The fate of Hepatitis E virus capsid protein is regulated by an Arginine-Rich Motif

Kévin Hervouet[a,$], Martin Ferrié[a,$], Maliki Ankavay[a,$,#], Claire Montpellier[a], Charline Camuzet[a], Virginie Alexandre[a], Aïcha Dembélé[a], Cécile Lecoeur[a], Arnold Thomas Foe[a], Peggy Bouquet[b], David Hot[b], Thibaut Vausselin[a], Jean-Michel Saliou[a], Sophie Salomé-Desnoulez[b], Alexandre Vandeputte[b], Laurent Marsollier[c], Priscille Brodin[a,b], Marlène Dreux[d], Yves Rouillé[a], Jean Dubuisson[a], Cécile-Marie Aliouat-Denis[a], Laurence Cocquerel[a,*]

[a] University of Lille, CNRS, INSERM, CHU Lille, Pasteur Institute of Lille, U1019-UMR 9017-CIIL- Center for Infection and Immunity of Lille, F-59000 Lille, France

[b] Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, UMR2014 - US41 - PLBS-Plateformes Lilloises de Biologie & Santé, F-59000 Lille, France.

[c] Equipe ATIP AVENIR, CRCINA, INSERM, University of Angers, Angers, France.

[d] CIRI - Centre International de Recherche en Infectiologie, Univ Lyon, Université Claude Bernard Lyon 1, Inserm, U1111, CNRS, UMR5308, ENS Lyon, 69007, Lyon, France.

[$] These authors contributed equally to this work

[#]Present address: Division of Gastroenterology and Hepatology, Institute of Microbiology, Lausanne, Switzerland

[*] Address requests to Dr Laurence Cocquerel, Molecular & Cellular Virology, CIIL, CNRS-UMR9017 & Inserm-U1019, Institut Pasteur de Lille, Bâtiment IBL, 1 rue du Pr. Calmette, CS50447, 59019 Lille cedex, France, laurence.cocquerel@ibl.cnrs.fr

Abstract

Producing multifunctional proteins is one of the major strategies developed by viruses to condense their genetic information. Here, we investigated the molecular determinants of the multifunctionality of hepatitis E virus (HEV) ORF2 capsid protein. We previously identified 3 isoforms of ORF2 which are partitioned in different subcellular compartments to perform distinct functions. Notably, the infectious ORF2 (ORF2i) protein is the structural component of the virion, whereas the genome-free secreted and glycosylated ORF2 proteins likely act as a humoral immune decoy. We identified a 5 amino acid Arginine-Rich Motif (ARM) located in the ORF2 N-terminal region as a central regulator of the subcellular localizations and functions of ORF2 isoforms. We showed that the ARM controls ORF2 nuclear translocation, promoting regulation of host antiviral responses. This motif also regulates the dual topology and functionality of ORF2 signal peptide, leading to the production of either cytosolic infectious ORF2i or reticular non-infectious glycosylated ORF2 forms. Furthermore, the ARM likely serves as a cleavage site of the glycosylated ORF2 protein. Finally, it promotes ORF2 membrane association that is likely essential for particle assembly. In conclusion, our observations highlight ORF2 ARM as a unique central regulator of ORF2 addressing that finely controls the HEV lifecycle.

Introduction

Hepatitis E virus (HEV) is the most common cause of acute viral hepatitis worldwide and is an emerging problem in industrialized countries. This virus causes about 20 million infections annually [1]. While HEV infection is asymptomatic for most patients, some human populations including pregnant women and immunocompromised patients have higher risk to develop severe forms and chronic infections, respectively. HEV strains infecting humans have been classified into 4 main distinct genotypes (gt) belonging to a single serotype [2]. Gt1 and gt2 that infect humans only, are primarily transmitted through contaminated drinking water and are responsible for waterborne hepatitis outbreaks in developing countries. In contrast, gt3 and gt4 are zoonotic and are largely circulating in industrialized countries. They are mainly transmitted by contact with swine and consumption of inadequately heated pork products [3]. There is no specific treatment nor universal vaccine to fight against HEV.

HEV is a quasi-enveloped [4,5], positive-sense RNA virus expressing three open reading frames (ORFs): ORF1, ORF2 and ORF3 [6]. ORF1 encodes the ORF1 non-structural polyprotein that is the viral replicase [7]. ORF2 encodes the ORF2 viral capsid protein and ORF3 encodes a small protein that is involved in virion morphogenesis and egress [8]. Since studying the HEV lifecycle has long been hampered by the absence of efficient systems to amplify HEV, many steps of the HEV lifecycle remain poorly understood [9]. By combining the gt3 p6 strain [10] and a highly transfectable subclone of PLC/PRF/5 cells (PLC3 cells), we previously described an efficient HEV cell culture system [11]. This model notably enabled the pioneering demonstration that, during its lifecycle, HEV produces at least 3 forms of the ORF2 capsid protein: infectious ORF2 (ORF2i), glycosylated ORF2 (ORF2g), and cleaved ORF2 (ORF2c). The ORF2i protein is the structural component of infectious particles. It is not glycosylated and is likely derived from the assembly of the intracellular ORF2 (ORF2intra) form present in the cytosolic compartment. Importantly, we showed that a fraction of the ORF2intra form is translocated into the nucleus of infected cells [12]. In contrast, ORF2g and ORF2c proteins (herein referred to as ORF2g/c) are highly glycosylated and secreted in large amounts in culture supernatant (*i.e.,* about 1000x more than ORF2i [13]) and are the most abundant antigens detected in patient sera [11]. In addition, these proteins likely act as a humoral immune decoy that inhibits antibody-mediated neutralization [13]. How these different forms of ORF2 are generated during the HEV lifecycle has not yet been fully investigated. However, their sequence and post-translational modifications suggest that they might be produced

either by a distinct addressing into the secretory pathway and the nucleus [11,12], and/or by a differential translation process[13].

Here we investigated the mechanisms by which the ORF2 forms are produced and differentially addressed to cell compartments. We demonstrated that HEV has set up a nucleo-cytoplasmic transport mechanism of its capsid protein to modulate cell host immune responses. In addition, we found that during the HEV lifecycle, a fine-tuning of ORF2 partitioning occurs between cytosolic, reticular and nuclear compartments. Importantly, we identified a stretch of 5 amino acids (herein referred to as ARM, Arginine-Rich Motif) in the N-terminal region of the ORF2 protein that drives nuclear translocation and tightly modulates the stoichiometry between the different ORF2 forms, especially by regulating the functionality of the ORF2 signal peptide.

Results

The ORF2 protein transits through the nucleus in the early phase of infection.

We and others previously showed that the ORF2 protein is translocated into the nucleus of infected cells of patient liver biopsies [14] and in cell culture system [12]. Here, we first analyzed the ORF2 expression in PLC3 cells electroporated with the infectious p6 strain (PLC3/HEV-p6), at different time post-electroporation (p.e.) by immunofluorescence (**Supplementary Fig.1**). ORF2 staining and quantification of nuclear fluorescence showed that nuclear translocation of ORF2 takes place at early time points p.e. (*i.e.,* 18h) and is then followed by a nuclear export process, indicating that HEV has developed mechanisms for ORF2 nuclear import and export.

The ORF2 protein displays an Arginine-Rich Motif (ARM) that functions as a Nuclear Localization Signal (NLS).

To decipher the molecular mechanisms of ORF2 nuclear import, we first analyzed its amino acid (aa) sequence with the NLSTradamus prediction program [15]. We identified a potential Nuclear Localization Signal (NLS) corresponding to a conserved Arginine-Rich Motif (ARM, 5 Arginine residues: RRRGRR) in the N-terminal region of ORF2, downstream of its signal peptide (SP) (**Fig.1a**). We next generated a series of ORF2 mutants in the p6 strain that are depicted in **Fig.1a**. We characterized their expression and subcellular localization (**Fig.1b, Fig.1c** and **Supplementary Fig.2**), and their impact on the HEV lifecycle (**Fig.1d**).

The replacement of arginine residues by alanine (3R/3A, 2R/2A and 5R/5A mutants) led to a drastic reduction of ORF2 nuclear localization compared to the wt protein (**Fig.1b** and **Fig.1c**, Nuclear extract), indicating that the ARM is likely a functional NLS. Interestingly, the reduced nuclear localization of these mutants was associated with an accumulation of ORF2 in the Golgi apparatus (**Supplementary Fig.2**) and a reduced association with cellular membranes (**Fig.1c**, Membrane extract), indicating that these mutated proteins are likely soluble in the Golgi lumen. In addition, high molecular weight forms of ORF2intra in the soluble fraction as well as an increase of ORF2g/c secretion were observed (**Fig.1c**), suggesting a higher translocation into the secretory pathway associated to an improved functionality of ORF2 SP. Quantification of intracellular RNAs showed that replication was not altered in these mutants (**Fig.1d**). In addition, ARM mutations did not affect ORF3 expression (**Fig.1c** and **Supplementary**

Fig.3). However, these mutants no longer produced infectious viral progeny (**Fig.1d**). Thus, our results suggest that the ARM drives the ORF2 nuclear translocation and plays important functions in the HEV lifecycle, notably in the assembly of infectious particles.

We also generated mutants for which the PSG residues were replaced by 3 arginine residues (PSG/3R mutant), and alternatively SP was fully or partially deleted (ΔSP and ΔSP1 mutants, respectively) (**Fig.1a**). The PSG/3R mutant showed a marked nuclear localization (**Fig.1b** and **Fig.1c**, Nuclear extract) but was impaired in ORF2g/c secretion (**Fig.1c,** Supernatant), as observed for the SP deletion mutants, indicating that the addition of arginine residues strengthens the NLS function of ARM but inhibits the functionality of ORF2 SP. The PSG/3R mutant expressed the ORF3 protein but displayed lower intracellular replication levels and was no longer infectious (**Fig.1d**). The increased nuclear localization of this mutant is therefore likely responsible for the reduction of HEV RNA replication and assembly of infectious particles.

The full (ΔSP) or partial (ΔSP1) deletion of the ORF2 SP led to a total inhibition of ORF2 secretion (**Fig.1c**, Supernatant), as expected due to the absence of reticular translocation. Interestingly, the ORF2 protein still exhibited a nuclear localization (**Fig.1b**), indicating that the nuclear translocation process is independent of the reticular translocation. Because ORF2 and ORF3 are overlapping, and ORF3 is essential to particle secretion, the SP deletion mutants did not express the ORF3 protein (**Fig.1c**) and displayed reduced extracellular titers (**Fig.1d**). Intracellular titers were also lowered in SP mutants (**Fig.1d**), indicating that the ORF2 SP likely plays an important role in the assembly of infectious particles.

Lastly, the highly conserved Gly31 residue was also mutated (**Fig.1a**, G/A mutant). This mutant displayed a subcellular distribution similar to that of wt, and expressed the ORF3 protein. Although to a lesser extent than wt, G/A mutant produced intracellular particles, but showed reduced extracellular RNA and infectious levels. This indicates that the G/A mutation affects particle secretion.

We next carried out a comparative study of NLS sequences in viral proteins and their importin. We found that the ORF2 ARM is similar to the Epstein-Barr nuclear antigen leader protein (EBNA-LP) arginine-rich NLS (RRVRRR) that interacts with Importin-α1 [16]. Interestingly, ORF2 and Importin-α1 co-localized in the nucleus of infected cells with a Pearson correlation coefficient (PCC) of 0.670 (**Supplementary Fig.4**), and the mutation of arginine residues drastically reduced this colocalization.

Taken together, our results indicate that ORF2 colocalizes with Importin-α1 thanks to its ARM that serves as a functional NLS. In addition, these results suggest that the ARM is involved in the fine-tuning of the addressing and stoichiometry of the ORF2 protein between the nuclear, cytosolic and reticular pathways. This stoichiometry is likely essential to the HEV lifecycle.

The nuclear translocation of ORF2 down-regulates the NF-κB-related signaling.

Our results demonstrated that ORF2 localizes in the nucleus (ORF2ni) and the ARM is pivotal for nuclear translocation. Our results also suggested that ORF2ni is readily detected as early as 18h p.e. in PLC3/HEV-p6 cells while its nuclear targeting is transient and starts to decrease after 48h (**Supplementary Fig.1**). This observation prompted us to address the impact of early nuclear translocation of ORF2 on the regulation of host genes. We performed a transcriptomic analysis by microarrays (Agilent SurePrint Technology) in PLC3/HEV-p6-wt, PLC3/HEV-p6-5R/5A, PLC3/HEV-p6-ΔORF3 and PLC3 mock cells at 18h p.e.. Interestingly, in PLC3/HEV-p6-wt and PLC3/HEV-p6-ΔORF3 cells, we observed a significant inhibition of expression of 7 genes related to the TNFα, IL-17 and NF-κB-mediated signaling as well as inflammatory responses (*i.e.,* NOD-like receptor-induced response) (**Fig.1e** and **Supplementary Fig.5**). In contrast, no gene expression inhibition was observed in PLC3/HEV-p6-5R/5A cells, reflecting the significance of ORF2 nuclear translocation in the observed inhibition. Of note, while some reports suggested that ORF3 expression modulates the host responses [17–22], no marked difference was observed when comparing ΔORF3 mutant to wt.

Altogether our results suggest that the ORF2 ARM, which notably regulates ORF2 nuclear translocation, is a pivotal viral determinant for the modulation of host pathways and, especially, genes of the NF-κB-induced signaling upon infection. Further studies will be required to define precisely the impact of this HEV-driven host regulation on immune cell responses.

Nuclear export of the ORF2 protein

The observation that ORF2 nuclear targeting is transient and decreases after 48h (**Supplementary Fig.1**) then prompted us to investigate the mechanisms of ORF2 nuclear export. We treated PLC3/HEV-p6 cells with nuclear export inhibitors, Leptomycin B (LepB) and Verdinexor (Verd). These compounds are irreversible (LepB) and reversible (Verd) inhibitors of the ubiquitous transport receptor chromosome maintenance protein 1 (CRM1/Exportin 1), which recognizes hydrophobic leucine-rich export signals [23].

Treated cells displayed a highly significant nuclear accumulation of ORF2, as compared to control cells (**Fig.2a**). In addition, co-localization studies revealed that ORF2 partially or transiently colocalizes with CRM1 in untreated cells whereas they significantly colocalize upon treatment with nuclear export inhibitors (**Supplementary Fig.6**). These results indicate that ORF2 undergoes a nuclear export to the cytoplasm by a CRM1-dependent mechanism.

CRM1 recognizes hydrophobic leucine-rich export signals. By analyzing the ORF2 sequence, we identified at least 12 potential nuclear export signals (NES) into the ORF2 sequence. We replaced the hydrophobic residues in these conserved motifs by alanine residues and characterized the generated mutants as described above. Three of them, named NES9, NES10 and NES12 (**Fig.2b**) led to a highly significant accumulation of ORF2 inside the nucleus (**Fig.2c** and **Fig.2d**), as observed for cells treated with nuclear export inhibitors (**Fig. 2a**). NES9 and NES10 mutants were no longer infectious, whereas the NES12 mutant with the lowest nuclear accumulation (**Fig.2c**) still exhibited some intracellular and extracellular infectivity (**Fig.2e**). Of note, intracellular replication was not altered by NES mutations (**Fig.2f**), indicating that the loss of infectious particle assembly is due to the differential subcellular localization of the mutants and not to a replication defect.

These results indicate that HEV has set up a nuclear export system for its ORF2 capsid protein. This mechanism involves CRM1 which recognizes three NES on the ORF2 sequence. Moreover, these results suggest again that a fine balance between the nuclear, cytosolic and reticular pathways is likely essential to the HEV lifecycle.

Translocation and maturation of the glycosylated ORF2 forms

Next, we investigated the mechanisms of translocation and maturation of the highly secreted and glycosylated ORF2g/c isoforms. First, we treated PLC3/HEV-p6 and Mock cells with Mycolactone, an inhibitor of Sec61 translocon, the membrane embedded protein complex responsible for the translocation of newly synthetized polypeptides into the ER lumen (reviewed in [24]). Interestingly, we observed a dose-dependent reduction of ORF2g/c secretion in Mycolactone-treated PLC3/HEV-p6 cell supernatants (**Fig.3a**), indicating that reticular translocation of the ORF2g/c forms is Sec61-dependent. Previously, we demonstrated that the first residues of ORF2i, ORF2g and ORF2c proteins are Leu14, Ser34 and Ser102, respectively [11,12] (**Fig.2b**). Therefore, the first 20 aa of the ORF2i protein are not present in the ORF2g/c isoforms. Furthermore, the ORF2i protein is not glycosylated whereas ORF2g/c

proteins are highly glycosylated [12] (**Fig.3b**). Thanks to these features, we generated a murine monoclonal antibody (P1H1) that recognizes the N-terminus of ORF2i (**Fig.3b**). P1H1 specifically immunoprecipitates the ORF2i protein without cross-reacting with the highly secreted and glycosylated ORF2g/c proteins (**Fig.3c**, SN). We also generated the P3H2 antibody that recognizes the different isoforms of ORF2. Both antibodies recognize the intracellular ORF2 form (**Fig.3c,** Cells).

These antibodies were used to evaluate the effects of three furin inhibitors and related proprotein convertases (PC) on ORF2g/c maturation. PC cleave the multibasic motifs R-X-R/K/X-R in the precursor proteins [25]. The presence of ARM and RRR motif upstream of the ORF2g/c N-termini (**Fig.2b**), respectively, suggests that a PC might be involved in the maturation of these ORF2 forms. Therefore, we treated PLC3/HEV-p6 cells with three potent furin/PC inhibitors (decanoyl-RVKR-chloromethylketone [CMK], hexa-D-arginine amide [D6R], and SSM3 trifluoroacetate [SSM3]) [26] and immunoprecipitated ORF2 proteins in cell supernatants with P1H1 and P3H2 antibodies. Intracellular contents were probed by WB for ORF2intra, cleavage of cellular αV-pro-integrin (a substrate of intracellular furin) and tubulin (**Fig.3d-f**). In these experiments, immunoprecipitation of ORF2g by P1H1 antibody was used as a read-out of the inhibition of ORF2g maturation (**Fig.3b,** ORF2g*). In treated cells, we observed a dose-dependent immunoprecipitation of ORF2g* by P1H1 (**Fig.3d-f**), indicating that furin/PC inhibitors abrogated ORF2g maturation. Of note, the cell-permeable CMK and SSM3 inhibitors showed a strong inhibition of ORF2g and αV-pro-integrin maturation, whereas the cell membrane impermeable D6R inhibitor showed a moderate effect on ORF2g maturation. Together, these results indicate that a furin/PC present in the secretory pathway is likely involved in the ORF2g/c maturation process.

The ORF2 ARM is the regulator of ORF2 addressing

To further analyze the molecular mechanisms by which ORF2 is differentially addressed to the cytosolic, nuclear or reticular pathways, we next generated chimeric and mutant constructs between ORF2 and the CD4 glycoprotein, as a reporter protein. These constructs were expressed in Huh-7 cells stably expressing the T7 RNA-polymerase [27]. We selected the 5R/5A and PSG/3R mutations for their marked phenotype (**Fig.1**) and generated an ARM-deleted mutant (ΔARM, deletion of Gln24 to Arg33, **Fig.4**). The full-length ORF2wt, $ORF2^{5R/5A}$ and $ORF2^{PSG/3R}$ proteins displayed a similar pattern and phenotype as observed in the infectious system (**Fig.1**), by immunofluorescence and WB (**Fig.4**). The effect of the

5R/5A mutation or the ARM deletion on the ORF2 secretion, membrane association and nuclear localization confirmed that the ARM located downstream of the SP negatively regulates ORF2 reticular translocation but is important for nuclear translocation and membrane association. Conversely, the PSG/3R mutation showed an increased nuclear localization and membrane association, whereas ORF2 secretion was fully blocked, confirming that positively charged residues negatively regulate the functionality of ORF2 SP but mediate nuclear translocation and membrane association.

Next, to define the impact of ARM on another SP, we exchanged the ORF2 SP by the one of CD4 that is a functional model SP [28] (**Fig.4,** Chimeras C1). Interestingly, the chimera C1 displayed a subcellular distribution similar to that of ORF2wt but was no longer secreted (**Fig.4c**, SN), indicating that the ARM inhibits the functionality of CD4 SP. Indeed, the mutation ($C1^{5R/5A}$) or deletion of ARM ($C1^{\Delta ARM}$) restored secretion of the chimera C1. The observation that ORF2wt is secreted in the presence of the ARM whereas it is not with the CD4 SP, suggests the existence of an interplay between ORF2 SP and ARM. The chimera $C1^{PSG/3R}$ showed an increased nuclear localization and membrane association, whereas its secretion was abolished (**Fig.4**), confirming that positively charged residues downregulate the functionality of CD4 SP and mediate nuclear translocation and membrane association. It should be noted that the chimera $C1^{PSG/3R}$ showed a marked reticular staining (**Fig.4a**). As described below, we hypothesized that the CD4 SP might disturb the maturation of ORF2i and anchors the protein into the membrane on the cytosolic side.

We also generated an additional group of ORF2 constructs in which the SP of ORF2 was partially deleted (C2) (**Supplementary Fig.7**). The characterization of these constructs confirmed that ARM mediates ORF2 nuclear translocation and membrane association independently of the reticular translocation.

In order to specifically study the impact of ARM on the functionality of the ORF2 SP independently of the ORF2 ectodomain, CD4 chimeras containing the SP ($CD4^{SPORF2}$), the N-terminus (Chimeras C4) or the ARM of ORF2 (Chimeras C5) were also generated and characterized as previously (**Supplementary Fig.8**). Thanks to the $CD4^{SPORF2}$ construct, we confirmed that the SP of ORF2 is a functional SP, as illustrated by its subcellular pattern (**Supplementary Fig.8a**) and its efficient secretion (**Supplementary Fig.8d**). Interestingly, the chimera C4 showed an intracellular distribution different from that of CD4wt, with a significant nuclear localization (**Supplementary Fig.8a** and **b**). In addition, WB analysis revealed a major decrease in C4 secretion as well as the appearance of a 40kDa band in

the soluble fraction (**Supplementary Fig.8d**). This band corresponds to the non-N-glycosylated CD4 ectodomain, which contains 2 N-glycosylation sites, indicating that the CD4 ectodomain is poorly translocated into the ER lumen when fused to ORF2 N-terminus. The same observations were made for the chimera $C4^{PSG/3R}$. In contrast, the 5R/5A mutations restored the secretion but abolished the nuclear translocation of C4 ($C4^{5R/5A}$, **Supplementary Fig.8a**, **b** and **d**). Characterization of the chimeras C5, which contain only the ORF2 ARM, showed results similar to the chimeras C4 (**Supplementary Fig.8b**, **c** and **d)**. However, unlike C4, the chimera C5 was no longer secreted (**Supplementary Fig.8d)**, supporting the hypothesis of an interplay between ORF2 SP and ARM. Moreover, the chimera C5 displayed a reticular staining in addition to nuclear staining (**Supplementary Fig.8c)**. This observation is in line with our hypothesis that the CD4 SP does not undergo the same maturation as ORF2 SP, and anchors the protein into the membrane with a cytosolic orientation.

Taken together, these results demonstrate that the ORF2 ARM on its own is capable of regulating the functionality of ORF2 or CD4 SP, as well as the nuclear translocation of the protein that carries it.

The ORF2 ARM regulates the topology of ORF2 SP

Positively charged residues, such as arginine residues, are known to function as determinants of membrane protein topology, which is reflected in a statistical rule of membrane topology, *i.e.* the positive-inside rule of membrane proteins [29]. Notably, positive charges determine the orientation of the signal sequences and contribute to membrane spanning of the SP H-segment translocating through the translocon [29]. Since our results suggest that the ARM regulates the SP functionality and membrane association of ORF2, we next analyzed the topology of our different ORF2 and CD4 constructs by immunofluorescence (**Fig.5** and **Supplementary Fig.9**). We used low concentrations of digitonin that selectively permeabilize the plasma membrane. Triton X-100-permeabilized cells were analyzed in parallel as a control. The differential detection of two epitopes on the ER-membrane associated Calnexin (CNX) was used as a control of permeabilization. We observed that the ORF2wt, the chimera C1 and the PSG/3R mutants displayed a staining in both Triton X-100 and Digitonin-permeabilized cells whereas the ΔARM and 5R/5A constructs showed a labelling only in Triton X-100 permeabilized cells (**Fig.5**). These accessibility differences in association with the secretion efficiencies (**Fig.4c** and **Supplementary Fig.8d**), which reflect the reticular translocation, allowed us to infer the membrane orientation of each construct as well as the SP topology (**Fig.5** and **Supplementary Fig.9**). Thus, in the

presence of the ARM, the ORF2 SP likely adopts a double topology (ORF2wt) whereas the CD4 SP is not functional (C1) and sticks the protein to the cytosolic side of membranes. The ORF2 SP and CD4 SP are fully functional when the ARM is deleted (ΔARM) or mutated (5R/5A) whereas they are nonfunctional when the ARM is coupled to the PSG/3R mutations. The same observations were done for the C4 and C5 chimeras (**Supplementary Fig.9**).

Thus, our findings demonstrate that the ORF2 SP and the ARM act together to direct the fate of ORF2 capsid protein. Thanks to the ARM, the ORF2 SP is likely able to adopt a dual topology leading to either reticular translocation or membrane integration to the cytosolic side.

Discussion

In the present study, we analyzed a series of mutants of the HEV ORF2 capsid protein to gain insight into how a same primary sequence can generate several ORF2 isoforms with distinctive sequences, post-translational modifications, subcellular localizations and functions in the HEV lifecycle. Several important conclusions can be drawn from our analyses. The first is that the ORF2 protein early transits through the nucleus during infection to control specific cellular functions *i.e.* antiviral responses of the infected cell. We identified the determinants of the ORF2 nuclear import and export. Notably, an ARM in the N-terminal region of ORF2 mediates nuclear import. Importantly, we showed that the mutation of this motif abolishes ORF2 nuclear translocation but also affects ORF2 addressing into membrane, cytosolic and reticular compartments, which was deleterious for the HEV lifecycle. This brings us to the second important finding, the ARM is pivotal in the fine-tuning of the partitioning and stoichiometry of the ORF2 protein between the nuclear, cytosolic and reticular pathways that are essential to the HEV lifecycle. The last significant finding in this study is the manner by which the SP and ARM cooperate to control the fate of ORF2 protein. Indeed, in addition to mediate the targeting of ORF2 to the ER membrane, the SP is likely able to adopt a reverse signal-anchor topology. This topology inversion would be driven by flanking charged residues of ARM according to the positive-inside rule [30,31] and leads to the anchoring of the ORF2i protein to the cytosolic side of membranes.

Previously, we [11] and others [13] demonstrated that HEV produces several isoforms of the ORF2 capsid protein. The ORF2i protein is the structural component of infectious particles. It is likely derived from the assembly of the ORF2intra form present in the cytosolic compartment. The ORF2i and ORF2intra proteins are not glycosylated and display the same sequence starting at Leu14 corresponding to the middle of the SP, indicating that an intra-membrane protease might be involved in their maturation. Further investigation is required to identity this intramembrane protease. In contrast, ORF2g/c proteins are highly glycosylated and secreted, but are not associated with infectious material. We identified the first residues of ORF2g/c as Ser34 and Ser102, respectively [11,12]. The nature of the sequences upstream of the ORF2g/c N-termini and our experiments using PC inhibitors, suggest that a PC, such as Furin, might be involved in their maturation[32]. Intriguingly, the ARM is right upstream of the N-terminus of ORF2g, indicating that this motif also serves as recognition site for ORF2g maturation. Further experiments are needed to thoroughly investigate the processing of ORF2 forms.

We demonstrated that early during infection, the ORF2 protein transits through the nucleus likely to control antiviral responses of the infected cell. Although we identified the determinants of nuclear import and export in the ORF2 sequence, further studies are now required to identify the cellular partners of nuclear ORF2 and define precisely the impact of ORF2 host regulation on immune cell responses. In line with our results, a previous study demonstrated the significance of ORF2 N-terminal Arginine residues in the HEV interference with the host innate immunity via an inhibition of TBK1-mediated IRF3 phosphorylation [33]

Based on our findings, we propose a model of ORF2 production. Firstly, when engaged with the translocon, the ORF2 SP initially inserts head-on in an N_{exo}/C_{cyt} orientation (**Fig.6a**). Two mechanisms can then take place. In one side, ORF2 SP inverts orientation to N_{cyt}/C_{exo} to integrate the ER membrane as cleavable signal. The C-terminal end of signal is exposed to the ER lumen and cleaved by signal peptidase, liberating the ORF2 ectodomain in the ER lumen where it undergoes glycosylation and protease maturation. This pathway leads to the production of ORF2g/c proteins that are abundantly secreted and likely serve as immunological baits (**Fig.6b**). On the other side, ORF2 SP does not invert, keeps a N_{exo}/C_{cyt} orientation and serves as a reverse signal anchor. Next, the ORF2 protein anchored to the cytosolic side of membranes is likely processed by an intramembrane protease (**Fig.6c**). This pathway leads to the production of the ORF2intra protein that is early translocated into the nucleus to play immunomodulatory functions and/or is then assembled into viral particles in the cytosolic compartment. Of note, the dual topology was exclusively observed for the ORF2 and CD4 constructs containing both ORF2 SP and ARM (ORF2wt and chimera C4), reflecting the specific interplay between ORF2 SP and ARM and no other sequence determinant in this process.

Due to the size constraint of their extracellular phase, viruses are under strong pressure to minimize the size of their genome. Overlapping genes represent an adaptive strategy developed by many viruses to condense a maximum amount of information into short nucleotide sequences. In addition to this gene overlap strategy exploited by HEV for the ORF2 and ORF3 expression[34] and ORF4 in gt1 [35], HEV developed a master strategy of information condensation into five amino acids that control the fate and function of its capsid protein. Hence, the ORF2 ARM controls (i) the ORF2 nuclear localization and hereby controls cellular functions promoting regulation of host antiviral responses, (ii) the functionality of ORF2 SP leading to the production of either cytosolic infectious ORF2i or reticular non-infectious ORF2g/c forms, (iii) maturation of the ORF2g protein, and (iv) membrane association that is likely

1 essential to particle assembly. Therefore, we conclude that the ORF2 ARM is a central regulator of the
2 HEV lifecycle.
3

Methods

Cell cultures. PLC3 [11]and Huh-7.5 [36] cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% inactivated fetal calf serum and 1% of Non-Essential amino acids (Life Technologies) at 37 °C. Transfected PLC3 cells were maintained at 32 °C in a medium containing DMEM/M199 (1v:1v), 1 mg/ml of lipid-rich albumin (Albumax I™), 1% of Non-Essential amino acids and 1% of pyruvate sodium (Life Technologies).

The Huh-7-derived H7-T7-IZ cells stably expressing the T7 RNA polymerase ([27]; kindly provided by Ralf Bartenschlager, University of Heidelberg, Germany) were maintained in a medium supplemented with 50 µg/ml of Zeocin. They were used for the transfection of the T7 promoter-driven pTM expression vectors.

Plasmids and transfection. The plasmid pBlueScript SK(+) carrying the DNA of the full length genome of adapted gt3 Kernow C-1 p6 strain, (GenBank accession number JQ679013, kindly provided by S.U Emerson) was used as a template[10]. Mutants of the ORF2 ARM or NES sites were generated by site directed mutagenesis. Individual mutations were introduced by sequential PCR steps, as described previously [12], using the Q5 High-Fidelity 2X Master Mix (New England Biolabs, NEB), then digestions with restriction enzymes and ligation were performed. All the mutations were verified by DNA sequencing. The primers used for the generation of ORF2 mutants are listed in **Supplementary Table 1.** The ORF3-null mutant of HEV-p6 (HEV-p6-ΔORF3) was generated as described in [37]

To prepare genomic HEV RNAs (capped RNA), pBlueScript SK(+) HEV plasmids were linearized by digestion with the MluI restriction enzyme (NEB) and transcribed with the mMESSAGE mMACHINE kit (Ambion). Capped RNAs were next delivered to PLC3 cells by electroporation using a Gene Pulser Xcell™ apparatus (Bio-Rad).

The plasmids pTM-ORF2 (kindly provided by J. Gouttenoire, University of Lausanne, Switzerland) [14] and pTM/CD4 have been previously described [14,38]. The pTM/CD4 contains the DNA sequence coding for the secreted ectodomain of CD4 (aa 1-371). The primers used for the generation of ORF2/CD4 chimeras/mutants are listed in **Supplementary Table 1.** For some constructs, PCR amplifications were performed by multiple heat pulses [39]. The pTM plasmids were transfected into H7-T7-IZ cells using ViaFect™ Transfection Reagent (Promega) following the manufacturer's recommendations.

Antibodies. Primary antibodies used in this study are listed in **Supplementary Table 2**. Secondary antibodies (Cyanine-3-Goat anti-Mouse; Alexa Fluor-488-Goat anti-Rabbit; Alexa Fluor-488-Donkey anti-Goat; Cyanine-3-Donkey Anti-mouse IgG2b; Alexa Fluor-488-Donkey Anti-mouse IgG1) were from Jackson ImmunoResearch.

Indirect immunofluorescence. Cells were fixed with 3% of Paraformaldehyde (PFA) for 20 minutes (min). Cells were next washed twice with phosphate-buffered saline (PBS) and permeabilized for 5 min with cold methanol and then with 0.5% Triton X-100 for 30 min. Cells were incubated in PBS containing 10% goat serum for 30 min at room temperature (RT) and stained with primary antibodies for 30 min at RT followed by secondary antibodies for 20 min at RT. The nuclei were stained with DAPI (4′,6-dia midino-2-phenylindole) and cell outlines with CellMask™ Green (Invitrogen). After 2 washes with PBS, coverslips were mounted with Mowiol 4–88 (Calbiochem) on glass slides and analyzed with a LSM 880 confocal laser-scanning microscope (Zeiss) using Plan Apochromat 63xOil/1.4N.A. and EC Plan Neofluar 40xOil/1.4N.A. objectives. The images were then processed using ImageJ and Fiji softwares. For selective permeabilization experiments, cells were fixed with 2% of PFA, washed twice with PBS and permeabilized for 30min at 4°C with either 0.01% of Digitonin (Sigma) in buffer containing 20 mM HEPES pH6.9, 0.3 M sucrose, 0.1 M KCl, 2.5 mM MgCl2 and 1 mM EDTA or 0.5% Triton X-100 in PBS. Cells were next stained as described above with buffers containing Digitonin and Triton, respectively.

Quantification of the ORF2 protein nuclear fluorescence. The method was adapted from McCloy *et al*. [40]. Briefly, the ORF2 protein nuclear fluorescence was determined using ImageJ software. The regions of interest (ROI) were drawn around the nuclei of cells using ImageJ ROI tools. Area, integrated density and mean gray values were measured. Then, corrected total cell fluorescence (CTCF) was calculated by the following formula: CTCF = integrated density – (area of selected electroporated cells x mean of background fluorescence around the cells). The exact nuclear fluorescence was = CTCF-the mean of the integrated density of non-infected cells.

For nuclear/cytosolic fluorescence intensity ratio, cells were co-stained with CellMask™ Green (Invitrogen) and analyzed using ImageJ software. ROI were drawn around the whole cells and the nuclei. Area, integrated density, mean values and the exact cell and nuclear fluorescence were measured and calculated as described above. For each cell, the nuclear/cytosolic fluorescence intensity ratio was

calculated by the following formula: exact fluorescence intensity of nucleus / (exact fluorescence intensity of whole cell - exact fluorescence intensity of nucleus).

Pearson's correlation coefficient (PCC) determination. Colocalization studies were performed by calculating the PCC using the JACoP plugin of ImageJ software. The PCC examines the relationship between the intensities of pixels from two channels in the same image. For each calculation, at least 30 cells were analyzed to obtain a PCC mean. A PCC of 1 indicates perfect correlation, 0 no correlation, and -1 a perfect anti-correlation.

Virus production and intracellular viral particles preparation. PLC3 cells were electroporated with HEV-p6 RNAs as previously described [11]. Supernatant of confluent T75 flasks of HEV producing cells were harvested, centrifuged for 10min at 800 rpm and stored at -80°C until experiment. For intracellular particles, the procedure was adapted from [41]. Briefly, cells were trypsinized and centrifuged for 10 min at 1500 rpm. Cells were washed thrice with PBS. Intracellular viral particles were extracted by resuspending cells in 1 ml of sterile MilliQ water at room temperature. Cells were vortexed vigorously for 20 min and then 110 µl of sterile 10X PBS were added. Samples were clarified by centrifugation 2 min at 14000 rpm. The supernatants containing intracellular particles were collected and stored at −80°C until analysis.

Infectious titer determination. Huh7.5 cells seeded in 96-well plates were infected with serial dilutions of supernatants or intracellular viral particles from HEV producing cells. The inoculum was removed after 8h and cells were overlaid with fresh medium. Three days post-infection, cells were fixed and processed for indirect immunofluorescence with 1E6 anti-ORF2 antibody. ORF2-positive cells were quantified using an InCell 6000 confocal analyzer (GE Healthcare) and the Columbus image analysis software (Perkin Elmer). The number of infected cells was defined for each dilution and use to define infectious titers in focus forming unit (FFU/mL).

Cell treatments with chemicals. Leptomycin B (Cell Signaling) was dissolved in absolute ethanol to generate a 200µM stock and diluted in culture medium to generate a 20nM solution. This solution was added to cells during 16h. Verdinexor (AdooQ® Biosciences) was dissolved in DMSO to generate a

20μM stock and diluted in culture medium to generate a 100nM solution. This solution was added to cells during 16h. Mycolactone A/B toxin was stored at -20°C in dark glass tubes at 6.7 mg/ml. Cells were treated for 24h with the following concentrations of mycolactone: 5nM, 10nM, 15nM, 30nM and 50 nM. Decanoyl-RVKR-chloromethylketone [CMK] (Sigma), hexa-D-arginine amide [D6R] (Sigma) and SSM3 trifluoroacetate (Tocris) were dissolved in DMSO and next diluted in culture medium to generate solutions at indicated concentrations (in μM). Dose-response curves of PLC3 cells treated with the different drugs are shown in **Supplementary Figure 10.**

RNA extraction and quantification. HEV RNA levels were quantified by RT-qPCR using primers (5'-GGTGGTTTCTGGGGTGAC-3' (F) and 5'-AGGGGTTGGTTGGATGAA-3' (R)) and a probe (5'-FAM-TGATTCTCAGCCCTTCGC-TAMRA-3') that target a conserved 70 bp region in the ORF2/3 overlap. In **Supplementary Figure 5,** HEV RNA levels were quantified by RT-qPCR using primers (5'-: AAGACATTCTGCGCTTTGTT-3' (F) and 5'- TGACTCCTCATAAGCATCGC-3' (R)) and a probe (5'-FAM- CCGTGGTTCCGTGCCATTGA-TAMRA-3') that target a conserved region of ORF1. HEV RNAs were extracted from culture surpernatants with the QIAmp viral RNA mini kit (Qiagen) and from cells with the Nucleospin RNA Plus kit (Macherey & Nagel). Retrotranscription was performed using the AffinityScript Multiple temperature cDNA synthesis Kit (Agilent Technologies) according manufacturer's instructions. Amplifications were done with a Quant Studio 3 apparatus (Applied Biosystems) and Taqman universal master mix no AmpErase UNA (Applied Biosystems).

Cellular gene RNA levels were quantified by RT-qPCR using in-home primers (see **Supplementary Table 3**) and standards. Total cellular RNAs were extracted using TRIzol (Invitrogen) according manufacturer's instructions and processed for retrotranscription using the High capacity reverse transcription kit (Applied Biosystems). Amplifications were done with a Quant Studio 3 apparatus (Applied Biosystems) and SYBRGreen PCR Master Mix (Applied Biosystems).

Western blotting analysis. Cells were lysed in buffer containing 10 mM TrisHCl (pH 7), 150 mM NaCl, 2 mM EDTA, 0.5% Triton X-100, 1 mM PMSF and protease inhibitor cocktail (Complete, Roche). Supernatants and cell lysates were stored at −80 °C until analysis. Protein concentrations were determined by Bicinchoninic Acid Protein Assay kit (Sigma) according to manufacturer's instructions. Western blotting analyses were performed as described previously [11]. Briefly, supernatants and lysates

were heated for 20 min at 80 °C in the presence of reducing Laemmli buffer. Samples were then separated by 10% SDS-PAGE and transferred onto nitrocellulose membranes (Hybond-ECL, Amersham). The targeted proteins were detected with specific antibodies (**Supplementary Table 2**) and corresponding peroxidase-conjugated secondary antibodies. The detection of proteins was done by chemiluminescence analysis (ECL, Amersham).

Immunoprecipitations. P1H1 and P3H2 antibodies were bound to magnetic Dynabeads® M-270 Epoxy beads (Thermofisher) overnight at 37°C following the manufacturer's recommendations. Beads were washed and then incubated for 1h at room temperature with heat-inactivated supernatants. Beads were washed and then heated at 80 °C for 20 min in Laemmli buffer. Proteins were separated by SDS-PAGE and ORF2 proteins were detected by WB using the 1E6 MAb.

Cell viability assay. PLC3 cells were seeded in 96-well plates and incubated at 37°C for 24 h. Subconfluent cultures were treated with different concentrations of inhibitors for 24h or 72h. A MTS [3-(4, 5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium, inner salt] based assay (CellTiter 96 aqueous nonradioactive cell proliferation assay, Promega) was used to evaluate cell viability. Cells treated with DMSO or ethanol served as controls. Dose-response curves of PLC3 cells treated with the different drugs are shown in **Supplementary Figure 10.**

Subcellular extraction. Confluent cells were harvested with trypsin-EDTA. Next, they were centrifuged at 4000 rpm for 5 min and washed thrice with PBS. Cytoplasmic, membrane and nuclear soluble proteins were extracted using the Subcellular protein fractionation kit for cultured cells (Thermo scientific) following the manufacturer's recommendations. The cytoplasmic extracts were ultra-centrifugated at 100 000 g during 1h at 4°C. Anti-β-tubulin (cytoplasmic), anti-Calnexin (membranous) and anti-SP1 or anti-Lamin B1 (nuclear soluble) antibodies were used to control the quality of extractions.

Transcriptomic analysis. PLC3 cells were electroporated with HEV-p6-wt, HEV-p6-5R/5A, HEV-p6-ΔORF3 RNAs or no RNA (mock). At 18h.p.e. total cellular RNAs were extracted using TRIzol (Invitrogen) according manufacturer's instructions. RNA integrity and purity were verified using the Agilent Bioanalyzer system (Agilent Technology). Two µg of total RNA were treated with 2 units of DNaseI (Sigma Aldrich) during 10 minutes before purification on Nucleomag NGS cleanup beads (Macherey

Nagel). Oligonucleotide microarrays for human whole genome (G4858A design 072363, 8x60k chips SurePrint G3 unrestricted GE, Agilent Technologies) were used for global gene expression analysis. Two hundred ng of total RNA was used in the Agilent Quick-Amp Labeling kit according to the manufacturer's instructions. After purification using an RNeasy Mini Kit (Qiagen), cRNA yield and incorporation efficiency (specific activity) into the cRNA were determined using a NanoDrop 2000 (Thermo Scientific) spectrophotometer. For each sample, a total of 600 ng of cRNA was fragmented and hybridized overnight at 65°C. After hybridization, slides were washed before being scanned on a SureScan Microarray Scanner (Agilent Technologies) and further processed using Feature Extraction v10.7.3.1 software. The resulting text files were uploaded into language R v4.0.3 and analyzed using the LIMMA package (Linear Model for Microarray Data) [42,43]. A 'within-array' normalization was performed using LOWESS (locally weighted linear regression) to correct for dye and spatial effects [44]. Moderate *t*-statistic with empirical Bayes shrinkage of the standard errors [45] was then used to determine significantly modulated genes. Statistics were corrected for multiple testing using a false-discovery rate approach. Protein-protein interactions network was generated using STRING database [46]. Gene ontology enrichment was performed using Metascape resource [47] (www.metascape.org) on the significantly modulated genes to identify pathways significantly modulated by either wild-type or mutants.

Statistical analyses. Statistical analyses were performed with the software RStudio version 1.2.5001 combined with R version 3.6.1. For all statistical tests, reported p values were two-sided. A test was declared statistically significant for any p value below 0.05. For comparing more than three groups of unpaired data, ANOVA or its non-parametric equivalent test, the Kruskal-Wallis test, was used. ANOVA was preferred when the distributions in each group followed a normal distribution, and the assumption of equality of the variances between each group was verified. When tests showed a significant difference between the groups, post hoc tests (available in the R package PMCMRplus) were performed. The Dunnett test (dunnettTest function with default option) followed an ANOVA, and the Conover's test (kwManyOneConoverTest function with pvalues adjusted by the Benjamini-Hochberg procedure) followed the Kruskal-Wallis test. Each group was compared to a reference. For the kinetic experiment, instead of using ANOVA for paired data, the Friedman test was preferred because of a lack of normality

and homogeneity of variance. The post hoc Nemenyi test (frdManyOneNemenyiTest function from the package PMCMRplus) was used and each time point was compared to data observed at 18 hours.

References

1. Horvatits, T., Wiesch, J. S. zur, Lütgehetmann, M., Lohse, A. W. & Pischke, S. The Clinical Perspective on Hepatitis E. *Viruses* **11**, 617–19 (2019).

2. Wang, S. *et al.* Rabbit and human hepatitis E virus strains belong to a single serotype. *Virus research* **176**, 101–106 (2013).

3. Doceul, V., Bagdassarian, E., Demange, A. & Pavio, N. Zoonotic Hepatitis E Virus: Classification, Animal Reservoirs and Transmission Routes. *Viruses* **8**, 270 (2016).

4. Feng, Z. *et al.* A pathogenic picornavirus acquires an envelope by hijacking cellular membranes. *Nature* **496**, 367–371 (2013).

5. Yin, X., Li, X. & Feng, Z. Role of Envelopment in the HEV Life Cycle. *Viruses* **8**, (2016).

6. Tam, A. W. *et al.* Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome. *Virology* **185**, 120–131 (1991).

7. LeDesma, R., Nimgaonkar, I. & Ploss, A. Hepatitis E Virus Replication. *Viruses* **11**, 719–17 (2019).

8. Nimgaonkar, I., Ding, Q., Schwartz, R. E. & Ploss, A. Hepatitis E virus: advances and challenges. *Nature reviews. Gastroenterology & hepatology* **15**, 96–110 (2018).

9. Wißing, M. H., Brüggemann, Y., Steinmann, E. & Todt, D. Virus–Host Cell Interplay during Hepatitis E Virus Infection. *Trends Microbiol* **29**, 309–319 (2020).

10. Shukla, P. *et al.* Adaptation of a genotype 3 hepatitis E virus to efficient growth in cell culture depends on an inserted human gene segment acquired by recombination. *J Virol* **86**, 5697–5707 (2012).

11. Montpellier, C. *et al.* Hepatitis E Virus Lifecycle and Identification of 3 Forms of the ORF2 Capsid Protein. *Gastroenterology* **154**, 211-223.e8 (2018).

12. Ankavay, M. *et al.* New insights into the ORF2 capsid protein, a key player of the hepatitis E virus lifecycle. *Sci Rep.* **9**, 6243 (2019).

13. Yin, X. *et al.* Origin, antigenicity, and function of a secreted form of ORF2 in hepatitis E virus infection. *Proc Natl Acad Sci USA* **3**, 201721345–6 (2018).

14. Lenggenhager, D. *et al.* Visualization of hepatitis E virus RNA and proteins in the human liver. *J Hepatol* **67**, 471–479 (2017).

15. Ba, A. N. N., Pogoutse, A., Provart, N. & Moses, A. M. NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics* **10**, 202 (2009).

16. Nakada, R. & Matsuura, Y. Crystal structure of importin-α bound to the nuclear localization signal of Epstein-Barr virus EBNA-LP protein. *Protein Science* **26**, 1231–1235 (2017).

17. Dong, C. *et al.* Suppression of interferon-α signaling by hepatitis E virus. *Hepatology* **55**, 1324–1332 (2012).

18. Lei, Q. *et al.* HEV ORF3 downregulates TLR7 to inhibit the generation of type I interferon via impairment of multiple signaling pathways. *Sci Rep.* **8**, 8585 (2018).

19. Wang, M., Huang, Y., He, M., Peng, W.-J. & Tian, D.-Y. Effects of hepatitis E virus infection on interferon production via ISG15. *World J Gastroentero* **24**, 2173–2180 (2018).

20. He, M. *et al.* The ORF3 Protein of Genotype 1 Hepatitis E Virus Suppresses TLR3-induced NF-κB Signaling via TRADD and RIP1. *Sci Rep.* **6**, 27597 (2016).

21. Nan, Y. *et al.* Enhancement of Interferon Induction by ORF3 Product of Hepatitis E Virus. *J Virol* **88**, 8696–8705 (2014).

22. Xu, J. *et al.* Open reading frame 3 of genotype 1 hepatitis E virus inhibits nuclear factor-κappa B signaling induced by tumor necrosis factor-α in human A549 lung epithelial cells. *PLoS ONE* **9**, e100787 (2014).

23. Mathew, C. & Ghildyal, R. CRM1 Inhibitors for Antiviral Therapy. *Frontiers in Microbiology* **8**, 10995–20 (2017).

24. Demangel, C. & High, S. Sec61 blockade by mycolactone: A central mechanism in Buruli ulcer disease. *Biology of the cell / under the auspices of the European Cell Biology Organization* **110**, 237–248 (2018).

25. Seidah, N. G. & Prat, A. The biology and therapeutic targeting of the proprotein convertases. *Nat Rev Drug Discov* **11**, 367–383 (2012).

26. Cheng, Y.-W. *et al.* Furin Inhibitors Block SARS-CoV-2 Spike Protein Cleavage to Suppress Virus Production and Cytopathic Effects. *Cell Reports* **33**, 108254 (2020).

27. Romero-Brey, I. *et al.* Three-dimensional architecture and biogenesis of membrane structures associated with hepatitis C virus replication. *PLoS Pathogens* **8**, e1003056 (2012).

28. Lumangtad, L. A. & Bell, T. W. The signal peptide as a new target for drug design. *Bioorganic & Medicinal Chemistry Letters* **30**, 127115–8 (2020).

29. Heijne, G. von. Membrane-protein topology. *Nature Reviews Molecular Cell Biology* **7**, 909–918 (2006).

30. Nilsson, J., Persson, B. & Heijne, G. von. Comparative analysis of amino acid distributions in integral membrane proteins from 107 genomes. *Proteins* **60**, 606–616 (2005).

31. Heijne, G. von. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature* **341**, (1989).

32. Izaguirre, G. The Proteolytic Regulation of Virus Cell Entry by Furin and Other Proprotein Convertases. *Viruses* **11**, 837–19 (2019).

33. Lin, S. *et al.* The Capsid Protein of Hepatitis E Virus Inhibits Interferon Induction via Its N-terminal Arginine-Rich Motif. *Viruses* **11**, 1050–17 (2019).

34. Graff, J., Torian, U., Nguyen, H. & Emerson, S. U. A Bicistronic Subgenomic mRNA Encodes both the ORF2 and ORF3 Proteins of Hepatitis E Virus. *J Virol* **80**, 5919–5926 (2006).

35. Nair, V. P. *et al.* Endoplasmic Reticulum Stress Induced Synthesis of a Novel Viral Factor Mediates Efficient Replication of Genotype-1 Hepatitis E Virus. *PLoS Pathogens* **12**, e1005521 (2016).

36. Blight, K. J., Mckeating, J. A. & Rice, C. M. Highly Permissive Cell Lines for Subgenomic and Genomic Hepatitis C Virus RNA Replication. *J Virol* **76**, 13001–13014 (2002).

37. Graff, J. *et al.* The Open Reading Frame 3 Gene of Hepatitis E Virus Contains a cis-Reactive Element and Encodes a Protein Required for Infection of Macaques. *J Virol* **79**, 6680–6689 (2005).

38. Cocquerel, L., Wychowski, C., Minner, F., Penin, F. & Dubuisson, J. Charged Residues in the Transmembrane Domains of Hepatitis C Virus Glycoproteins Play a Major Role in the Processing, Subcellular Localization, and Assembly of These Envelope Proteins. *J Virol* **74**, 3623–3633 (2000).

39. Orpana, A. K., Ho, T. H. & Stenman, J. Multiple Heat Pulses during PCR Extension Enabling Amplification of GC-Rich Sequences and Reducing Amplification Bias. *Anal Chem* **84**, 2081–2087 (2012).

40. McCloy, R. A. *et al.* Partial inhibition of Cdk1 in G 2phase overrides the SAC and decouples mitotic events. *Cell Cycle* **13**, 1400–1412 (2014).

41. Emerson, S. U., Nguyen, H., Torian, U. & Purcell, R. H. ORF3 Protein of Hepatitis E Virus Is Not Required for Replication, Virion Assembly, or Infection of Hepatoma Cells In Vitro. *J Virol* **80**, 10457–10464 (2006).

42. Brownstein, M. J., Khodursky, A., Smyth, G. K., Yang, Y. H. & Speed, T. Functional Genomics, Methods and Protocols. *Methods Mol Biology Clifton N J* **224**, 111–136 (2003).

43. Ihaka, R. & Gentleman, R. R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics* **5**, 299–314 (1996).

44. Yang, Y. H. *et al.* Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* **30**, e15–e15 (2002).

45. Lönnstedt, I. & Speed, T. Replicated Microarray Data. *Statistica Sinica* **12**, 31–46 (2002).

46. Snel, B., Lehmann, G., Bork, P. & Huynen, M. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res* **28**, 3442–3444 (2000).

47. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523 (2019).

Acknowledgements

This work was supported by a grant from the French agency ANRS-Maladies infectieuses émergentes. This work was also supported by the Pasteur Institute of Lille, Région Haut-de-France and Inserm-transfert. K.H., M.A. and T.V. were supported by a fellowship from the ANRS. M.F. was supported by a fellowship from the Pasteur Institute of Lille and Région Hauts-de-France. The PLBS platform used in this work was supported by the ANR (ANR-10-EQPX-04-01) and Feder (12001407 [D-AL] EquipEx ImagInEx BioMed). We thank Olivia Beseme for her technical contribution. We thank Suzanne U. Emerson (NIH, USA), Jérôme Gouttenoire (University of Lausanne) and Ralph Bartenschlager (University of Heidelberg) for providing us with reagents. We thank François-Loïc Cosset (University of Lyon) for critical reading of the manuscript.

Author Contributions

K.H., M.F., M.A., C.M., C.C., V.A., A.D., C.L., A.T.F., P.B., D.H., T.V., J-M.S., S.S-D., A.V., P.B., M.D., Y.R., J.D., C-M.A., and L.C. performed research and/or analyzed data.

L.M., M.D., Y.R. contributed to reagents or analytic tool.

L.C. wrote the paper.

Competing Interests statement

The authors declare no competing interests.

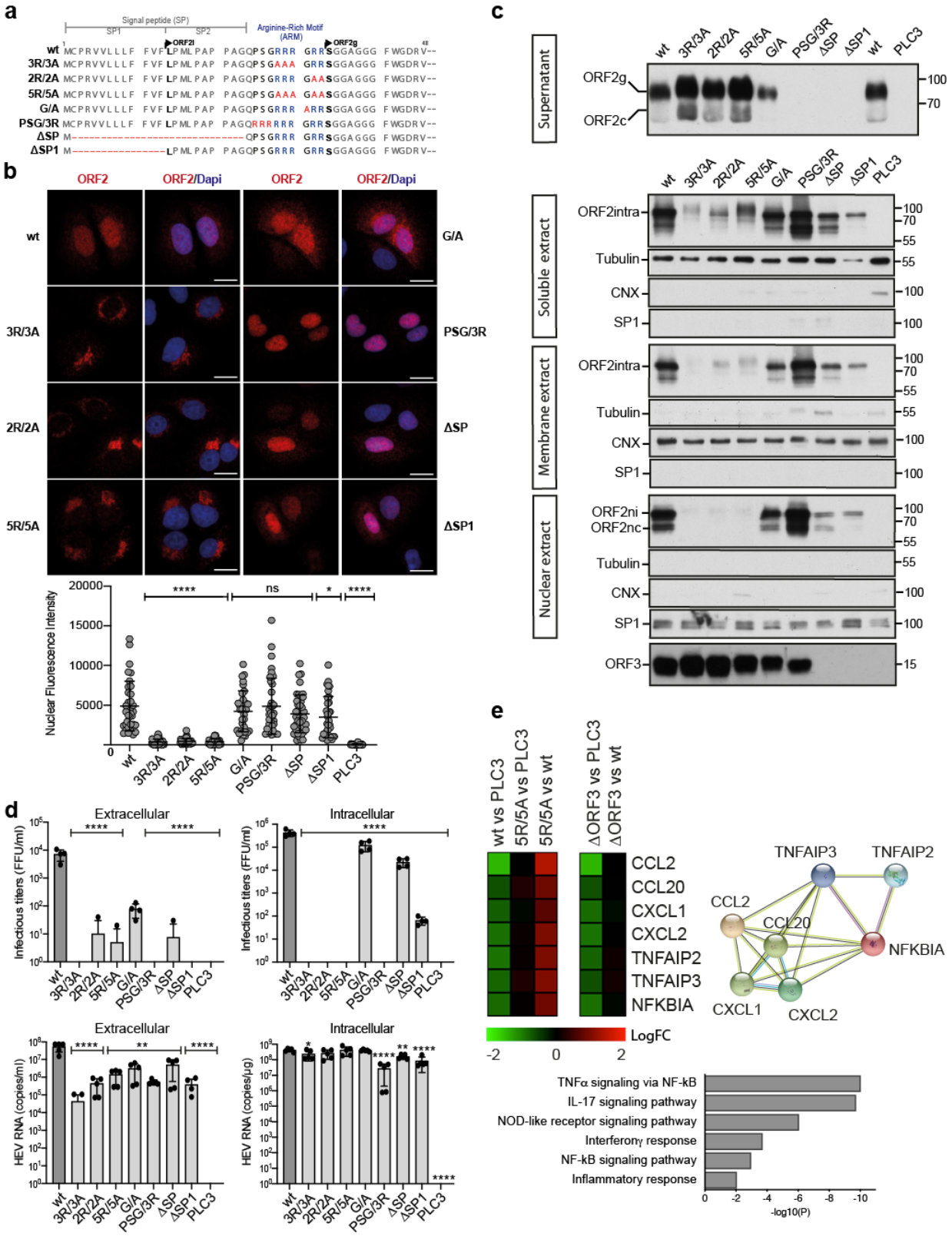

Figure 1: ORF2 contains an Arginine-Rich Motif (ARM) that is important for its nuclear localization and host gene expression regulation. a, Schematic sequence alignment of ORF2wt and ARM/SP mutants. **b,** Subcellular localization of ORF2wt and ARM/SP mutants. PLC3 cells were electroporated with wt and mutant HEV-p6 RNAs. At 18h p.e, cells were processed for indirect

immunofluorescence using the 1E6 anti-ORF2 antibody (Ab) and analyzed by confocal microscopy (magnification x63). Red = ORF2; Blue = DAPI. Scale bar, 20μm. Nuclear fluorescence intensities quantification was done using ImageJ software (mean ± S.D., $n \geq 30$ cells, Kruskal-Wallis with Conover's test). *$p < 0.05$, ****$p < 0.0001$ **c,** Subcellular fractionation of PLC3/HEV-p6 expressing ORF2wt and ARM/SP mutants at 10 d.p.e. Fractionation was done using the subcellular protein fractionation kit for cultured cells. ORF2 proteins were detected by WB with 1E6 Ab. Glycosylated ORF2 (ORF2g), cleaved ORF2 (ORF2c), intracellular ORF2 (ORF2intra), nuclear ORF2intra (ORF2ni), nuclear and cleaved ORF2intra (ORF2nc) are indicated. ORF3 protein in cell lysates was detected with a rabbit anti-ORF3 Ab. Tubulin, ER marker Calnexin (CNX) and the transcription factor SP1 used as a nuclear marker, were also detected to check the quality of fractionation. Molecular mass markers are indicated on the right (kDa). **d,** Infectious titer determination and HEV RNA quantification in PLC3/HEV-p6 expressing ORF2wt or mutant proteins. Extra- and intracellular viral particles were extracted at 10 d.p.e and used to infect naïve Huh7.5 cells for 3 days. Cells were next processed for indirect immunofluorescence. ORF2-positive cells were counted and each positive cell focus was considered as one FFU. Results were expressed in FFU/ml (n=4). Extra- and intracellular viral RNAs were quantified at 10 d.p.e by RT-qPCR ($n \geq 5$) (mean ± S.D., Kruskal-Wallis with Conover's test). *$p < 0.05$, **$p < 0.01$, ****$p < 0.0001$ **e,** Transcriptomic analysis of PLC3 cells expressing HEV-p6 wt, mutants or mock cells performed with microarrays (Agilent SurePrint). Left - Heatmap of gene expression in PLC3 cells expressing HEV-p6 wt, mutants or mock cells at 18 h.p.e. Color-code represents the log fold-change (logFC) of gene expression in the indicated comparisons. Right – STRING representation of the gene network specifically modulated by ORF2. Bottom – Signaling pathways preferentially induced by the nuclear translocation of ORF2 at 18h p.e. Transcriptomic results stem from 4 independent electroporation experiments. Data are provided in the accompanying Source Data file.

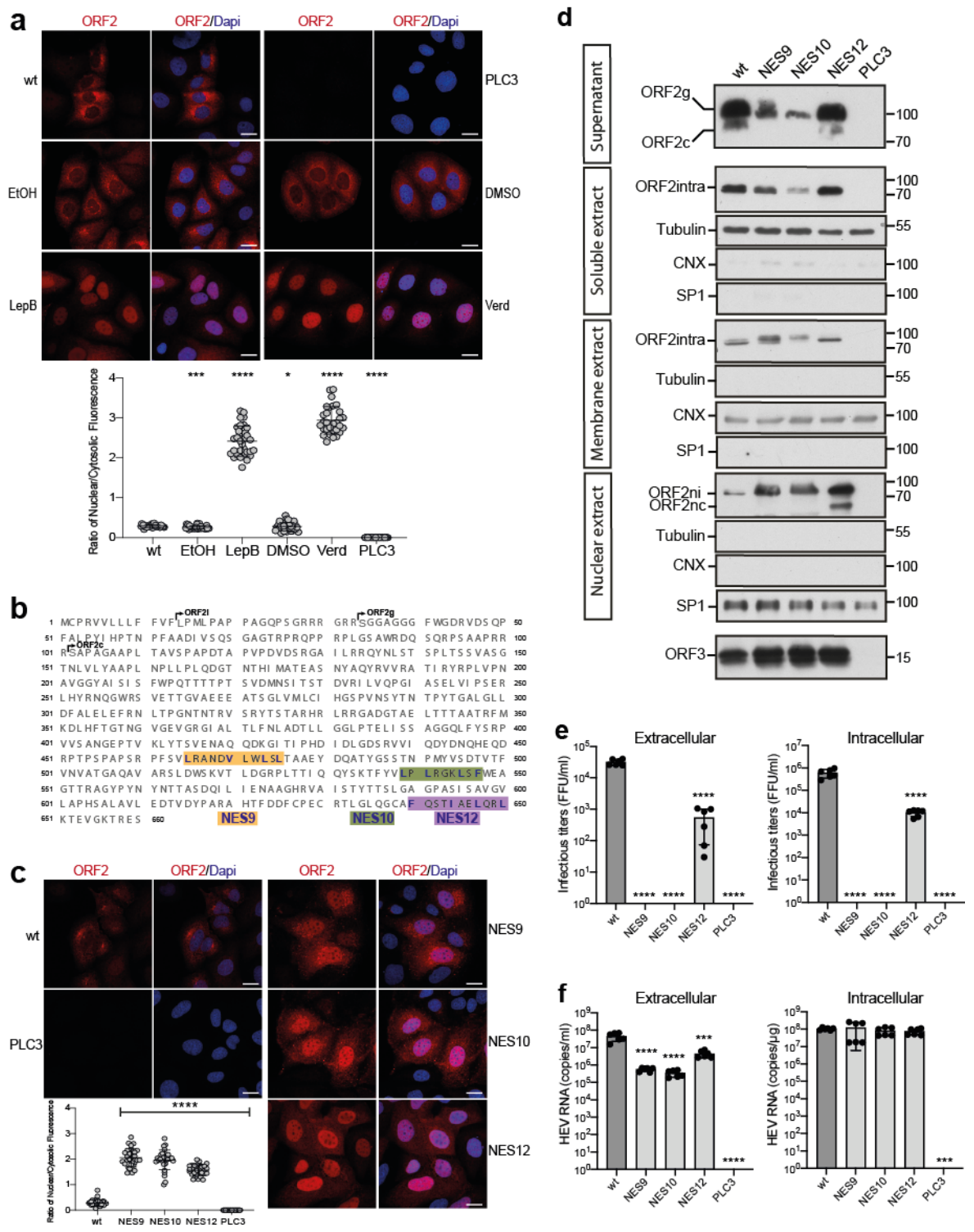

Figure 2: ORF2 active export from the nucleus is regulated by three nuclear export signal (NES) motifs. a, Analysis of nuclear export in inhibitors treated-PLC3/HEV-p6 cells. Cells were treated at 32 h.p.e with 20nM of Leptomycin B (LepB), 100nM of Verdinexor (Verd) or diluent (EtOH or DMSO, respectively) for 16h. Cells were processed for indirect immunofluorescence with the 1E6 anti-ORF2 Ab and analyzed by confocal microscopy (magnification x63). Red = ORF2; Blue = DAPI. **b,** Schematic representation of HEV-p6 ORF2 protein sequence highlighting the three studied NES motifs (*i.e.,* NES9,

NES10 and NES12). **c,** Subcellular localization of ORF2 NES mutants at 48h p.e. Red = ORF2; Blue = DAPI. In **a** and **c**, the scale bars correspond to 20μm, and nuclear/cytosolic fluorescence intensity quantification was done using ImageJ software (mean ± S.D., $n \geq 30$ cells, Kruskal-Wallis with Conover's test). **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$. **d,** Subcellular fractionation of PLC3/HEV-p6 expressing ORF2wt and NES mutants at 4 d.p.e. Fractionation was done using the subcellular protein fractionation kit for cultured cells. ORF2 proteins were detected by WB with 1E6 Ab. Glycosylated ORF2 (ORF2g), cleaved ORF2 (ORF2c), intracellular ORF2 (ORF2intra), nuclear ORF2intra (ORF2ni), nuclear and cleaved ORF2intra (ORF2nc) are indicated. ORF3 protein in cell lysates was detected with a rabbit anti-ORF3 Ab. Tubulin, ER marker Calnexin (CNX) and the transcription factor SP1 used as a nuclear marker, were also detected to check the quality of fractionation. Molecular mass markers are indicated on the right (kDa). **e,** Infectious titer determination in PLC3/HEV-p6 expressing ORF2wt or NES mutants. Extra- and intracellular viral particles were extracted at 10 d.p.e and used to infect naïve Huh7.5 cells for 3 days. Cells were next processed for indirect immunofluorescence. ORF2-positive cells were counted and each positive cell focus was considered as one FFU. Results were expressed in FFU/ml. **f,** HEV RNA quantification in PLC3/HEV-p6 expressing ORF2wt or NES mutants. Extra- and intracellular viral RNAs were quantified at 10 d.p.e by RT-qPCR. In **e** and **f**, n=6, mean ± S.D., Kruskal-Wallis with Conover's test, ***$p < 0.001$, ****$p < 0.0001$. Data are provided in the accompanying Source Data file.

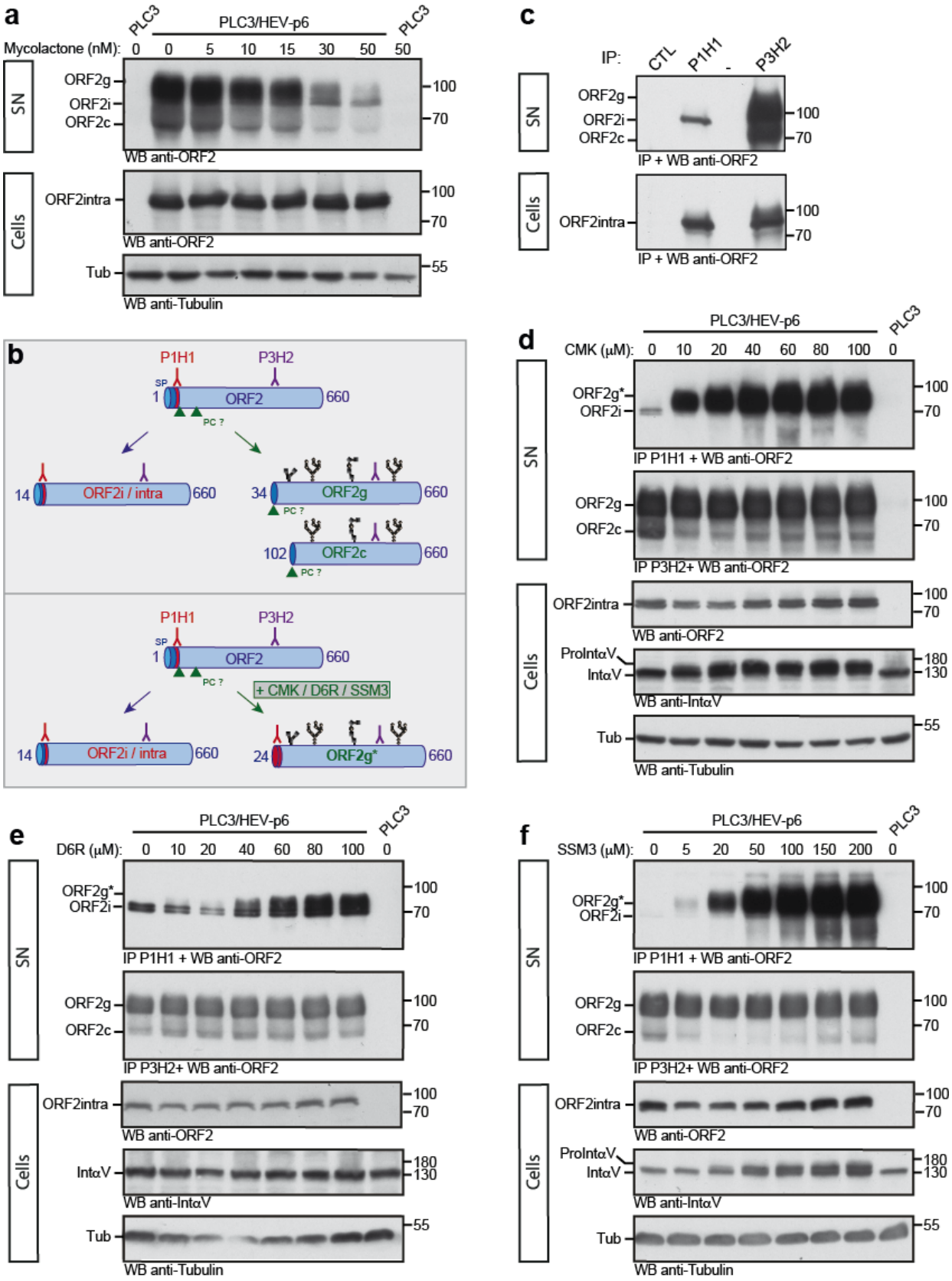

Figure 3: ORF2g/c forms are likely translocated through the Sec61 translocon and processed by a furin/proprotein convertase. a, Dose-response inhibition of ORF2g/c secretion in mycolactone-treated cells. PLC3/HEV-p6 and mock cells were treated for 24h with the indicated concentrations of

mycolactone (in nM) or maximal volume of the vehicle, ethanol (indicated as 0 nM). Supernatants (SN) and lysates (Cells) were collected and ORF2 proteins were detected by WB using the 1E6 Ab. Tubulin served as control protein loading. **b**, Schematic representation of ORF2i/g/c proteins and recognition sites of P1H1 and P3H2 antibodies used to discriminate the different ORF2 forms. SP, signal peptide. PC, proprotein convertase. Glycans are in black. **c**, Immunoprecipitation of ORF2 proteins in SN and lysates of PLC3/HEV-p6 cells by P1H1, P3H2 and isotype control (CTL) antibodies immobilized on magnetic beads. ORF2 proteins were detected by WB using the 1E6 Ab. **d-f**, PLC3/HEV-p6 cells were treated for 72h with the indicated concentrations of CMK, D6R or SSM3 (in μM) or DMSO diluent (indicated as 0 μM). Supernatants (SN) and lysates (Cells) were collected. SN were immunoprecipitated with P1H1 and P3H2 antibodies and ORF2 proteins were detected by WB using the 1E6 Ab. ORF2intra, αV-Integrin (IntαV) and Tubulin (Tub) were detected in cell lysates. αV-pro-integrin (ProintαV) corresponds to the non-maturated αV-integrin. ORF2g* corresponds to the ORF2g immunoprecipitated by the P1H1 Ab. Molecular mass markers are indicated on the right (kDa). Data are provided in the accompanying Source Data file.

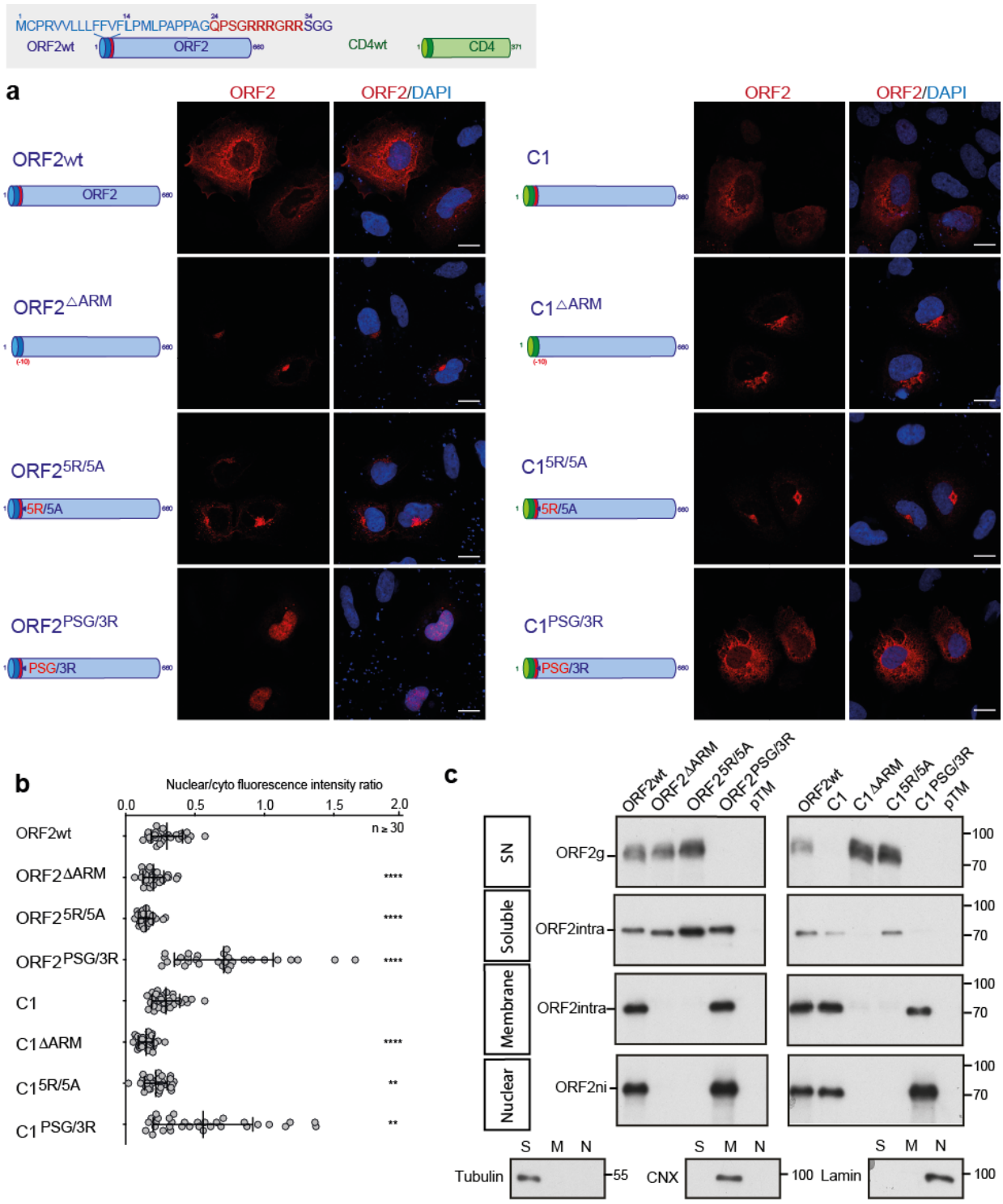

Figure 4: ORF2 addressing is regulated by its ARM. Schematic representation of ORF2wt and CD4wt proteins. ORF2 sequences are in blue. ARM residues are highlighted in red. CD4 sequences are in green **a**, H7-T7-IZ cells were transfected with pTM plasmids expressing wt, mutant or chimeric ORF2 proteins. Twenty-four hours post-transfection, cells were fixed and processed for ORF2 staining (in red). Nuclei are in blue. Representative confocal images are shown together with ORF2/DAPI merge images (magnification x63). Blue dots observed in some pictures are DAPI-stained transfected plasmids. A

schematic representation of each construct is shown on the left. Scale bar, 20μm. **b**, Nuclear to cytoplasmic ORF2 staining ratio in H7-T7-IZ cells expressing mutant and chimeric ORF2 proteins. Quantification was done using ImageJ software (mean ± S.D., $n \geq 30$ cells, Kruskal-Wallis with Conover's test). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$. **c,** Subcellular fractionation of H7-T7-IZ cells expressing mutant and chimeric ORF2 proteins at 24h post-transfection. Fractionation was done using the subcellular protein fractionation kit for cultured cells. ORF2 proteins were detected by WB with 1E6 Ab. Tubulin, Calnexin (CNX) and Lamin B1 were also detected to control the quality of fractionation. Molecular mass markers are indicated on the right (kDa). Data are provided in the accompanying Source Data file.

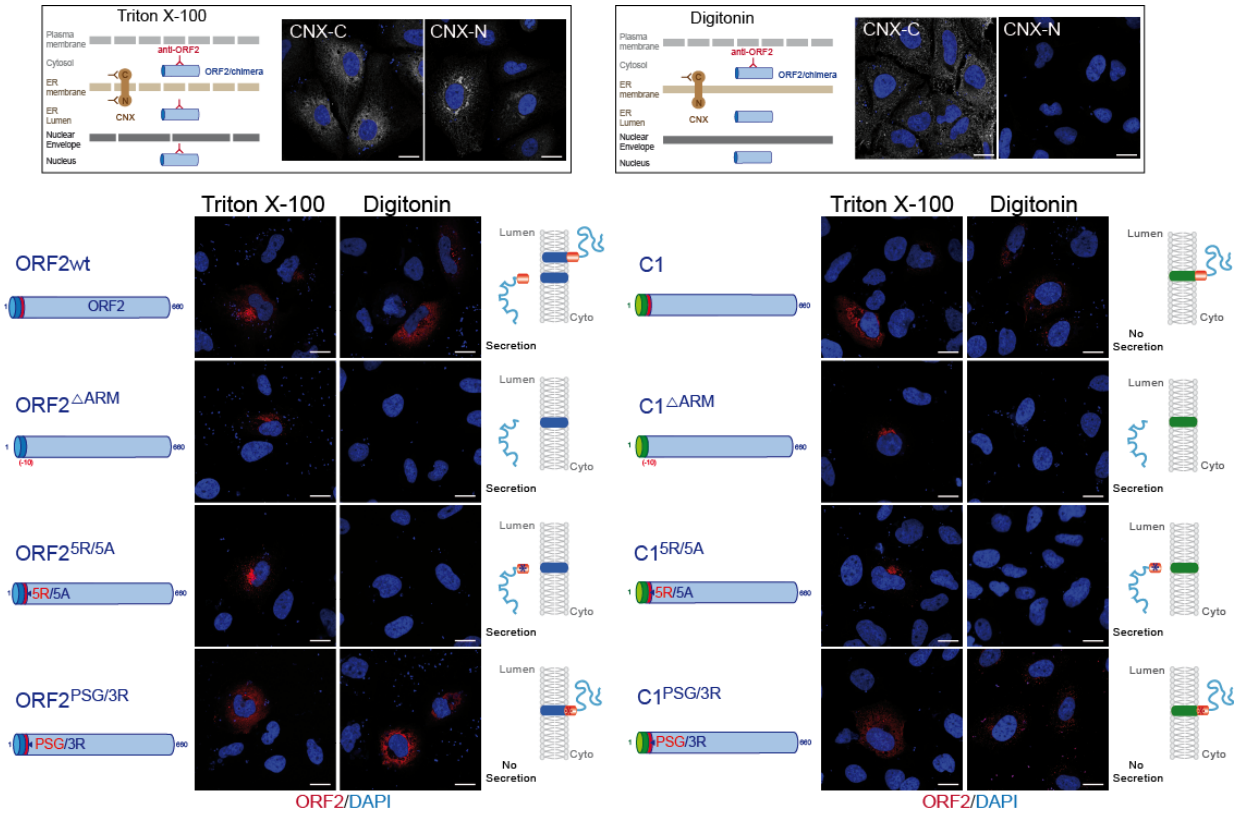

Figure 5: The ORF2 ARM controls the topology of ORF2 SP. A schematic representation of differential permeabilization process with Triton X-100 and Digitonin is shown. Representative images of the differential detection of two epitopes on the ER-membrane associated Calnexin (CNX) used to assess the permeabilization conditions are shown. H7-T7-IZ cells were transfected with pTM plasmids expressing wt, mutant or chimeric ORF2 proteins. Twenty-four hours post-transfection, cells were fixed, permeabilized with either Triton X-100 or Digitonin, and processed for ORF2 staining (in red). Nuclei are in blue. Representative confocal merge ORF2/DAPI images are shown (magnification x63). Blue dots observed in some pictures are DAPI-stained transfected plasmids. A schematic representation of each construct is shown on the left and its predicted topology on the right. Blue and red asterisks correspond to 5R/5A and PSG/3R mutations, respectively. Scale bar, 20μm.

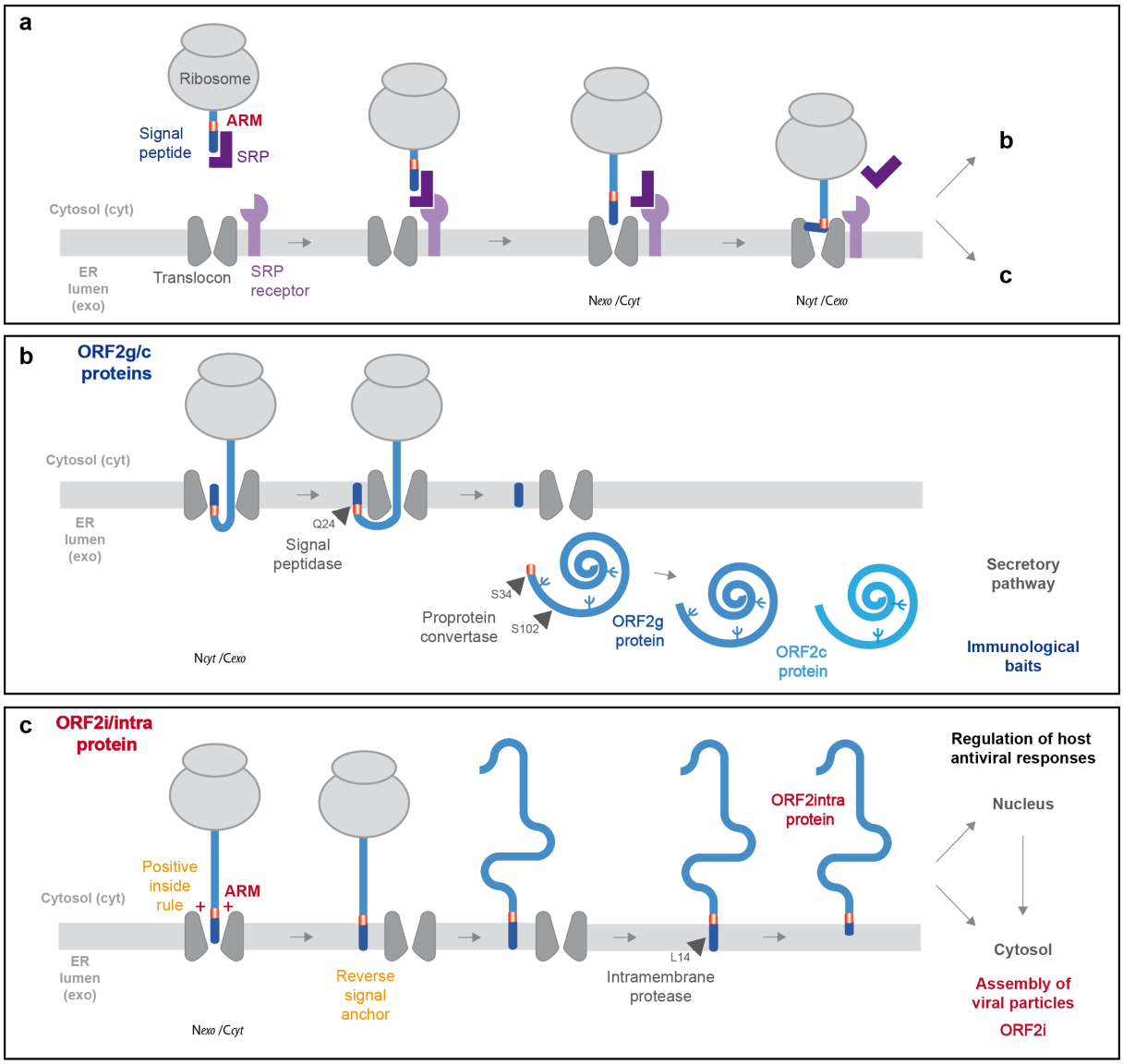

Figure 6: Model of ORF2 addressing regulation by ARM. a, The signal recognition particle (SRP) recognizes the hydrophobic signal peptide (SP) of the ORF2 nascent chain as it emerges from a translating ribosome. The ribosome-nascent chain-SRP complex is targeted to the membrane and interacts with the SRP receptor, resulting in release of the SP and docking of the ribosome–nascent chain complex to the Sec61 translocon. The ORF2 SP initially inserts head-on in an N_{exo}/C_{cyt} orientation, then inverts its orientation to N_{cyt}/C_{exo}. The C-terminal end of SP is exposed to ER lumen and is cleaved by signal peptidase, generating a new N-terminus. Translation then resumes, and the nascent ORF2 protein is translocated into the ER lumen where it is glycosylated and likely undergoes maturation by a proprotein convertase. This pathway generates the ORF2g/c forms (**b**). For a fraction of ORF2 nascent polypeptide chains, the ARM leads the ORF2 SP to retain its N_{exo}/C_{cyt} orientation and integrates as reverse signal-anchor, according to the positive-inside rule (**c**). The ORF2 protein anchored to the

1 cytosolic side of membrane is likely processed by an intramembrane protease to generate the
2 ORF2i/ORF2 intra protein that is translocated into the nucleus and assembles into viral particles.
3