law, Y.-S., Utt, A., Tan, Y.B., Zheng, J., Wang, S., Chen, M.W., Griffin, P.R., Merits, A., and Luo,  
945 D. (2019). Structural insights into RNA recognition by the Chikungunya virus nsP2 helicase.  
946 *Proceedings of the National Academy of Sciences* 116, 9558–9567. 10.1073/pnas.1900656116.
- 947 89. Deng, Z., Lehmann, K.C., Li, X., Feng, C., Wang, G., Zhang, Q., Qi, X., Yu, L., Zhang, X., Feng, W.,  
948 et al. (2014). Structural basis for the regulatory function of a complex zinc-binding domain in a  
949 replicative arterivirus helicase resembling a nonsense-mediated mRNA decay helicase. *Nucleic Acids*  
950 *Research* 42, 3464–3477. 10.1093/nar/gkt1310.
- 951 90. Tan, Y.B., Lello, L.S., Liu, X., Law, Y.-S., Kang, C., Lescar, J., Zheng, J., Merits, A., and Luo, D.  
952 (2022). Crystal structures of alphavirus nonstructural protein 4 (nsP4) reveal an intrinsically dynamic  
953 RNA-dependent RNA polymerase fold. *Nucleic Acids Research* 50, 1000–1016.  
954 10.1093/nar/gkab1302.
- 955 91. Bressanelli, S., Tomei, L., Roussel, A., Incitti, I., Vitale, R.L., Mathieu, M., De Francesco, R., and  
956 Rey, F.A. (1999). Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Proc.*  
957 *Natl. Acad. Sci. U.S.A.* 96, 13034–13039.
- 958 92. Caillet-Saguy, C., Lim, S.P., Shi, P.-Y., Lescar, J., and Bressanelli, S. (2014). Polymerases of  
959 hepatitis C viruses and flaviviruses: structural and mechanistic insights and drug development.  
960 *Antiviral Res.* 105, 8–16. 10.1016/j.antiviral.2014.02.006.
- 961 93. Metzger, K., Bentaleb, C., Hervouet, K., Alexandre, V., Montpellier, C., Saliou, J.-M., Ferrié,  
962 M., Camuzet, C., Rouillé, Y., Lecoq, C., et al. (2022). Processing and Subcellular Localization of the  
963 Hepatitis E Virus Replicase: Identification of Candidate Viral Factories. *Frontiers in Microbiology* 13.
- 964 94. Bressanelli, S., Tomei, L., Rey, F.A., and De Francesco, R. (2002). Structural analysis of the  
965 hepatitis C virus RNA polymerase in complex with ribonucleotides. *J. Virol* 76, 3482–3492.
- 966 95. Shin, G., Yost, S.A., Miller, M.T., Elrod, E.J., Grakoui, A., and Marcotrigiano, J. (2012).  
967 Structural and functional insights into alphavirus polyprotein processing and pathogenesis. *PNAS* 109,  
968 16534–16539. 10.1073/pnas.1210418109.
- 969 96. Fogeron, M.-L., Lecoq, L., Cole, L., Harbers, M., and Böckmann, A. (2021). Easy Synthesis of  
970 Complex Biomolecular Assemblies: Wheat Germ Cell-Free Protein Expression in Structural Biology.  
971 *Front Mol Biosci* 8, 639587. 10.3389/fmolb.2021.639587.

972 97. Metzger, K., Bentaleb, C., Hervouet, K., Alexandre, V., Montpellier, C., Saliou, J.-M., Ferrié,  
973 M., Camuzet, C., Rouillé, Y., Lecoœur, C., et al. (2022). Processing and Subcellular Localization of the  
974 Hepatitis E Virus Replicase: Identification of Candidate Viral Factories. *Frontiers in Microbiology* *13*.

975 98. Kanade, G.D., Pingale, K.D., and Karpe, Y.A. (2018). Activities of Thrombin and Factor Xa Are  
976 Essential for Replication of Hepatitis E Virus and Are Possibly Implicated in ORF1 Polyprotein  
977 Processing. *J Virol* *92*, e01853-17. 10.1128/JVI.01853-17.

978 99. Fogeron, M.-L., Badillo, A., Jirasko, V., Gouttenoire, J., Paul, D., Lancien, L., Moradpour, D.,  
979 Bartenschlager, R., Meier, B.H., Penin, F., et al. (2015). Wheat germ cell-free expression: Two  
980 detergents with a low critical micelle concentration allow for production of soluble HCV membrane  
981 proteins. *Protein Expr. Purif.* *105*, 39–46. 10.1016/j.pep.2014.10.003.

982 100. van der Heijden, M.W., and Bol, J.F. (2002). Composition of alphavirus-like replication  
983 complexes: involvement of virus and host encoded proteins. *Arch. Virol.* *147*, 875–898.  
984 10.1007/s00705-001-0773-3.

985 101. Martelli, G.P., Adams, M.J., Kreuze, J.F., and Dolja, V.V. (2007). Family Flexiviridae: a case  
986 study in virion and genome plasticity. *Annu Rev Phytopathol* *45*, 73–100.  
987 10.1146/annurev.phyto.45.062806.094401.

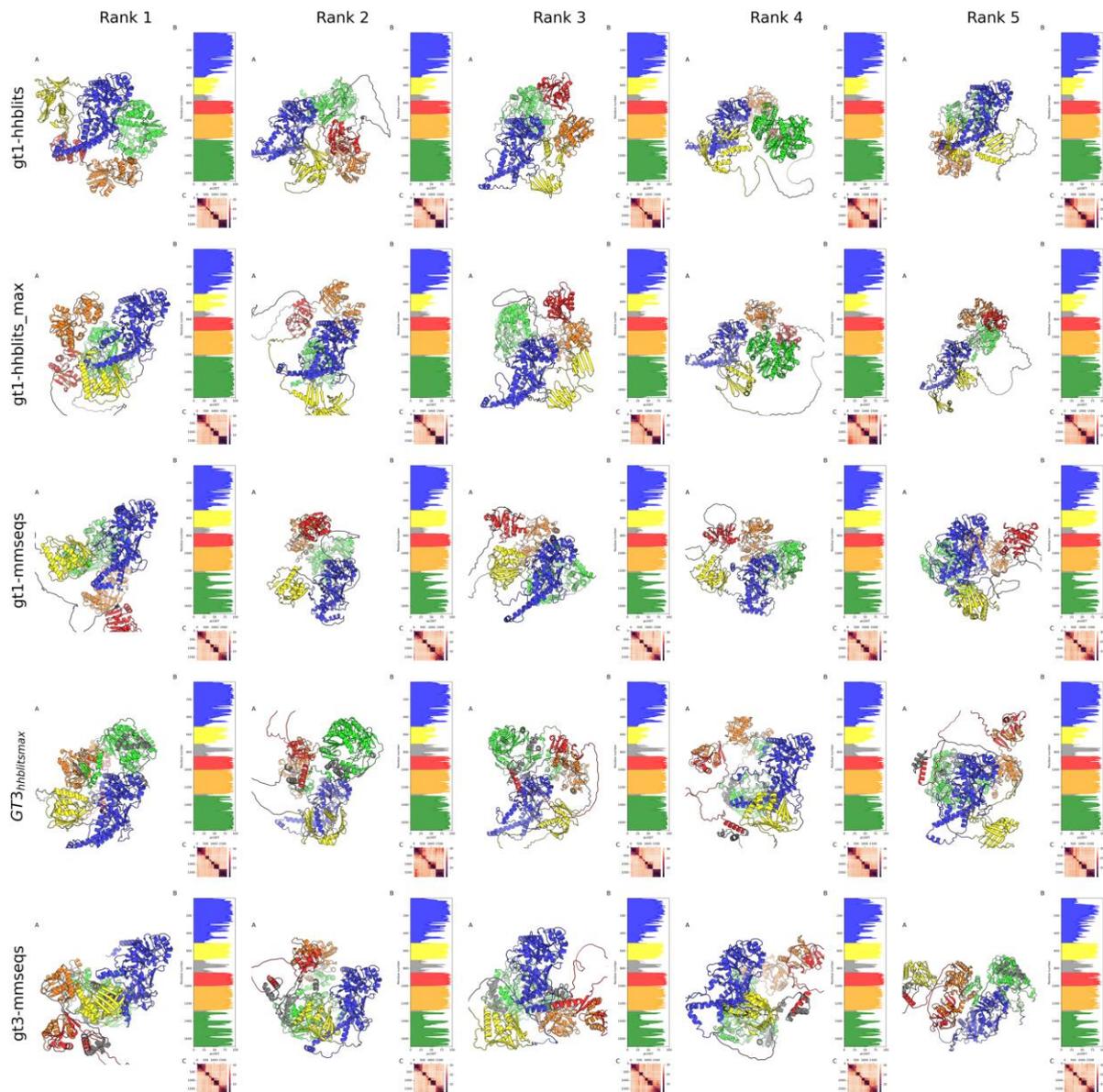
988 102. Venter, P.A., and Schneemann, A. (2008). Recent insights into the biology and biomedical  
989 applications of Flock House virus. *Cell. Mol. Life Sci.* *65*, 2675–2687. 10.1007/s00018-008-8037-y.

990 103. Schrödinger, LLC (2015). The PyMOL Molecular Graphics System, Version 2.5.

991

992

993



995

996

997 **Fig. S1. AlphaFold2 models of gt1 and gt3 HEV ORF1 replication polyprotein.**

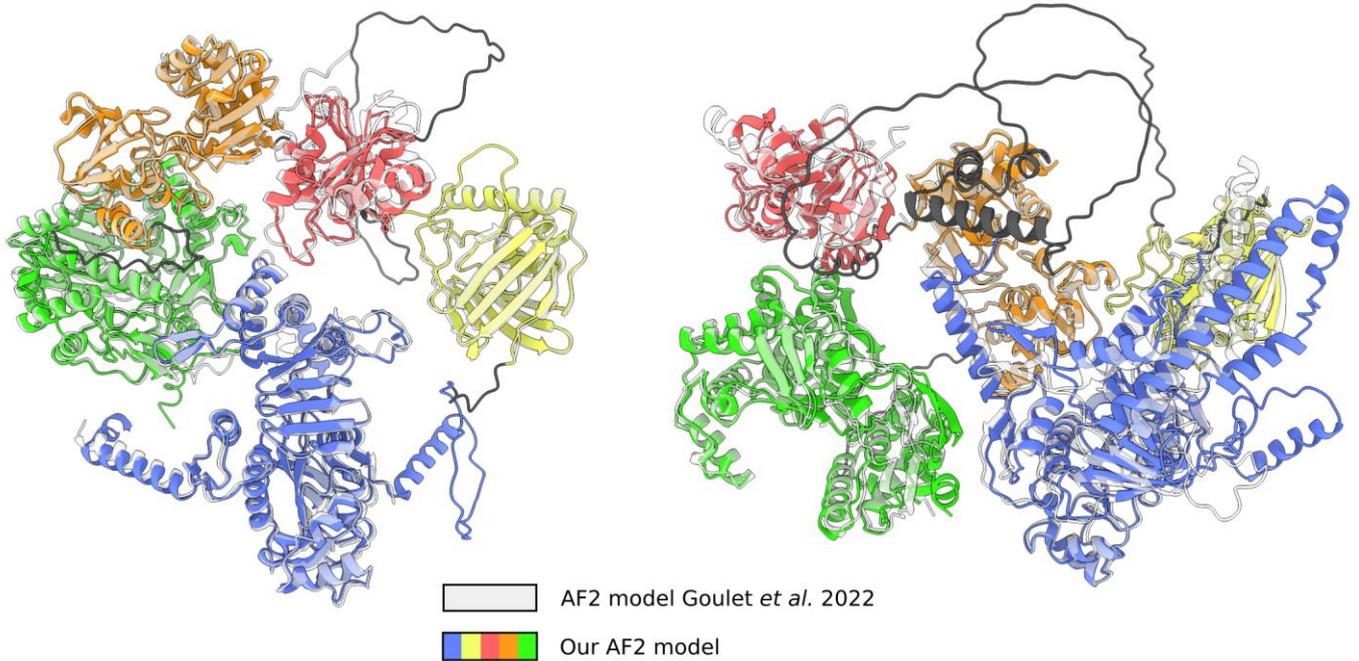
998 AlphaFold2 models of gt1 and gt3 (A) with corresponding pLDDT (B) and PAE matrix (C). A and B are  
 999 colored according to their domain decompositions: MetY in blue, FABD-like in yellow, X-  
 1000 macrodomain in red, helicase in orange and polymerase in green. Linkers and the hypervariable  
 1001 regions are colored in gray. All models were aligned using TM-align<sup>41</sup> with its dedicated PyMOL Addon.  
 1002 Structure rendered with PyMOL<sup>103</sup>, charts in Python.

1003

1004

gt1 vs. gt3 Goulet *et al.* 2022

gt3 vs. gt3 Goulet *et al.* 2022



1005

1006 **Fig. S2. Comparison between our AlphaFold2 complete model with the one obtained by Goulet *et***  
1007 ***al.***

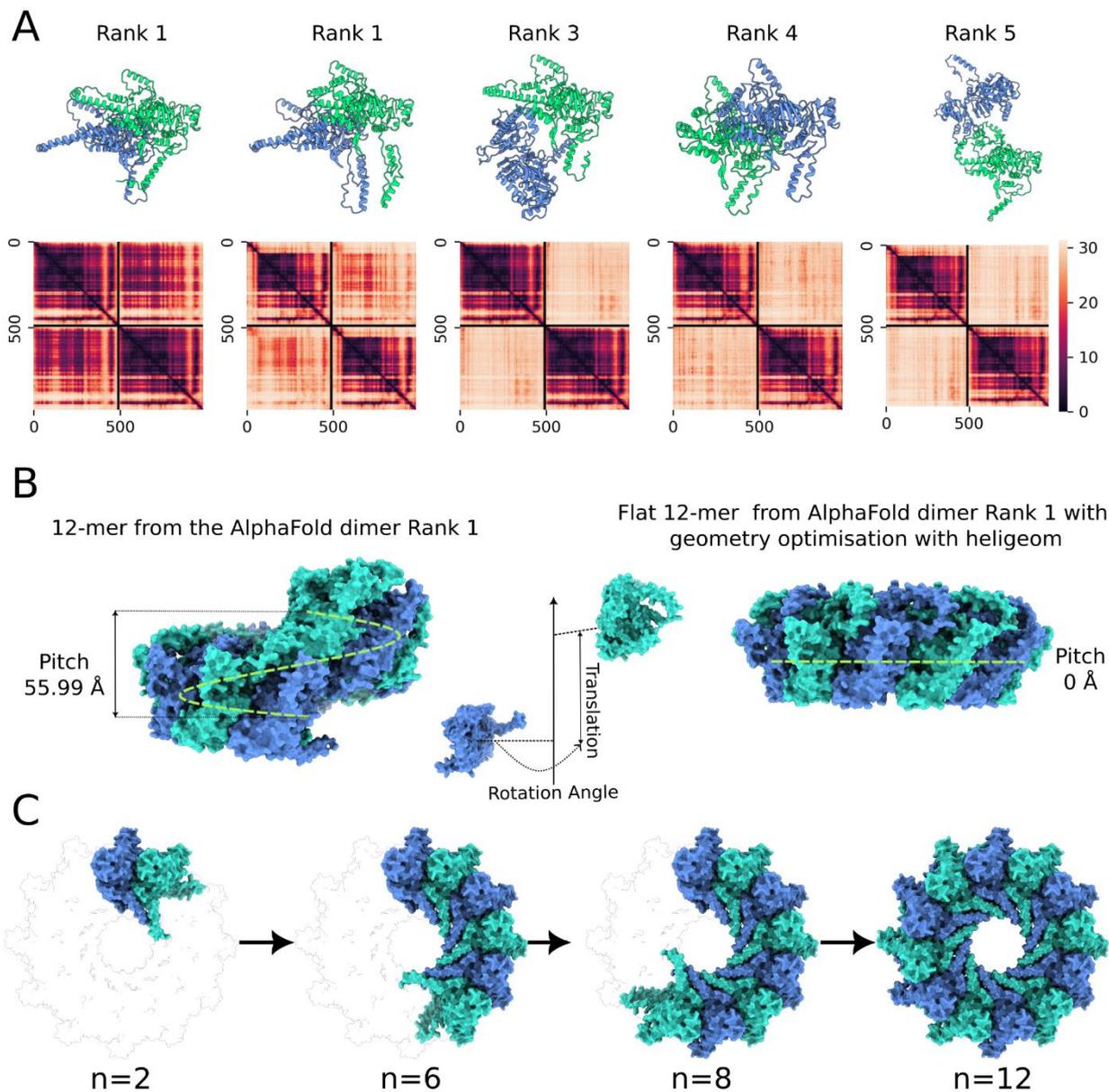
1008 Goulet *et al.* individual models<sup>52</sup> are represented with a gray semi-transparent cartoon while our gt1  
1009 and gt3 AlphaFold2 complete models are colored per domain: MetY in blue, FABD-like in yellow, X-  
1010 macrodomain in red, HEL in orange and RdRp in green. Rendered with ChimeraX<sup>37,38</sup>.

1011

1012

1013

1014

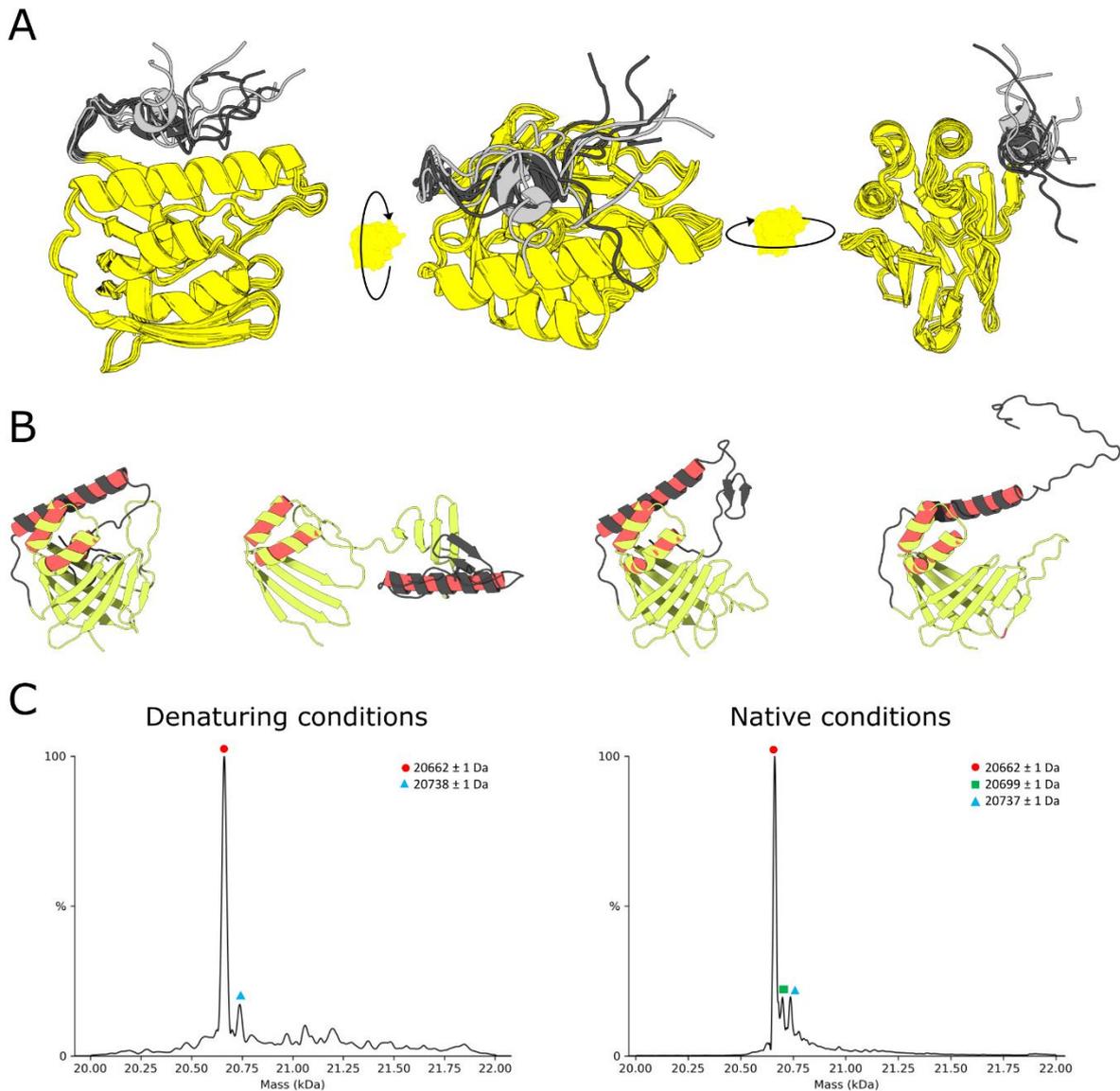


1015 **Fig. S3. MetY dimer modelization and its hypothetical dodecamerization.**

1016 (A) MetY dimer models with respective PAE matrices. Chain A is blue and B in green. Models aligned  
 1017 on chain A with TM-align<sup>41</sup> with its dedicated PyMoL Addon, rendered with ChimeraX<sup>37,38</sup> and trimed  
 1018 using Python. (B) Definition of pitch, translation and rotation angle on the rank 1 alphafold dimer(C)  
 1019 Example of dodecamerization with 2, 6, 8 and 12 sub-units. Rendered with ChimeraX<sup>37,38</sup>.

1020

1021

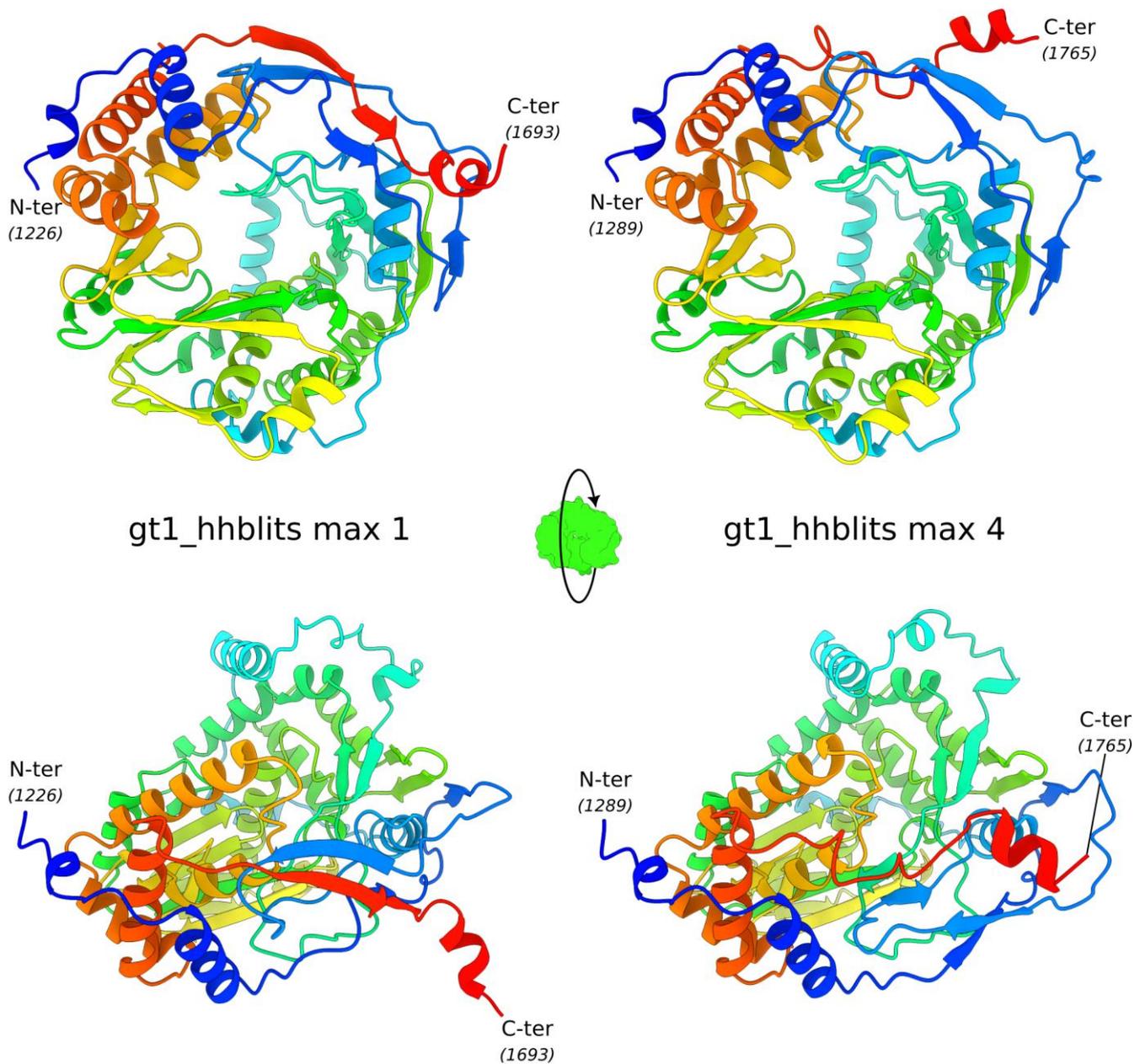


1022

1023 **Fig. S4. Flexibility of the FABD-like domain.**

1024 (A) Position of the C-terminal (691-707) region of the FABD-like in light gray for gt1 and dark gray for  
 1025 gt3 (MMSEQS models only). Rendered with PyMol<sup>103</sup>. (B) Different FABD-like conformation made with  
 1026 AF2. Rendered with Protein Imager<sup>42</sup>. (C) Deconvoluted mass spectrum of gt3 HEV FABD-like domain  
 1027 in denaturing conditions (left) and in native conditions (right). The main population (in red)  
 1028 corresponds to the expected mass of the protein without any ligand (< 10 ppm in denaturing  
 1029 conditions, left; < 30 ppm in native conditions, right). Additional masses with an increment of 76 Da (in  
 1030 blue, left) or 37 and 76 Da (in green and blue, respectively, right) were detected, not corresponding to  
 1031 a metal ion (expected mass of 65-67 Da for a Zn ion).

1032

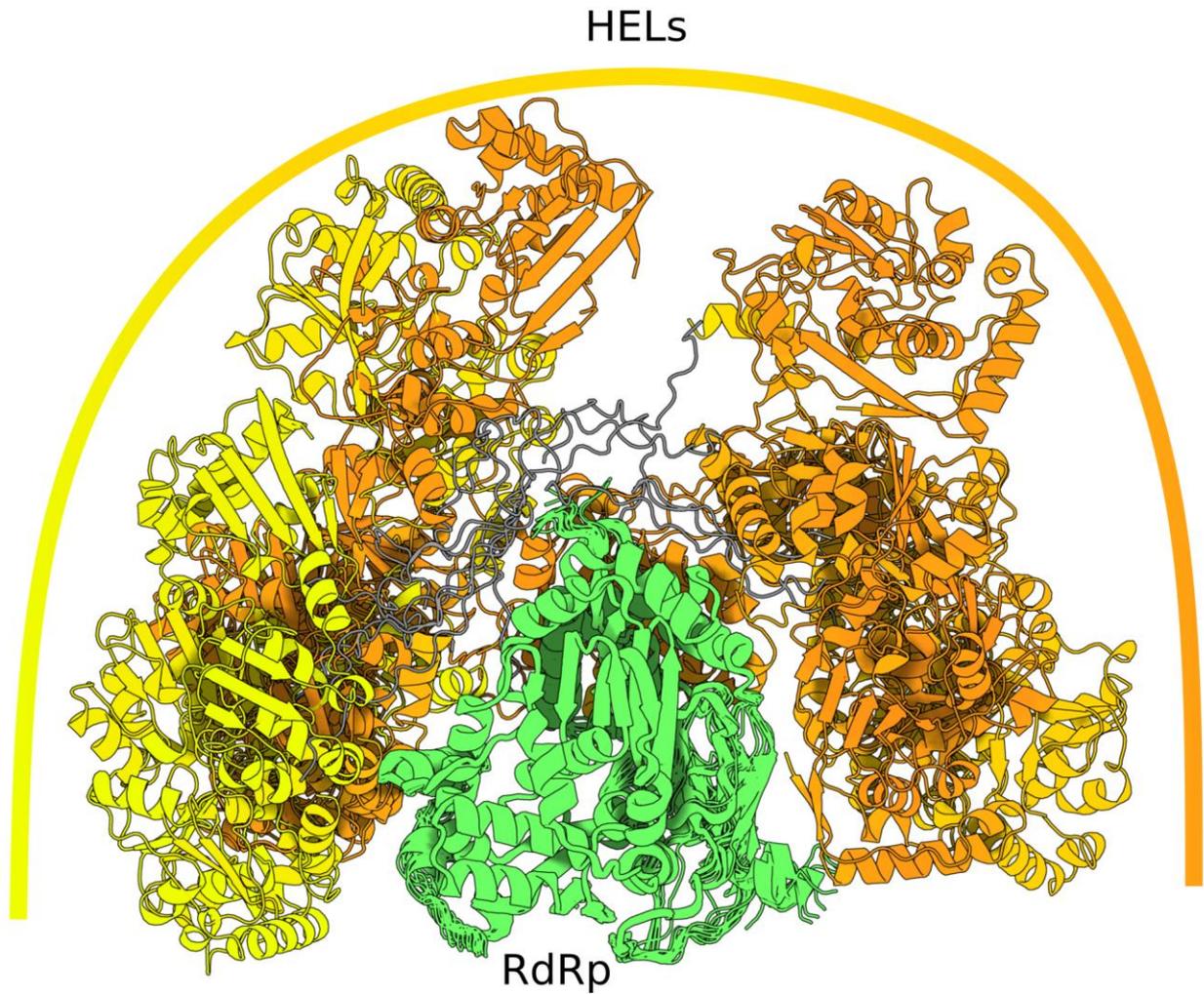


1033

1034 **Fig. S5. RdRp model 1 and 4 (from the HHBlits max alignment) in different orientations.**

1035 Cartoon colored from blue (N-ter) to red (C-ter). Rendered with ChimeraX <sup>37,38</sup>.

1036



1037

1038 **Fig. S6. Relative orientation of the HEL domains to the RdRp domains.**

1039 HEL domains are represented from a yellow to orange gradient and the RdRp domain in green. All  
1040 structures were aligned on the RdRp domain using the “super” method in PyMol <sup>103</sup>. AlphaFold2  
1041 models selected are HHblits-default and HHblits-MAX for ORF1-gt1 and HHblits-MAX for gt3. Rendered  
1042 with ChimeraX <sup>37,38</sup>.

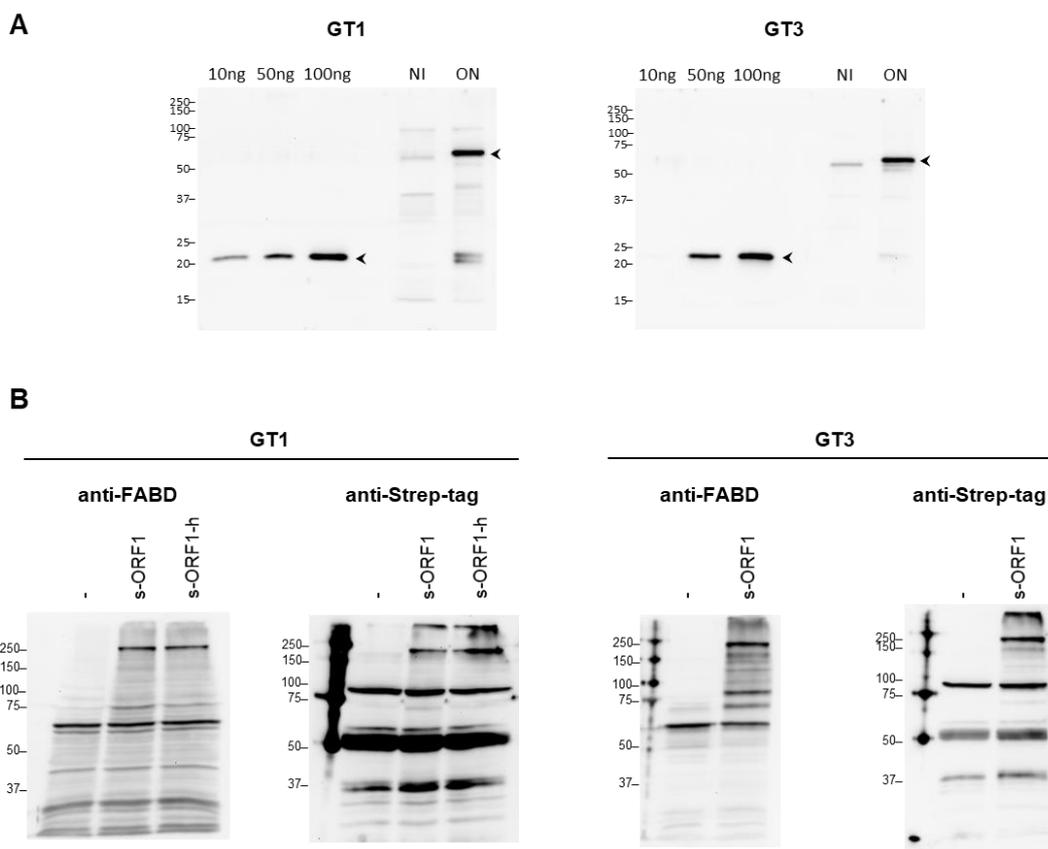
1043

1044

1045

1046

1047



1049

1050

1051

1052

1053 **Fig. S7. Production and characterization of polyclonal antibodies against the HEV FABD-like domain.**

1054 (A) The anti-FABD-like antibodies specifically recognize the pure recombinant protein expressed in *E.*

1055 *coli* (left panel) and the protein present in the bacterial extract after expression (NI, non-induced; ON,

1056 overnight MBP-FABD-like fusion protein expression induced by IPTG) (right). (B) The anti-FABD-like

1057 antibodies specifically recognize the HEV ORF1 expressed by *in vitro* translation, with a single (s-ORF1)

1058 or double tag (s-ORF1-h). Western-blot with anti-Strep-tag are also presented for comparison.

1059 Molecular weights are indicated.

1060