Received: 9 April 2025 | Revised: 3 September 2025 | Accepted: 10 October 2025

DOI: 10.1002/pro.70355

RESEARCH ARTICLE

THE PROTEIN SOCIETY WILEY

Structural insights into SARS-CoV-2 nonstructural protein 4 (nsp4) biogenesis

José M. Acosta-Cáceres[1] | **Fabio Lolicato**[2,3] | **Laura Gadea-Salom**[1] | **Juan Ortiz-Mateu**[1] | **Diego Belda**[1] | **Luis Martínez-Gil**[1] | **María Jesús García-Murria**[1] | **Ismael Mingarro**[1]

[1]Departament de Bioquímica i Biologia Molecular, Faculty of Biological Sciences, Institut Universitari de Biotecnologia i Biomedicina (BIOTECMED), Universitat de València, Burjassot, Spain

[2]Heidelberg University Biochemistry Center, Heidelberg, Germany

[3]Department of Physics, University of Helsinki, Helsinki, Finland

Correspondence
Ismael Mingarro, Departament de Bioquímica i Biologia Molecular, Faculty of Biological Sciences, Institut Universitari de Biotecnologia i Biomedicina (BIOTECMED), Universitat de València, E-46100 Burjassot, Spain.
Email: ismael.mingarro@uv.es

Funding information
Generalitat Valenciana, Grant/Award Number: CIPROM/2022/062; Spanish Ministry of Science and Innovation, Grant/Award Numbers: PID2020-119111GB-I00, PID2023-152568NB-I00; Deutsche Forschungsgemeinschaft, Grant/Award Number: SFB-1638/1-511488495-Z01

Review Editor: Colin J Jackson

Abstract

The SARS-CoV-2 virus—responsible for the COVID-19 pandemic—requires a replication–transcription complex (RTC) for efficient RNA synthesis and viral propagation. One critical RTC component is nonstructural protein 4 (nsp4), a multipass transmembrane (TM) protein implicated in endoplasmic reticulum (ER) membrane rearrangements and double-membrane vesicle (DMV) formation. The membrane topology and functional role of nsp4 in SARS-CoV-2 remain unclear. Here we determined that SARS-CoV-2 nsp4 contains a partially cleaved signal peptide, three TM segments, an extracellularly oriented N-terminus (towards the ER lumen in human cells), and a cytoplasm-facing C-terminus. The non-canonical glycosylation sequon ($N^{131}IC$) in nsp4 is not glycosylated in mammalian cells. Molecular dynamics simulations based on these findings refined the structural predictions and folding of individual nsp4 molecules in lipid bilayers, elucidating its membrane disposition. This study advances our understanding of nsp4 membrane topology and its contributions to the formation of DMV and double-membrane-spanning pores, which are essential for viral RNA transport.

KEYWORDS
coronavirus, membrane insertion, nonstructural protein 4, SARS-CoV-2, topology

1 | INTRODUCTION

SARS-CoV-2—the causative agent of the devastating COVID-19 pandemic—is a *Betacoronavirus* from the Coronaviridae family. Coronaviruses are enveloped single-stranded positive-sense RNA viruses with large genomes, and SARS-CoV-2 contains one of the largest genomes among all known RNA viruses. The first two-thirds of the genome contains two large open reading frames (ORFs), ORF1a and ORF1b (Bar-On et al. 2020), which encode 16 nonstructural proteins (nsps) and are immediately translated upon viral entry and uncoating. The resultant polyproteins are pp1a, which contains nsps 1–11 and is translated from ORF1a; and pp1ab, which contains nsps 1–10 and 12–16 and is translated from both ORF1a and ORF1b after a −1 frameshift that allows continued translation beyond the stop codon of ORF1a (Brant et al. 2021; Malone et al. 2022). The nonstructural polyproteins are co- and post-translationally processed via proteolytic cleavage by two cysteine proteases that are located within nsp3 (papain-like protease; PLpro) and nsp5 (3C-like protease; 3CLpro) (Finkel et al. 2021; V'kovski et al. 2021). After cleavage, the nonstructural proteins form the viral replication–transcription complex (RTC) and participate in double-membrane vesicle (DMV) formation (Mohan

and Wollert 2021; Wolff et al. 2020a). The nsp3, nsp4, and nsp6 transmembrane (TM) proteins rearrange endoplasmic reticulum (ER) membranes and induce DMV formation (Angelini et al. 2013; Oudshoorn et al. 2017), which affects the regulation of host immunity and the modification of host cell endomembranes into replication organelles for viral replication and exit (V'kovski et al. 2021; Wolff et al. 2020a).

Notably, nsp4 is the largest TM nsp and the second largest membrane protein of SARS-CoV-2 (after the spike protein) and has been previously associated with membrane rearrangements in other coronaviruses (Angelini et al. 2013; Hagemeijer et al. 2014; Sakai et al. 2017). Amino acid substitutions in the nsp4 protein of murine hepatitis virus (MHV) were found to cause a temperature-sensitive phenotype, and further analysis of this infectious clone revealed dramatic decreases in DMVs and viral titre (Clementz et al. 2008). Subsequent analyses revealed that MHV nsp4 mutations resulted in reduced viral RNA synthesis and growth defects. Electron microscopy showed abnormal DMVs in cells infected with nsp4 mutants, and the degree of irregular DMVs was correlated with impaired RNA synthesis and viral growth. These findings highlight the importance of nsp4 in DMV organization and stability, which are crucial for efficient viral RNA synthesis and replication of MHV (Beachboard et al. 2015; Gadlage et al. 2010).

Furthermore, the nsp4 of SARS-CoV reportedly interacts with nsp3, and this interaction is necessary and sufficient to induce convoluted membranes and DMVs. The effects of mutations in the nsp3–nsp4 interacting region of nsp4 (H120N and F121L) have indicated that nsp3–nsp4 binding and induction of these membrane rearrangements correlate with viral propagation (Sakai et al. 2017). SARS-CoV-2 nsp3 and nsp4 are the minimal viral components required to induce DMV formation and to form a double-membrane-spanning pore that is essential to transport newly synthesized viral RNAs (Huang et al. 2024; Tabata et al. 2021; Wolff et al. 2020b; Zimmermann et al. 2023). Interaction between the luminal domains of these two proteins (stoichiometry 1:1) has been modeled (Klatte et al. 2022), and the impact of the TM helices has been analyzed in simulations. These simulations revealed that some models might represent artifacts, likely arising from inaccurate placement of the TM segments. More recent analyses show that nsp4 is the key organizer of DMV pores in hCoV-229E infected cells, spanning the double membrane and forming most of the pore lining, allowing newly synthesized mRNAs to exit the DMV into the cytoplasm (Chen et al. 2024). However, there is not yet a consensus regarding the membrane topology of nsp4, which hinders detailed structural insights regarding the monomeric protein and consequently regarding its functions in RTC formation and viral propagation.

In the present study, we analyze nsp4 topology in bacterial and human cells, and find that this protein contains a partially cleaved signal peptide-like sequence, three TM segments between residues 279 and 362, a large extracellularly oriented N-terminus (towards the ER lumen in human cells), and a bulky C-terminus (~140 residues) facing the cytoplasm. We also observe that the non-canonical glycosylation sequon (Asn-Ile-Cys) in the SARS-CoV-2 nsp4 sequence ($N^{131}IC$) is not glycosylated in mammalian cells. These data are used to generate a visual model of the monomeric structure relaxed through all-atom molecular dynamics (MD) simulations in the presence of lipid bilayers, providing insight into its possible conformation and membrane anchoring.

2 | RESULTS AND DISCUSSION

2.1 | Computer-assisted analysis of nsp4 sequence

The SARS-CoV-2 nsp4 protein has a length of 500 amino acids (residues 2764–3263 in pp1a and pp1ab), and an estimated molecular mass of approximately 56 kDa. Computer-assisted analysis of the nsp4 sequence for TM segments showed variable results depending on the method used, highlighting the importance of experimental demonstration (Duart et al. 2024). UniProt entry (P0DTC1) indicated six TM segments, while seven popular membrane protein prediction algorithms indicated between 3 and 7 TM segments (Figure 1a). The first TM segment has been suggested to function as a cleavable signal sequence in the previous SARS-CoV nsp4 (Oostra et al. 2007), and this prediction is supported by both MEMSAT and SPOCTOPUS (Figure 1b); however, it cannot function as an actual signal due to the presence of the preceding nsps (nsp1–nsp3) in the pp1a polypeptide. Notably, noncanonical cryptic (non-N-terminal or "internal") signal peptidase activities have been described as part of the maturation of viral polyproteins (Snapp et al. 2017; Zhang et al. 2016). Topology predictions displayed a higher level of agreement, with all algorithms (except Protter) predicting an amino-terminal (N-terminal) intracellular orientation, and all methods (except HMMTOP) predicting a carboxyl-terminal (C-terminal) intracellular (cytoplasmic) orientation (Figure 1b). Similarly to SARS-CoV nsp4, the homologous protein from SARS-CoV-2 contains a non-canonical glycosylation motif (Asn-Ile-Cys) at position N131, which has been suggested to be glycosylated in the SARS-CoV protein (Oostra et al. 2007).

2.2 | Nsp4 topology in bacterial cells

The membrane topology of nsp4 was first experimentally analyzed using a dual *pho-lac* reporter system, which comprises the *E. coli* alkaline phosphatase

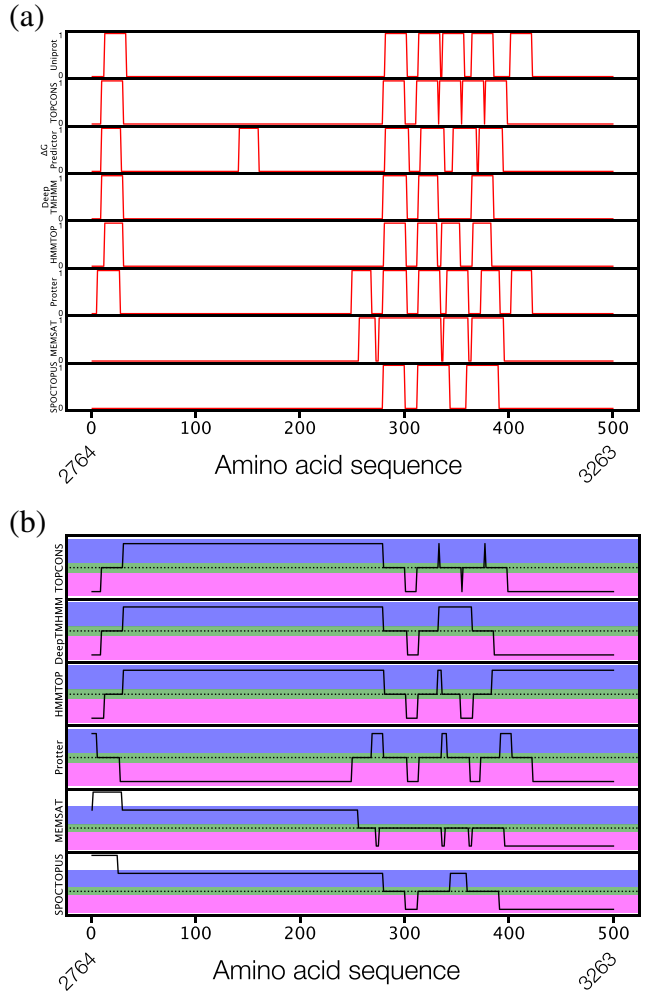

FIGURE 1 Computer analysis of the SARS-CoV-2 nsp4 amino acid sequence topology. (a) Comparison of predictions from seven different topology prediction tools, regarding the transmembrane (TM) regions of the SARS-CoV-2 nsp4 protein, based on its FASTA sequence. Predicted TM regions are assigned a value of 1, and non-TM regions a value of 0. (b) The overall topology predictions generated using the same set of tools. The predicted regions are color-coded, with blue indicating the periplasmic region, green the membrane-embedded segments, and pink the cytosolic region. The sequence is displayed using the nsp4 numbering system (1–500).

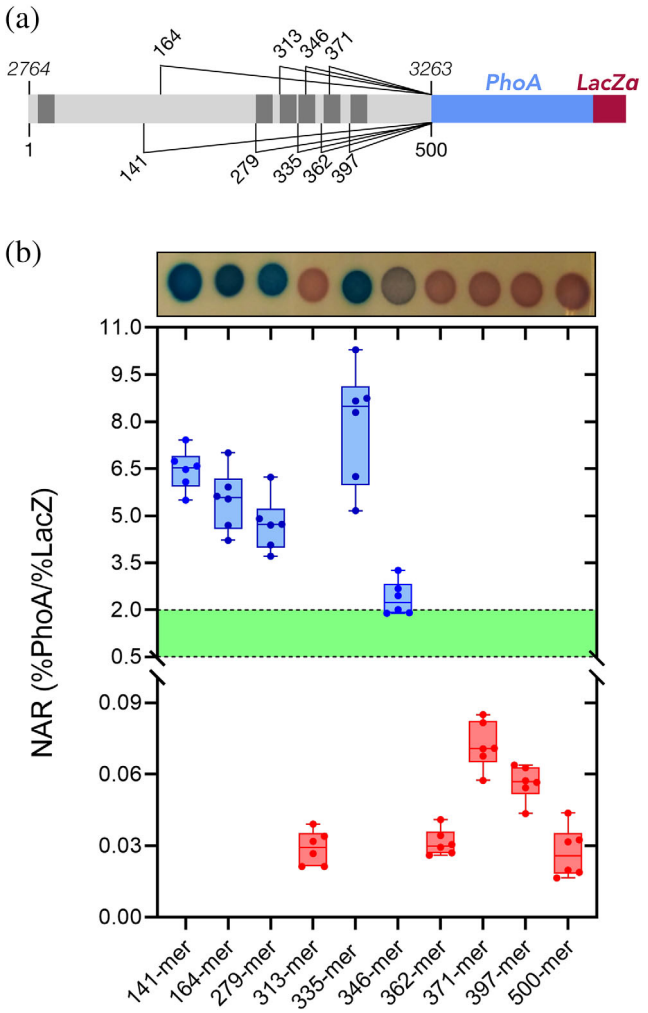

FIGURE 2 Pho-Lac dual reporter topology analysis in bacterial cells. (a) A Pho-Lac dual reporter is constructed using two topological markers with complementary properties: the *E. coli* alkaline phosphatase PhoA, which is active in the bacterial periplasm, and the β-galactosidase LacZ, which is active in the cytoplasm. When fused in-frame to the C-terminus of nsp4 variants, the dual nature of the reporter allows identification of the subcellular location of the fusion site. The positions of different fusions with the Pho-Lac reporter are indicated using nsp4 sequence numbering. The residue numbering for nsp4 in the pp1a polyprotein sequence is shown in italics. Transmembrane segments predicted at UniProt (P0DTC1) are highlighted in dark gray. (b) Experimental determination of nsp4 membrane topology. (top) *E. coli* DH5α cells expressing the different nsp4/Pho-Lac fusions were plated on indicator medium with two chromogenic substrates: Red-Gal (for β-galactosidase activity) and X-Pho (for phosphatase activity). Blue coloration of the colonies indicates high phosphatase activity, and thus a periplasmic location of the fusion point. Red coloration of the colonies indicates high β-galactosidase activity, and thus a cytosolic location of the fusion point. (bottom) Quantitative enzymatic assays of various nsp4/Pho-Lac fusions. The box plot (blue box, phosphatase activity; red box, β-galactosidase activity) represents the normalized activities (NAR), measured from liquid cultures of DH5α expressing the indicated nsp4/Pho-Lac fusions (position of the fusion indicated on the abscissa). NAR values of >2 indicate periplasmic location, between 2 and 0.5 indicate TM or interfacial (juxtamembrane) location, and <0.5 indicate cytosolic location. Solid dots represent the results of individual experiments.

(PhoA) followed in-frame by the α-peptide of *E. coli* β-galactosidase (LacZα) (Alexeyev and Winkler 1999; Duart et al. 2024; Karimova et al. 2009). This dual reporter cassette was fused in-frame to the C-terminus of different nsp4 truncated variants (Figure 2a), allowing us to determine the protein topology. These PhoA-LacZα chimeric proteins display high phosphatase and low β-galactosidase activities when targeted to the periplasm, and high β-galactosidase and low phosphatase activities when targeted to the cytoplasm (Duart et al. 2024; Karimova and Ladant 2017). We also constructed in-frame fusions of the *pho-lac* reporter after 10 selected nsp4 codons (141-mer,

164-mer, 279-mer, 313-mer, 335-mer, 346-mer, 362-mer, 371-mer, 397-mer, and full-length 500-mer) in the plasmid pKTop (see section 4). Next, *E. coli* cells were transformed with plasmids expressing the different nsp4/*pho-lac* variants. The resulting transformants were analyzed on dual-indicator media containing a blue chromogenic substrate for phosphatase activity, and a red chromogenic substrate for β-galactosidase activity.

Cells expressing 141-mer, 164-mer, and 279-mer fusions displayed a blue phenotype, indicating the periplasmic location of the *pho-lac* reporter, and thus also of the corresponding nsp4 residues E141, G164, and D279 (Figure 2b, top). This localization suggests either the translocation (or cleavage) of the first hydrophobic region (residues 10–32), or the transmembrane insertion of this hydrophobic segment with an N-terminal cytoplasmic and C-terminal periplasmic (N_{cyt}/C_{per}) orientation. Fusion of the *pho-lac* reporter after H313 (313-mer) yielded red colonies on the dual-indicator media, indicating intracellular localization of H313, and suggesting the presence of a TM segment between residues 279 and 313. When the *pho-lac* reporter was fused after residue Y335 (335-mer), the chimeric protein exhibited high phosphatase activity (deep blue), consistent with periplasmic localization of the Y335 residue, and suggesting a TM segment between residues 313 and 335. Fusion of the *pho-lac* reporter after residue Y346 of nsp4 yielded blue-purple colonies on the dual-indicator medium, suggesting that residue Y346 is localized in (or closely associated with) the membrane, rather than in the periplasm or cytosol. When the *pho-lac* reporter was fused after residues H362, P371, Y397, or Q500 (full-length) of nsp4, the chimeric proteins exhibited high β-galactosidase activity (red), consistent with cytoplasmic localization of the corresponding nsp4 residues (Figure 2b), and suggesting a TM segment between residues 335 and 362, and a bulky C-terminus (residues 363–500) facing the cytoplasm.

We corroborated these data with quantitative measurements in liquid media, using the standard colorimetric substrates: p-nitrophenyl phosphate (pNPP) for phosphatase, and ortho-nitrophenyl-β-D-galactoside (ONPG) for β-galactosidase. The phosphatase and β-galactosidase enzymatic activities of each hybrid protein were normalized (to the highest value observed within each set of fusions) to obtain the so-called normalized activity ratio (NAR) (Alexeyev and Winkler 1999). As shown in Figure 2b (bottom), the quantitative data confirm the results observed on indicator agar plates (top), suggesting the presence of an N-terminal TM (or signal peptide-like) segment, followed by a large periplasmic loop, three TM segments (among residues 279–313, 313–335, and 346–362), and a cytoplasmic C-terminus (residues 363–500). Furthermore, our data suggest that the TM segments are sequentially inserted into the bacterial membrane. In other words, after exiting the ribosome exit tunnel in a helical conformation (Bañó-Polo et al. 2018), each TM segment is likely inserted into the hydrophobic core of the lipid bilayer by the multipass translocon (Hegde and Keenan 2024) in an ordered and sequential fashion (Whitley et al. 2021; Whitley and Mingarro 2014).

2.3 | The first hydrophobic region is recognized as transmembrane (or signal peptide-like) by the ER translocon

To investigate whether the first hydrophobic region of nsp4 is recognized by the ER translocon in eukaryotic cells, we used a glycosylation-based reporter system (C_LTM). The tested sequence (residues 10–32) was fused in-frame to the C-terminus of the constant domain of the antibody λ light chain C_L (Feige and Hendershot 2013), harboring an engineered glycosylation site (G1) (Duart et al. 2022), and followed by a C-terminal sequence carrying another glycosylation site (G2) and a c-myc epitope (Figure 3a, top). In this system, G1 is always glycosylated due to native translocation of the λ light chain, while G2 is glycosylated only upon translocation (non-insertion) of the analyzed sequence across the ER membrane (Figure 3a, bottom). When Hek293T cells were transfected with a construct harboring residues 10–32 from nsp4, "insulated" from the surrounding sequence by N- and C-terminal GGPG- and -GPGG tetrapeptides (Hessa et al. 2005), we predominantly observed singly glycosylated (reporting membrane integration) forms (Figure 3b, lane 6). Sugar moiety presence was confirmed by treatment with endoglycosidase H (EndoH) (Figure 3b, lane 5), which is a highly specific enzyme that cleaves *N*-linked oligosaccharides (Duart et al. 2021). Hek293T cells were transfected in parallel with constructs harboring either a TM sequence (Figure 3b, lanes 1 and 2) (Martinez-Gil et al. 2010) or a translocated sequence (Figure 3b, lanes 3 and 4) (Martinez-Gil et al. 2011). These findings indicated that the first hydrophobic region from nsp4 is recognized by the ER translocon like a truly TM segment (or signal sequence-like) in human cells.

2.4 | Nsp4 cleavage by signal peptidases and glycosylation status

To study the co-translational processing of nsp4, we transfected Hek293T cells with a plasmid containing the full-length nsp4 sequence, flanked by engineered glycosylation sites at the N- and C-terminus as reporters, and a c-myc tag (Figure 4a). MEMSAT and SPOCTOPUS predictions indicated that the first hydrophobic TM domain of nsp4 might be recognized as a cleavable sequence (Figure 1b). Signal peptides are widely believed to be co-translationally cleaved;

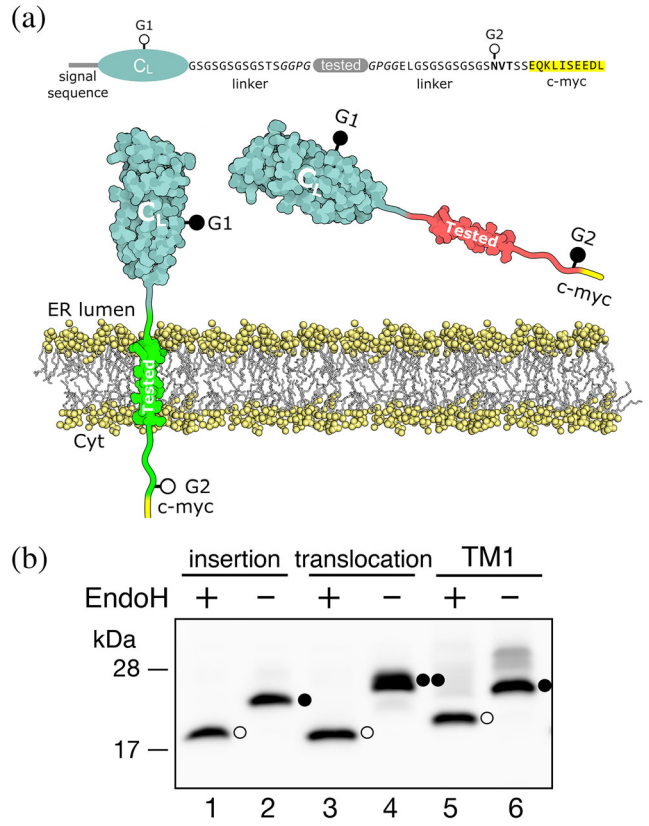

FIGURE 3 The first hydrophobic region is recognized by the translocon as a transmembrane segment. (a) (top) Schematic illustration of the utilized C_LTM construct. The antibody λ light chain (C_L) harboring a designed glycosylation site (G1) is preceded by its signal sequence, and connected by flexible linkers to the tested sequence, followed by a C-terminal glycosylation site (G2) and a *c*-myc tag (EQKLISEEDL, yellow). (bottom) Scheme depicting the main features of the C_LTM insertion assay. Black dots represent glycosylated sites. White dots represent non-glycosylated sites. (b) Exemplary western blot ($n = 3$) of Hek293T cells transfected with the indicated constructs. Cell lysates in the presence (+) or absence (–) of the glycan-removing enzyme endoglycosidase (EndoH) were immunoblotted with anti-*c*-myc antiserum. The experiment included insertion (KVLLVTGVLGLLLLIKWK, lanes 1 and 2) and translocation (YNLYIYAFFASSLVHGFI, lanes 3 and 4) control sequences (Martinez-Gil et al. 2010). Non-glycosylated proteins are indicated by white dots. Single or double glycosylated proteins are indicated by one or two black dots, respectively.

however, post-translational cleavage can occur for the maturation of viral polyproteins (Snapp et al. 2017). The N-terminal hydrophobic region of SARS-CoV nsp4 is (partially) processed by signal peptidase (Oostra et al. 2007). Additionally, the nsp4 sequence contains a non-canonical glycosylation motif (NXC) at position 131. It has been suggested that an equivalent glycosylation sequon functions as an *N*-glycan attachment site in the homologous nsp4 from SARS-CoV (Oostra et al. 2007).

As shown in Figure 4b (lane 1), cells that were lysed and processed for immunoprecipitation (anti-c-myc) yielded a double band of about 56 kDa. EndoH treatment of these samples confirmed that neither of these two protein bands corresponds to *N*-glycosylated forms of nsp4. To test whether the faster-migrating band might result from signal peptide-like cleavage of nsp4 (as predicted by some algorithms), we mutated the predicted cleavage region, replacing natural residues (^{28}TPVH^{31}) with alanines (nsp4^{Ala}; Figure 4b, lanes 3 and 4) and phenylalanines (nsp4^{Phe}; Figure 4b, lanes 5 and 6). When cells were transfected with these nsp4 variants, the faster-migrating band partially (nsp4^{Ala}) or totally (nsp4^{Phe}) disappeared, and the samples were not affected by EndoH treatment (Figure 4b, lanes 3–6). These findings demonstrated that nsp4 from SARS-CoV-2 is partially cleaved by signal peptidases in human cells.

To further corroborate the signal peptidase (SP) cleavage, we treated nsp4-transfected cells with the SP inhibitor cavinafungin (Estoppey et al. 2017; Zanotti et al. 2022). Western blot analysis revealed that cavinafungin treatment strongly reduced nsp4 processing, enhancing the proportion of full-size precursor (Figure 4c), together with mutations of the predicted cleavage site (Figure 4b), confirming membrane protein cleavage of the nsp4 and lack of glycosylation in eukaryotic cells. Notably, the first hydrophobic TM domain of SARS-CoV-2 nsp4 (residues 1–31) is absent in a recent cryo-ET structure (PDB: 8YAX) of DMV pores assembled in Hek293F cells (Huang et al. 2024), suggesting that SP cleavage removes this segment during or shortly after insertion but prior to nsp3–nsp4 pore assembly.

Comparative sequence analysis of the nsp4 of SARS-CoV-2 and the other six known human coronaviruses revealed that the non-canonical glycosylation motif was only present in the SARS coronaviruses (Figure 4d), creating an aggregation-prone region (Figure S1, Supporting Information). To confirm the non-glycosylation status of nsp4, we designed two mutants. One was designed to preclude *N*-glycosylation by replacing acceptor asparagine with non-acceptor glutamine (N131Q). The other mutant (C133S) was designed to convert the original non-canonical site (NXC) to a canonical one (NXS/T, where X represents any residue except Pro). Only the C133S mutant received the sugar moiety (Figure 4e), demonstrating that this region is facing the lumen of the ER and verifying that the wild-type nsp4 SARS-CoV-2 protein containing a non-canonical glycosylation site is not modified in human cells.

2.5 | Nsp4 topology in human cells

To experimentally determine the topology of nsp4 in eukaryotic cells, we applied a glycosylation mapping approach (Nilsson and von Heijne 1990; Tamborero et al. 2011). We prepared a series of protein truncates that contained an optimized C-terminal glycosylation

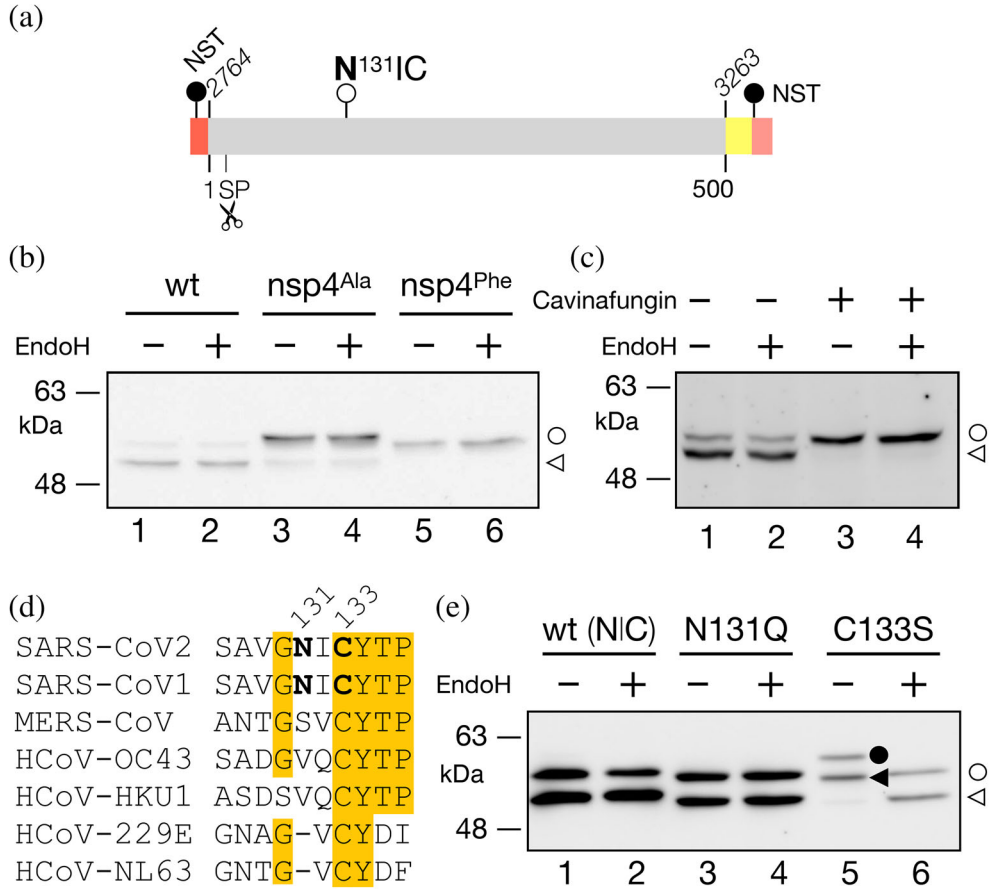

FIGURE 4 Nsp4 is partially processed and non-glycosylated in mammalian cells. (a) Schematic representation of the nsp4 sequence, with residue numbering in the pp1a polyprotein sequence shown in italics. Designed acceptor glycosylation sites (NST, Asn-Ser-Thr) at the N- and C-terminus are indicated as black dots. The non-canonical glycosylation site ($N^{131}IC$) is indicated as a white dot. The *c*-myc epitope is shown in yellow. (b) The nsp4 N- and C-terminal ends are located at the cytosol. To determine the N- and C-terminal location of the protein, Hek293T cells were transfected with the above c-myc-tagged nsp4 sequence harboring optimized glycosylation sites at the N- and C-terminal ends. Constructs encoding wild-type (wt; lanes 1 and 2) plus two mutant variants designed to preclude putative signal peptide cleavage ($nsp4^{Ala}$, lanes 3 and 4; and $nsp4^{Phe}$, lanes 5 and 6) were treated with EndoH (+) or mock treated (–), and then analyzed by SDS-PAGE and western blot (anti-*c*-myc). The white dot and triangle indicate non-glycosylated full-length and cleaved molecules, respectively. (c) Immunoblot of c-myc-tagged nsp4 in Hek293T cells treated with the SPC inhibitor cavinafungin (1 μM) and EndoH where indicated. Empty circle and empty arrowhead indicate full-length and SPC-processed nsp4 forms, respectively. (d) Among the seven human coronaviruses, the non-canonical glycosylation site ($N^{131}IC$) found in SARS-CoV-2 is only conserved in SARS-CoV. From top to bottom, the amino acid sequences for nsp4 from SARS-CoV-2, SARS-CoV (UniProt P59637), MERS-CoV (UniProt K9N5R3); HCoV-OC43 (UniProt Q4VID3), HCoV-HKU1 (UniProt Q0ZJ83), HCoV-229E (UniProt P19741), and HCoV-NL63 (UniProt Q5SBN7). Residues with over 50% conservation among human coronavirus are highlighted in orange. (e) The non-canonical glycosylation site is not modified in mammalian cells. Hek293T cells were transfected with the wild-type construct (lanes 1 and 2), a construct harboring the non-glycosylable N131Q mutant (lanes 3 and 4), or a mutant harboring a canonical glycosylation site (C133S, lanes 5 and 6). White dots and triangles indicate non-glycosylated full-length and cleaved molecules, respectively. Black dots and triangles indicate glycosylated full-length and cleaved molecules, respectively.

tag (NSTMGM) (Bañó-Polo et al. 2011), or a non-glycosylable tag (QSTMGM). The cells transfected with a construct containing the N-terminal half of nsp4 (279-mer) yielded three protein bands, with the slowest-migrating species containing *N*-linked sugars (Figure 5, lanes 1–3), suggesting that residue D279 is located in the ER lumen. Interestingly, the more intense band corresponded to the presence of partially signal peptide-digested glycosylated molecules and signal peptide-non-digested unglycosylated molecules. The addition of a glycan moiety increases the protein's molecular weight by about 2.5 kDa, which coincides with the decrease of molecular mass observed for the signal peptide-like cleaved protein forms. The observation of non-glycosylated truncated 313-mer polypeptides (lanes 4–6) indicated a cytoplasmic location of residue H313, and suggested the presence of a TM segment between residues 279 and 313, which agrees with the nsp4 topology observed in bacterial cells. On the other hand, truncated 335-mer and 346-mer polypeptides were efficiently glycosylated (lanes 7–9 and 10–12, respectively), demonstrating the ER luminal disposition of Y335 and Y346, and suggesting the presence of a TM domain between residues 313 and 335. Finally, we found that longer

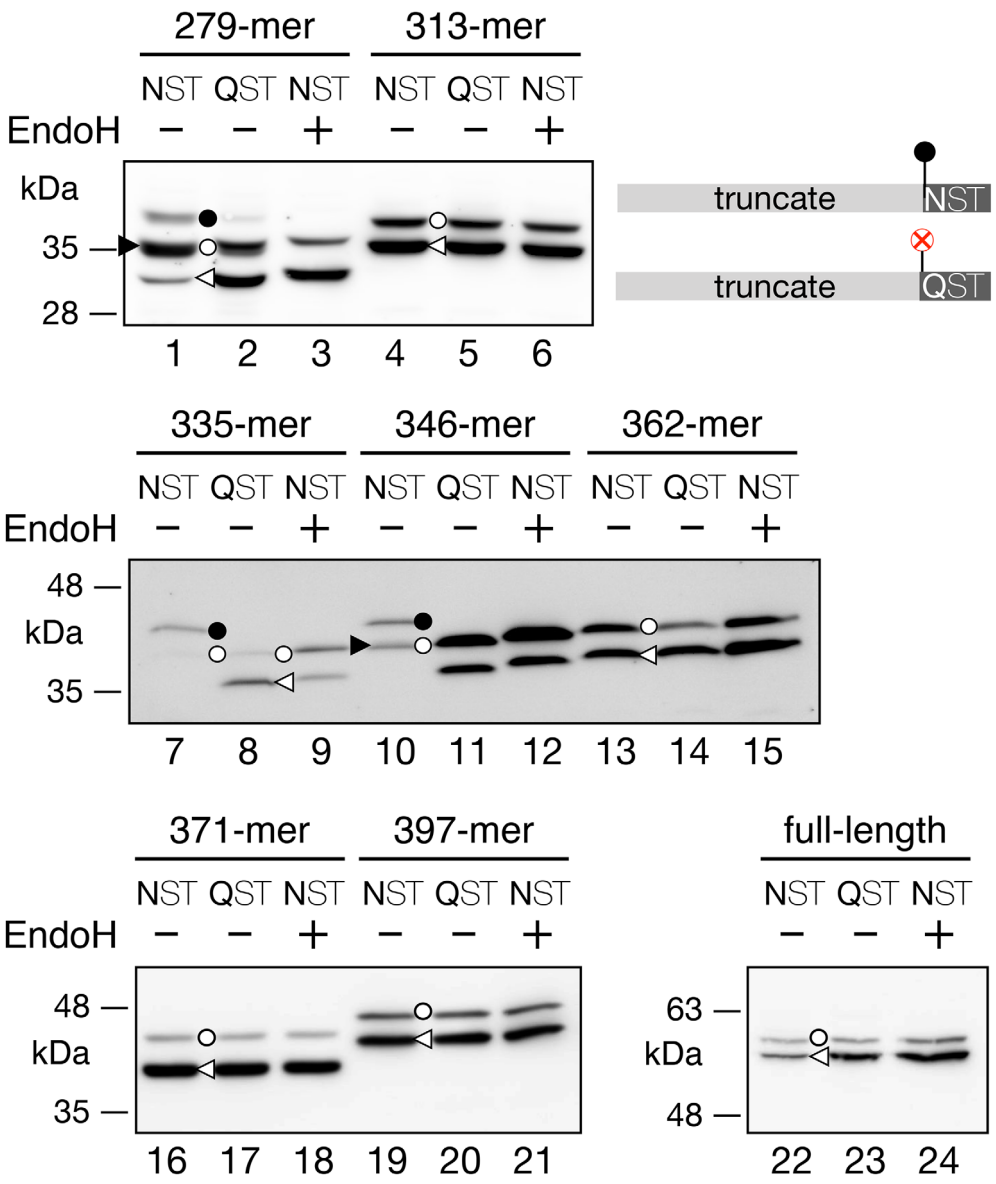

FIGURE 5 Nsp4 membrane topology in mammalian cells. Hek293T cells were transfected with truncated nsp4 constructs, in which fused C-terminal N-glycosylation (NST) or non-glycosylable (QST) tags provide a simple readout for topology. Exemplary western blots of relevant constructs are shown ($n > 4$). Samples were treated with EndoH (+) or mock treated (–), as indicated. White and black dots indicate non-glycosylated and glycosylated protein molecules, respectively. White and black triangles indicate non-glycosylated and glycosylated signal peptide-cleaved molecules, respectively.

truncated molecules (362-mer, 371-mer, 397-mer, and full length) rendered non-glycosylated forms (lanes 13–15, 16–18, 19–21, and 22–24, respectively), indicating that the C-terminus of nsp4 is a large region (~140 residues) located on the cytoplasmic face of the membrane, and that the last TM segment lies between residues 335 and 362. These results suggest that for nsp4 in eukaryotic cells, the TM segments are sequentially inserted into the ER membrane, such that the insertion of each TM segment does not require the presence of downstream TM domains.

2.6 | Computational structural modeling of nsp4 in membrane

We used AlphaFold2 (Jumper et al. 2021) to predict the structure of the SARS-CoV-2 nsp4 protein (Uniprot ID P0DTC1, residues 2764–3263), as a template. However, the AlphaFold2 (AF2) prediction does not account for the presence of the membrane, yielding an N-terminal α-helix configuration that is incompatible with proper membrane insertion, compared to other TM helices (Figure S2a). To address this issue, we truncated the protein (residues 1–271) and embedded the N-terminal helix into a POPC-CHOL (7:3) membrane. Next, we performed atomistic MD simulations and selected a snapshot in which the extra-cytoplasmic portion of the protein was near the membrane (Figure S2b, left), adopting a configuration suitable for seamless integration with the rest of the protein.

The AF2 prediction aligned well with our experimental data, with the exception of two residues: H362 and P371. Our topological experiments indicated that both residues should reside in the cytosolic region; however, in the AF2 model, H362 is positioned within the

membrane core, and P371 is located at the surface of the extra-cytoplasmic region. This discrepancy may be attributed to the inability of AF2 to account for the membrane environment. Specifically, at position D356—which is situated in the middle of the TM helix in the current setup—AF2 predicts a kink that shifts the helix containing H362 and P372 towards the extra-cytoplasmic region. We modeled this region by generating a straight helix for residues 335–373. The remaining portion of the protein (residues 374–500) was shifted along the z-axis to align with the new protein structure (Figure S2b, refined model). In this configuration, the helix spanning residues 374–398, which was previously predicted to be a TM helix, is exposed to the cytosolic region. Both the full-length model and the signal peptide-cleaved nsp4 variant (residues 1–32 removed) were embedded in two distinct membrane compositions (with and without 30 mol% cholesterol), with each system prepared in triplicate to ensure independent initial configurations. All systems were subjected to a 1100-ns atomistic MD simulation—with D356 modeled in its protonated state because it is positioned in the hydrophobic core of the membrane. This modeling now places the 374–398 helix in a position where it interacts with both the membrane surface and the cytosolic soluble domain (Figure S2c). Interestingly, the 374–398 helix is a highly interfacial sequence with an apparent free energy of interfacial disposition (ΔG^{int}_{app}) value of −5.8 kcal/mol, according to a recently determined scale for the translocon-mediated insertion of helices at the membrane interface (Grau et al. 2025). Then, this region could dynamically, and possibly post-translationally, partition from the interface into the membrane for subsequent DMV formation and nsp3–nsp4 pore assembly.

As shown in Figure S2c–d, the PPM server (Lomize et al. 2012), used to predict the orientation of proteins within membranes, places the protein in two distinct configurations depending on the presence or absence of the N-terminal signal peptide. This observation suggests a potential role for the soluble N-terminal domain in interacting with other membrane compartments, possibly facilitating DMV formation. To test the hypothesis that the N-terminal domain is capable of adopting both membrane-bound and detached states, we performed 1100-ns atomistic molecular dynamics (MD) simulations of the protein embedded in POPC membranes, both with and without 30 mol% cholesterol.

MD simulations revealed that, without a signal peptide-like sequence, the extra-cytoplasmic soluble domain exhibited increased mobility. While it remained predominantly associated with the membrane (maintaining a position similar to that observed prior to the removal of the N-terminal sequence), it also sampled configurations in which it was positioned farther away from the bilayer (Figure S3b). This dynamic behavior suggests that the soluble domain may transiently disengage from the membrane, potentially facilitating interactions with other membrane compartments and/or proteins like nsp3 as recently observed (Chen et al. 2024; Huang et al. 2024) and contributing to DMV formation. Notably, the *Z*-distance distribution (measured as the distance between the center of mass of residues 69–79 and the phosphate atoms of the upper leaflet) was slightly broader and shifted towards higher values in the presence of cholesterol (Figure S3c). This indicates increased mobility under cholesterol-rich conditions. One possible explanation is that cholesterol promotes a more upright orientation of the TM helices, reducing their tilt angle and enhancing domain freedom through membrane stiffening and tighter lipid packing.

3 | CONCLUSIONS

Coronaviruses replicate and transcribe their RNA using RTCs, which comprise nsps and are associated with DMVs. In coronaviral DMVs, the nsp3–nsp4 TM multiprotein complex forms the pore required for viral mRNA to exit into the cytoplasm. It has been proposed that nsp4 plays a central role in bringing the two membranes together (Chen et al. 2024). In infected cells, each nsp4 molecule is generated after cleavage of the pp1a or pp1ab polyproteins and becomes embedded in the ER membrane prior to complex formation. In the present study, we investigated SARS-CoV-2 nsp4 membrane topology, processing, and glycosylation. Moreover, we used the experimental results as constraints for monomeric nsp4 membrane structure prediction.

Our results demonstrated that the nsp4 is embedded in prokaryotic and eukaryotic membranes with the same topology. We used the experimental constraints to refine the structural model of the protein in the presence of membranes mimicking the ER lipid composition. Our modeling suggests that the protein is anchored into the membrane with its first hydrophobic domain (residues 3–28) inserted in a N_{cyt}/C_{out} orientation. This is followed by a large extra-cytoplasmic loop (residues 29–276), a second TM segment (residues 277–305) oriented N_{out}/C_{cyt}, a third TM segment (residues 312–332) with a reversed N_{cyt}/C_{out} orientation, a fourth TM segment (residues 338–367) oriented N_{out}/C_{cyt}, and a cytoplasm-exposed C-terminal soluble domain (residues 368–500). In the membrane embedded region, we observed notable discrepancies between our refined model and the arrangement of nsp4 in the recently solved nsp3–nsp4 DMV pore structure (Huang et al. 2024). In the pore structure—which closely matches the AlphaFold2 prediction (Figure S2b, top right)—TM3 and TM4 form a large re-entrant loop. In contrast, our refined model shows these elements merging into an extended helix (Figure S2b, bottom right), resulting in helix 374–398 being positioned at the membrane interface rather than spanning the membrane. We propose that our model may represent the

membrane topology of "monomeric" nsp4 molecules inserted into a single membrane, whereas the reported pore structure likely reflects the organization required for oligomerization and nsp3–nsp4 complex formation across double membranes.

Additionally, we demonstrated that the first TM segment is recognized and partially cleaved by cellular signal peptidases, generating nsp4 molecules having a significantly more flexible N-terminal extra-membrane domain, to interact with its nsp3 partner, and to thereby form viral DMVs. Finally, we proved that the non-canonical glycosylation site present in the SARS-CoV-2 nsp4 sequence (N^{131}IC) is not *N*-glycosylated upon integration of the protein into ER membranes, which is contrary to descriptions of the closely related SARS-CoV (Oostra et al. 2007). Further research is needed to fully elucidate the impact of our monomeric nsp4 structural model in nsp3–nsp4 pore complexes and, correspondingly, in future interventions for the development of antivirals for betacoronavirus.

4 | MATERIALS AND METHODS

4.1 | Enzymes and chemicals

The DNA sequence encoding nsp4 protein was a kind gift from Nevan Krogan (UCSF). All oligonucleotides were purchased from Macrogen (Seoul, South Korea). Selected DNA fragments were amplified by PCR using NZYTaq II polymerase (NZYTECH). The amplified DNA fragments were digested with the *Xba*I and *Sac*I restriction endonucleases (New England BioLabs Inc.), and then ligated with T4 DNA-ligase (NZYTECH) or In-Fusion® HD Cloning Kits (Japan). We used DNA plasmid clean-up and PCR purification kits from Thermo Scientific (USA). *Dpn*I was purchased from Thermo Scientific (USA). We obtained the QuikChange PCR mutagenesis kit from Stratagene (La Jolla, CA). All constructs were verified by DNA sequencing (Macrogen). Bacterial growth media (Condalab S.A.) was supplemented with ampicillin and kanamycin (Tokyo Chemical Industry), disodium 5-bromo-4-chloro-3-indolyl phosphate (BCIP), 6-Chloro-3-indolyl-β-D-galactoside (Red-Gal), and isopropyl β-d-1-thiogalactopyranoside (IPTG) (Acros Organics S.A.). The chemical reagents for M63 medium [KH_2PO_4, $(NH_4)_2SO_4$, Fe_2SO_4, and $MgSO_4$], β-galactosidase assay buffer PM2 (Na_2HPO_4, $NaHPO_4$, $MgSO_4$, and $MnSO_4$), wash buffer for phosphatase assay WB (Tris–HCl and $MgSO_4$), and phosphatase assay buffer PM1 (Tris–HCl, $ZnCl_2$, and iodoacetamide), as well as sodium dodecyl sulfate (SDS), ONPG, Na_2CO_3, p-nitrophenyl phosphate (pNPP), Tris–HCl, and NaOH were obtained from Sigma-Aldrich and Fisher Bioreagents (USA). HEK293T cells were maintained with Dulbecco's Modification of Eagle's Medium 1X with 4.5 g/L glucose, 0.25% Trypsin–EDTA 1X, and phosphate-buffered saline (PBS) pH 7.4 (1X) from Gibco (USA). Lipofectamine™ 2000 Transfection Reagent was obtained from Invitrogen (USA), rabbit anti-c-myc and anti-rabbit-HRP conjugate antibodies from SIGMA (USA), and nitrocellulose blotting membrane from Amersham™ (Germany).

4.2 | Bacterial strains, cell lines, and plasmids

Table 1 describes the cells and plasmids used in this study. The *E. coli* strain DH5α (which is capable of α-complementation of LacZ and *phoA*-deficient Δ*phoA*) was used in all routine cloning experiments and as a host for investigations of membrane protein topology. We used the pKTop plasmid carrying the dual pho-lac reporter gene—which encodes the *E. coli* alkaline phosphatase fragment $PhoA_{22–472}$, fused in-frame with the α-peptide of *E. coli* β-galactosidase $LacZ_{4–60}$—allowing protein topology determination in vivo. In this system, periplasmic localization of the reporter leads to high alkaline phosphatase activity and low β-galactosidase activity, whereas cytosolic localization results in high β-galactosidase activity and low alkaline phosphatase activity. The HEK293T cell line was used for transfection experiments and as a host to study membrane protein topology. We also used the pCAGGS plasmid carrying the c-myc tag, which can be detected by chemiluminescence western blotting.

TABLE 1 Cell lines and plasmids used in this study.

Strain	Genotype	Source or reference
DH5α	F^- *gyrA96 endA1 relA1 thi-1 hsdR17 glnV44(AS) deoR Δ(argF-lac) U169 recA1 Φ80Δ(lacZ)M15*	Laboratory collection
HEK293T	Human embryonic kidney 293 cells and SV40 T-antigen	Laboratory collection
Plasmid	**Relevant feature(s)**	**Source or reference**
pKTop	Vector expressing dual reporter $PhoA_{22-472}$/$LacZ_{4-60}$, p15 ori; Km^r	Laboratory collection
pCAGGS	Vector expressing, SV40 promoter, Amp^r	Laboratory collection
pSVLd	Vector expressing, SV40 promoter, Amp^r	Laboratory collection

TABLE 2 Oligonucleotides used to clone truncated nsp4 SARS-CoV-2 variants in pKTop.

Oligonucleotide	Sequence (5′ → 3′)
FORWARD	CGGTCTAGAGATGAAGATTGTAAACAAT
REVERSE	
E141 (pKTop-NSP4$_{1\text{-}141}$)	CGGGAGCTCGGTTCAATCAGCTTGCTTGG
G164 (pKTop-NSP4$_{1\text{-}164}$)	CGGGAGCTCGGTCCGCTCGCATCCTTAAA
D279 (pKTop-NSP4$_{1\text{-}279}$)	CGGGAGCTCGGATCCAGTGCGCCAATGGG
H313 (pKTop-NSP4$_{1\text{-}313}$)	CGGGAGCTCGGGTGGGAATATTCCCCGAA
Y335 (pKTop-NSP4$_{1\text{-}335}$)	CGGGAGCTCGGGTAGACAGGAGTCAAACA
Y346 (pKTop-NSP4$_{1\text{-}346}$)	CGGGAGCTCGGGTAAATAACTGAGTATAC
H362 (pKTop-NSP4$_{1\text{-}362}$)	CGGGAGCTCGGGTAAATAACTGAGTATAC
P371 (pKTop-NSP4$_{1\text{-}371}$)	CGGGAGCTCGGTGGCGTAAACATCACCAT
Y397 (pKTop-NSP4$_{1\text{-}397}$)	CGGGAGCTCGGGTAGTTACTAAAGAACCA
Q 500 (pKTop-NSP4$_{1\text{-}500}$)	CGGGAGCTCGGCTGGAGTACGGCTGAGGT

4.3 | Computer-assisted analysis of nsp4 sequence

TM segments were predicted using up to seven of the most common methods available on the Internet: TOPCONS (Tsirigos et al. 2015) (http://topcons.net/), ΔG Predictor (Hessa et al. 2007) (https://dgpred.cbr.su.se/index.php?p=home), deepTMHMM (Hallgren et al. 2022; Jumper et al. 2021) (https://services.healthtech.dtu.dk/services/DeepTMHMM-1.0/), HMMTop (Tusnády and Simon 1998; Tusnády and Simon 2001) (http://www.enzim.hu/hmmtop), Protter (Omasits et al. 2014) (http://wlab.ethz.ch/protter), MEMSAT (Nugent and Jones 2009) (http://bioinf.cs.ucl.ac.uk/psipred/), and SPOCTOPUS (Viklund et al. 2008) (http://octopus.cbr.su.se/). All user-adjustable parameters were left at their default values.

4.4 | Bacterial membrane topology assays

Selected DNA fragments were amplified by PCR using NZYTaq II polymerase (NZYTECH) and the oligonucleotides shown in Table 2. The amplified DNA fragments were digested with *Xba*I and *Sac*I, and then ligated with T4 DNA ligase (NZYTECH) at 16°C into the pKTop plasmid, which had been previously digested with the same enzymes. Thus, the dual pho-lac reporter was fused in-frame after selected DNA fragments encoding the following 11 truncated proteins: 89-mer (pKTop-NSP4$_{1–89}$), 141-mer (pKTop-NSP4$_{1–141}$), 164-mer (pKTop-NSP4$_{1–164}$), 279-mer (pKTop-NSP4$_{1–279}$), 313-mer (pKTop-NSP4$_{1–313}$), 335-mer (pKTop-NSP4$_{1–335}$), 346-mer (pKTop-NSP4$_{1–346}$), 362-mer (pKTop-NSP4$_{1–362}$), 371-mer (pKTop-NSP4$_{1–371}$), 397-mer (pKTop-NSP4$_{1–397}$), and 500-mer (pKTop-NSP4$_{1–500}$).

To study the topology of the nsp4 protein in vivo, we used a methodology based on a dual reporter: the pKTop plasmid encoding alkaline phosphatase, and α peptide from β galactosidase (Karimova and Ladant 2017). In the pKTop plasmid, we cloned DNA sequences encoding 11 protein truncates (89-mer, 141-mer, 164-mer, 279-mer, 313-mer, 335-mer, 346-mer, 362-mer, 371-mer, 397-mer, and 500-mer) or the full-length protein. Each construct was transformed into competent *E. coli* DH5α cells, which were grown at 37°C on agar plates with kanamycin 50 μg/mL.

To qualitatively analyze the topology, a freshly transformed colony was selected, and the cells were grown to an OD_{600nm} of 0.5 in Luria-Bertani (LB) broth (0.5% yeast extract, 1% tryptone, and 0.5% NaCl). Next, 7 μL was spotted onto plates containing LB broth agar, supplemented with 80 μg/mL disodium 5-bromo-4-chloro-3-indolyl phosphate (BCIP), 100 μg/mL 6-chloro-3-indolyl-β-D-galactoside (Red-Gal), 1 mM IPTG, and 50 μg/mL kanamycin. Plates were developed overnight at 37°C.

For quantitative analysis, the cells were grown to an OD_{600nm} of 0.5 in LB with 50 μg/mL kanamycin and were induced with 1 mM IPTG for 1 h with aeration at 37°C to induce truncates/PhoA-LacZα protein expression. To study β-galactosidase activity, the cells were harvested by centrifugation (5 min at 7000 rpm) and resuspended in M63 medium [100 mM KH_2PO_4, 15 mM $(NH_4)_2SO_4$, 1.7 mM Fe_2SO_4, and 1 mM $MgSO_4$, pH 7.0]. Next, 50 μL of the bacterial suspension was permeabilized with SDS/chloroform and then mixed with 100 μL of PM2 buffer and ONPG (0.15%) to initiate the enzymatic reaction. To study phosphatase activity, the cells were harvested by centrifugation (5 min at 7000 rpm) and resuspended in cold PM1 buffer (1M Tris–HCl, pH 8.0, 0.1 mM $ZnCl_2$, and 1 mM iodoacetamide). Next, 100 μL of the bacterial suspension was permeabilized with SDS/chloroform and then mixed with 50 μL of pNPP solution (0.15% in 1M Tris–HCl, pH 8.0) to initiate the enzymatic reaction.

The β-galactosidase and phosphatase reactions were stopped by adding 1M Na_2CO_3 and 2 N NaOH, respectively, and the enzymatic activities in relative units and the NAR were calculated as previously described (Karimova and Ladant 2017).

4.5 | C_LTMnsp4 construct design

Chimeric constructs were generated using a previously described λ light-chain C_L domain construct (Feige and Hendershot 2013). In this construct, a new glycosylation site (Q109N) was designed by site-directed mutagenesis (QuikChange), fused to the tested nsp4 sequence (TM1, residues 10–32), flanked by insulating GGPG- and -GPGG tetrapeptides (Duart et al. 2022) plus flexible linkers, and followed by an additional C-terminal glycosylation site.

4.6 | Nsp4 expression in mammalian cells

Nsp4 protein sequence variants and C_LTM constructs were tagged with a *c*-myc epitope (EQKLISEEDL) at their C-terminus, and then inserted into a pCAGGS-ampicillin plasmid. The *c*-myc epitope was followed by an optimized glycosylation site (NSTMGM) (Bañó-Polo et al. 2011; Tamborero et al. 2011) or a non-glycosylable (QSTMGM) sequence, as appropriate (Figure 5). After sequence verification, plasmids were transfected into HEK293-T cells using Lipofectamine 2000 (Life Technologies), following the manufacturer's protocol. Approximately 24 h post-transfection, the cells were harvested and washed with PBS buffer. After centrifugation (1000 rpm for 5 min), the cells were lysed by addition of 100 μL of lysis buffer (30 mM Tris–HCl, 150 mM NaCl, and 0.5% Nonidet P-40), and were sonicated for 10 min using a Bioruptor (Diagenode), in an ice bath, followed by another centrifugation. Total protein was quantified, and equal amounts of protein were subjected to Endo H treatment or mock treatment, followed by SDS-PAGE analysis and transfer onto a PVDF transfer membrane (ThermoFisher Scientific). Protein glycosylation status was analyzed by western blotting, using an anti-c-myc antibody (Sigma), anti-rabbit IgG-peroxidase conjugated antibody (Sigma), and ECL developing reagent (GE Healthcare). Chemiluminescence was visualized using an ImageQuantTM LAS 4000 mini Biomolecular Imager (GE Healthcare).

4.7 | Atomistic molecular dynamics simulations

We performed atomistic molecular dynamics (MD) simulations using the CHARMM36m force field for lipids and proteins, the CHARMM TIP3P model for water, and the standard CHARMM36 force field for ions (Huang et al. 2016). Simulations were conducted using the GROMACS 2024 simulation package (Abraham et al. 2015). The initial nsp4 structure was obtained from an AlphaFold2 (Jumper et al. 2021) prediction, using ColabFold v1.5.2 with default parameters (Mirdita et al. 2022).

For the full-length protein, the N- and C-termini were modeled as charged residues. On the other hand, the N-terminus was kept neutral for the signal peptide-cleaved. Histidine (HIS) residues were modeled as neutral. Glutamate (GLU), aspartate (ASP), arginine (ARG), and lysine (LYS) residues were modeled as charged, with the exception of D356, which was protonated (neutral) due to its position within the hydrophobic membrane core. Instead, the signal peptide-like cleaved nsp4 variant (residues 1–32 removed) was set to a neutral charge state, to minimize potential artifacts from unspecific electrostatic interactions. The CHARMM-GUI web server (Lee et al. 2016) was used to integrate the models into a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) model membrane with or without 30 mol% cholesterol (Figure S2c), with the positioning of the protein guided by PPM 2.0 server (Lomize et al. 2012) predictions. Each condition was generated three times to ensure independent initial configurations and run for at least 1100 ns. The PPM 2.0 server (Lomize et al. 2012) yields two distinct TM predictions depending on whether a signal sequence is included, which alters the spatial orientation of the N-terminal domain relative to the plasma membrane (Figure S2a,b).

To mimic experimental conditions, 150 mM potassium chloride was added. The system underwent energy minimization, followed by multiple equilibration steps using scripts provided by CHARMM-GUI, with gradual removal of restraints. Temperature was maintained at 310 K, using a V-rescale (Bussi et al. 2007) thermostat with a time constant of 1.0 ps. Pressure was kept constant at 1 atm, using a semi-isotropic pressure-coupling scheme with a C-rescale barostat (Bernetti and Bussi 2020) (time constant: 5.0 ps; isothermal compressibility: 4.5×10^{-5} bar^{-1}). Neighbor searching was performed using the Verlet algorithm (Verlet 1967), with updates every 20 steps. Electrostatic interactions were calculated using the Particle Mesh Ewald method (Darden et al. 1993), with periodic boundary conditions applied in all directions. Simulations were performed with a 2-fs integration time step.

All structures predicted using AlphaFold2, along with the initial and configuration files for the MD simulations associated with this study, have been deposited on ZENODO. They can be accessed at 10.5281/zenodo.16649058.

AUTHOR CONTRIBUTIONS

José M. Acosta-Cáceres: Investigation; writing – review and editing; methodology; data curation; formal analysis.

Fabio Lolicato: Investigation; methodology; software; writing – review and editing; formal analysis; data curation; visualization; resources; funding acquisition. **Laura Gadea-Salom:** Investigation; data curation. **Juan Ortiz-Mateu:** Investigation; resources; visualization. **Diego Belda:** Investigation; data curation. **Luis Martínez-Gil:** Formal analysis; supervision. **María Jesús García-Murria:** Supervision; data curation; formal analysis; validation. **Ismael Mingarro:** Conceptualization; funding acquisition; writing – original draft; writing – review and editing; project administration; formal analysis; supervision; validation; visualization; data curation.

ACKNOWLEDGMENTS

We thank Prof. Nevan J. Krogan (UCSF) for sharing the SARS-CoV-2 nsp4 DNA fragment. We thank Novartis for kindly providing cavinafungin under a material transfer agreement. This work was supported by grants PID2020-119111GB-I00 and PID2023-152568NB-I00 from the Spanish Ministry of Science and Innovation (MCIN/AEI/10.13039/501100011033), and CIPROM/2022/062 from the Generalitat Valenciana to I.M.; and by the Deutsche Forschungsgemeinschaft SFB-1638/1-511488495-Z01 to F.L. J.A.-C. was the recipient of a predoctoral fellowship from the Ministerio de Educación (Perú). The authors gratefully acknowledge the data storage service SDS@hd supported by the Ministry of Science, Research, and the Arts Baden-Württemberg (MWK), the German Research Foundation (DFG) through grant INST 35/1314-1 FUGG and INST 35/1503-1 FUGG. We acknowledge the computing resources provided by the CSC-IT Center for Science Ltd. (Espoo, Finland), and support from the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant INST 35/1597-1 FUGG.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Ismael Mingarro https://orcid.org/0000-0002-1910-1229

REFERENCES

Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1–2:19–25.

Alexeyev MF, Winkler HH. Membrane topology of the *Rickettsia prowazekii* ATP/ADP translocase revealed by novel dual pho-lac reporters. J Mol Biol. 1999;285:1503–13.

Angelini MM, Akhlaghpour M, Neuman BW, Buchmeier MJ. Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. MBio. 2013;4:e00524-13.

Bañó-Polo M, Baeza-Delgado C, Tamborero S, Hazel A, Grau B, Nilsson I, et al. Transmembrane but not soluble helices fold inside the ribosome tunnel. Nat Commun. 2018;9:5246.

Bañó-Polo M, Baldin F, Tamborero S, Marti-Renom MA, Mingarro I. N-glycosylation efficiency is determined by the distance to the C-terminus and the amino acid preceding an Asn-Ser-Thr sequon. Protein Sci. 2011;20:179–86.

Bar-On YM, Flamholz A, Phillips R, Milo R. SARS-CoV-2 (COVID-19) by the numbers. eLife. 2020;9:1715–87.

Beachboard DC, Anderson-Daniels JM, Denison MR. Mutations across murine hepatitis virus nsp4 alter virus fitness and membrane modifications. J Virol. 2015;89:2080–9.

Bernetti M, Bussi G. Pressure control using stochastic cell rescaling. J Chem Phys. 2020;153:114107.

Brant AC, Tian W, Majerciak V, Yang W, Zheng ZM. SARS-CoV-2: from its discovery to genome structure, transcription, and replication. Cell Biosci. 2021;11:136.

Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys. 2007;126:014101.

Chen A, Lupan A-M, Tong Quek R, Stanciu SG, Asaftei M, Stanciu GA, et al. A coronaviral pore-replicase complex links RNA synthesis and export from double-membrane vesicles. Sci Adv. 2024;10:eadq9580.

Clementz MA, Kanjanahaluethai A, O'Brien TE, Baker SC. Mutation in murine coronavirus replication protein nsp4 alters assembly of double membrane vesicles. Virology. 2008;375:118–29.

Darden T, York D, Pedersen L. Particle mesh Ewald: an N·log (N) method for Ewald sums in large systems. J Chem Phys. 1993;98:10089–92.

Duart G, Garcia-Murria MJ, Mingarro I. The SARS-CoV-2 envelope (E) protein has evolved towards membrane topology robustness. Biochim Biophys Acta. 2021;1863:183608. https://doi.org/10.1016/j.bbamem.2021.183608

Duart G, Graña-Montes R, Pastor-Cantizano N, Mingarro I. Experimental and computational approaches for membrane protein insertion and topology determination. Methods. 2024;226:102–19.

Duart G, Lamb J, Ortiz-Mateu J, Elofsson A, Mingarro I. Intra-helical salt bridge contribution to membrane protein insertion. J Mol Biol. 2022;434:167467.

Estoppey D, Lee CM, Janoschke M, Lee BH, Wan KF, Dong H, et al. The natural product cavinafungin selectively interferes with zika and dengue virus replication by inhibition of the host signal peptidase. Cell Rep. 2017;19:451–60.

Feige MJ, Hendershot LM. Quality control of integral membrane proteins by assembly-dependent membrane integration. Mol Cell. 2013;51:297–309.

Finkel Y, Mizrahi O, Nachshon A, Weingarten-Gabbay S, Morgenstern D, Yahalom-Ronen Y, et al. The coding capacity of SARS-CoV-2. Nature. 2021;589:125–30. https://doi.org/10.1038/s41586-020-2739-1

Gadlage MJ, Sparks JS, Beachboard DC, Cox RG, Doyle JD, Stobart CC, et al. Murine hepatitis virus nonstructural protein 4 regulates virus-induced membrane modifications and replication complex function. J Virol. 2010;84:280–90.

Grau B, Kormos R, Bañó-Polo M, Chen K, García-Murria MJ, Hajredini F, et al. Sequence-dependent scale for translocon-mediated insertion of interfacial helices in membranes. Sci Adv. 2025;11:eads6804.

Hagemeijer MC, Monastyrska I, Griffith J, van der Sluijs P, Voortman J, van Bergen en Henegouwen PM, et al. Membrane rearrangements mediated by coronavirus nonstructural proteins 3 and 4. Virology. 2014;458–459:125–35.

Hallgren J, Tsirigos KD, Damgaard Pedersen M, Juan J, Armenteros A, Marcatili P, et al. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. bioRxiv. 2022. https://doi.org/10.1101/2022.04.08.487609

Hegde RS, Keenan RJ. A unifying model for membrane protein biogenesis. Nat Struct Mol Biol. 2024;31:1009–17.
Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, Andersson H, et al. Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature. 2005;433:377–81.
Hessa T, Meindl-Beinker NM, Bernsel A, Kim H, Sato Y, Lerch-Bader M, et al. Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature. 2007;450:1026–30.
Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, De Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods. 2016;14:71–3.
Huang Y, Wang T, Zhong L, Zhang W, Zhang Y, Yu X, et al. Molecular architecture of coronavirus double-membrane vesicle pore complex. Nature. 2024;633:224–31.
Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583–9. https://doi.org/10.1038/s41586-021-03819-2
Karimova G, Ladant D. Defining membrane protein topology using pho-lac reporter fusions. Methods Mol Biol. 2017;1615:129–42.
Karimova G, Robichon C, Ladant D. Characterization of YmgF, a 72-residue inner membrane protein that associates with the Escherichia coli cell division machinery. J Bacteriol. 2009;191:333–46.
Klatte N, Shields DC, Agoni C. Modelling the transitioning of SARS-CoV-2 nsp3 and nsp4 lumenal regions towards a more stable state on complex formation. Int J Mol Sci. 2022;24:720. https://doi.org/10.3390/ijms24010720
Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, et al. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. J Chem Theory Comput. 2016;12:405–13.
Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Res. 2012;40:D370–6.
Malone B, Urakova N, Snijder EJ, Campbell EA. Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. Nat Rev Mol Cell Biol. 2022;23:21–33.
Martinez-Gil L, Bañó-Polo M, Redondo N, Sánchez-Martinez S, Nieva JL, Carrasco L, et al. Membrane integration of poliovirus 2B viroporin. J Virol. 2011;85:11315–24.
Martinez-Gil L, Johnson AE, Mingarro I. Membrane insertion and biogenesis of the Turnip Crinkle Virus p9 movement protein. J Virol. 2010;84:5520–7.
Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nat Methods. 2022;19:679–82.
Mohan J, Wollert T. Membrane remodeling by SARS-CoV-2—double-enveloped viral replication. Fac Rev. 2021;10:17.
Nilsson I, von Heijne G. Fine-tuning the topology of a polytopic membrane protein: role of positively and negatively charged amino acids. Cell. 1990;62:1135–41.
Nugent T, Jones DT. Transmembrane protein topology prediction using support vector machines. BMC Bioinf. 2009;10:911–80.
Omasits U, Ahrens CH, Müller S, Wollscheid B. Protter: interactive protein feature visualization and integration with experimental proteomic data. Bioinformatics. 2014;30:884–6.
Oostra M, te Lintelo EG, Deijs M, Verheije MH, Rottier PJM, de Haan CAM. Localization and membrane topology of coronavirus nonstructural protein 4: involvement of the early secretory pathway in replication. J Virol. 2007;81:12323–36.
Oudshoorn D, Rijs K, Limpens RWAL, Groen K, Koster AJ, Snijder EJ, et al. Expression and cleavage of middle east respiratory syndrome coronavirus nsp3-4 polyprotein induce the formation of double-membrane vesicles that mimic those associated with coronaviral RNA replication. MBio. 2017;8:e01658-17.
Sakai Y, Kawachi K, Terada Y, Omori H, Matsuura Y, Kamitani W. Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication. Virology. 2017;510:165–74.
Snapp EL, McCaul N, Quandte M, Cabartova Z, Bontjer I, Källgren C, et al. Structure and topology around the cleavage site regulate post-translational cleavage of the HIV-1 gp160 signal peptide. eLife. 2017;6:e26067.
Tabata K, Prasad V, Paul D, Lee JY, Pham MT, Twu WI, et al. Convergent use of phosphatidic acid for hepatitis C virus and SARS-CoV-2 replication organelle formation. Nat Commun. 2021;12:7276.
Tamborero S, Vilar M, Martínez-Gil L, Johnson AE, Mingarro I. Membrane insertion and topology of the translocating chain-associating membrane protein (TRAM). J Mol Biol. 2011;406:571–82.
Tsirigos KD, Peters C, Shu N, Käll L, Elofsson A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. Nucleic Acids Res. 2015;43:W401–7.
Tusnády GE, Simon I. Principles governing amino acid composition of integral membrane proteins: application to topology prediction. J Mol Biol. 1998;283:489–506.
Tusnády GE, Simon I. The HMMTOP transmembrane topology prediction server. Bioinformatics. 2001;17:849–50.
Verlet L. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. Phys Rev. 1967;159:98–103.
Viklund H, Bernsel A, Skwark M, Elofsson A. SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. Bioinformatics. 2008;24:2928–9.
V'kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus biology and replication: implications for SARS-CoV-2. Nat Rev Microbiol. 2021;19:155–70.
Whitley P, Grau B, Gumbart JC, Martínez-Gil L, Mingarro I. Folding and insertion of transmembrane helices at the ER. Int J Mol Sci. 2021;22:12778.
Whitley P, Mingarro I. Stitching proteins into membranes, not sew simple. Biol Chem. 2014;395:1417–24.
Wolff G, Limpens RWAL, Zevenhoven-Dobbe JC, Laugks U, Zheng S, de Jong AWM, et al. A molecular pore spans the double membrane of the coronavirus replication organelle. Science. 2020b;369:1395–8.
Wolff G, Melia CE, Snijder EJ, Bárcena M. Double-membrane vesicles as platforms for viral replication. Trends Microbiol. 2020a;28:1022–33.
Zanotti A, Coelho JPL, Kaylani D, Singh G, Tauber M, Hitzenberger M, et al. The human signal peptidase complex acts as a quality control enzyme for membrane proteins. Science. 2022;378:996–1000.
Zhang R, Miner JJ, Gorman MJ, Rausch K, Ramage H, White JP, et al. A CRISPR screen defines a signal peptide processing pathway required by flaviviruses. Nature. 2016;535:164–8.
Zimmermann L, Zhao X, Makroczyova J, Wachsmuth-Melm M, Prasad V, Hensel Z, et al. SARS-CoV-2 nsp3 and nsp4 are minimal constituents of a pore spanning replication organelle. Nat Commun. 2023;14:7894. https://doi.org/10.1038/s41467-023-43666-5

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Acosta-Cáceres JM, Lolicato F, Gadea-Salom L, Ortiz-Mateu J, Belda D, Martínez-Gil L, et al. Structural insights into SARS-CoV-2 nonstructural protein 4 (nsp4) biogenesis. Protein Science. 2025;34(11):e70355. https://doi.org/10.1002/pro.70355