Oligomeric assembly of the C-terminal and transmembrane region of SARS-CoV-2 nsp3

Marion Babot, Yves Boulard, Samira Agouda, Laura Pieri, Sonia Fieulaine, Stéphane Bressanelli, Virginie Gervais

▶ **To cite this version:**

Marion Babot, Yves Boulard, Samira Agouda, Laura Pieri, Sonia Fieulaine, et al.. Oligomeric assembly of the C-terminal and transmembrane region of SARS-CoV-2 nsp3. Journal of Virology, 2024, 98 (4), pp.e01575-23. 10.1128/jvi.01575-23 . hal-04731906

HAL Id: hal-04731906
https://hal.science/hal-04731906v1

Submitted on 11 Oct 2024

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

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

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology

Computational Biology | Full-Length Text

Oligomeric assembly of the C-terminal and transmembrane region of SARS-CoV-2 nsp3

Marion Babot,[1] Yves Boulard,[1] Samira Agouda,[1] Laura Pieri,[1] Sonia Fieulaine,[1] Stéphane Bressanelli,[1] Virginie Gervais[1]

AUTHOR AFFILIATION See affiliation list on p. 21.

ABSTRACT As for all single-stranded, positive-sense RNA (+RNA) viruses, intracellular RNA synthesis relies on extensive remodeling of host cell membranes that leads to the formation of specialized structures. In the case of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) coronavirus causing COVID-19, endoplasmic reticulum membranes are modified, resulting in the formation of double-membrane vesicles (DMVs), which contain the viral dsRNA intermediate and constitute membrane-bound replication organelles. The non-structural and transmembrane protein nsp3 is a key player in the biogenesis of DMVs and, therefore, represents an interesting antiviral target. However, as an integral transmembrane protein, it is challenging to express for structural biology. The C-terminus of nsp3 encompasses all the membrane-spanning, -interacting, and -remodeling elements. By using a cell-free expression system, we successfully produced the C-terminal region of nsp3 (nsp3C) and reconstituted purified nsp3C into phospholipid nanodiscs, opening the way for structural studies. Negative-stain transmission electron microscopy revealed the presence of nsp3C oligomers very similar to the region abutting and spanning the membrane on the cytosolic side of DMVs in a recent subtomogram average of the SARS-CoV-2 nsp3-4 pore (1). AlphaFold-predicted structural models fit particularly well with our experimental data and support a pore-forming hexameric assembly. Altogether, our data give unprecedented clues to understand the structural organization of nsp3, the principal component that shapes the molecular pore that spans the DMVs and is required for the export of RNA *in vivo*.

IMPORTANCE Membrane remodeling is at the heart of intracellular replication for single-stranded, positive-sense RNA viruses. In the case of coronaviruses, including severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), this leads to the formation of a network of double-membrane vesicles (DMVs). Targeting DMV biogenesis offers promising prospects for antiviral therapies. This requires a better understanding of the molecular mechanisms and proteins involved. Three non-structural proteins (nsp3, nsp4, and nsp6) direct the intracellular membrane rearrangements upon SARS-CoV-2 infection. All of them contain transmembrane helices. The nsp3 component, the largest and multi-functional protein of the virus, plays an essential role in this process. Aiming to understand its structural organization, we used a cell-free protein synthesis assay to produce and reconstitute the C-terminal part of nsp3 (nsp3C) including transmembrane domains into phospholipid nanodiscs. Our work reveals the oligomeric organization of one key player in the biogenesis of SARS-CoV-2 DMVs, providing basis for the design of future antiviral strategies.

KEYWORDS SARS-CoV-2, cell-free expression, nanodiscs, electron microscopy, nsp3, coronavirus, membrane remodeling

SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) is the coronavirus strain discovered in early December 2019 in the city of Wuhan in China (1), which

Editor Kanta Subbarao, The Peter Doherty Institute for Infection and Immunity, Melbourne, Australia

Address correspondence to Marion Babot, marion.babot@universite-paris-saclay.fr, or Virginie Gervais, virginie.gervais@i2bc.paris-saclay.fr.

The authors declare no conflict of interest.

Received 16 October 2023
Accepted 22 February 2024
Published 14 March 2024

Copyright © 2024 American Society for Microbiology. All Rights Reserved.

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

has since plunged the world into a fearsome pandemic with over 6.9 million deaths worldwide (January 2024). The deployment of multiple COVID-19 vaccines in record time has resulted in the administration of more than 13 billion COVID-19 vaccine doses worldwide, providing effective protection against severe disease and death, which resulted in a net reduction in hospital and intensive care unit admissions. However, the vaccines currently used are predominantly targeting the spike (S) protein of SARS-CoV-2, the key player in the receptor recognition and in cell membrane fusion process. While widespread vaccination coverage offers the hope of ending the pandemic, the frequency of mutations in S glycoprotein remains a concern and may alter the efficacy of current vaccines (2). Similarly, drugs targeting the SARS-CoV-2 main protease and polymerase have made it to the market in record times but still have limited use to cure SARS-CoV-2 infection and/or limit its spread (3). Thus, in the long-term fight against novel emerging variants of SARS-CoV-2, the need to extend vaccination and drug design to other targets remains, which implies further studies to determine functions and structures of proteins that are essential for the viral life cycle.

Like all coronaviruses, SARS-CoV-2 possesses a remarkably large RNA genome and displays a highly conserved genomic organization (4). It is characterized by a 30 kb single-stranded, non-segmented, polyadenylated, positive-sense RNA genome. In the 5′-region, two large ORFs (ORF1a and ORF1b) make up about two-thirds of the viral genome, encoding the pp1a and pp1ab polyproteins. Two protease activities encoded by the virus (papain-like proteinase, nsp3 and 3-chymotrypsine-like protease, nsp5) allow the cleavage of the two polyproteins into 16 mature non-structural proteins (nsp1 to nsp16), Fig. 1A. The last third of the genome encodes the ORFs which release the four structural proteins that include spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins as well as several virus-specific accessory proteins, presumably involved in the modulation of the host's response to viral infection (5).

The non-structural proteins have a wide range of biological functions and altogether, they interact with cellular factors to form the viral replication-transcription complex (RTC) (8). Three non-structural and transmembrane proteins (nsp3, nsp4, and nsp6) share membrane anchoring properties (9, 10). Replication of single-stranded, positive-sense RNA (+RNA) viruses takes place in the cytoplasm in specialized structures called replication organelles (ROs). Infection by +RNA viruses leads to the remodeling of host cell endomembranes to produce a diverse array of specialized structures associated with RNA synthesis (11). These rearranged structures emerge early in the infection process, providing an optimal and dynamic micro-environment where the entire replication machinery concentrates and is physically protected from the host's innate immune system (12). Within the Nidovirales order, arteriviruses and coronaviruses constitute a subset of +RNA viruses known for inducing double-membrane vesicles (DMVs) (13, 14). In well-characterized betacoronaviruses such as mouse hepatitis virus (MHV), severe acute respiratory syndrome coronavirus (SARS-CoV), middle east respiratory syndrome coronavirus (MERS-CoV), and now SARS-CoV-2, the double-membrane vesicles are coming from endoplasmic reticulum (ER). DMVs act as central hubs for viral RNA synthesis (15). Many studies have reported an essential role of nsp3 and nsp4 in the membrane remodeling process, particularly in the DMVs genesis of coronaviruses (14, 16–18) and arteriviruses (where the counterparts of coronavirus nsp3 and nsp4 are nsp2 and nsp3, respectively) (19).

With its 1945 amino acids, nsp3 is the largest protein encoded by the SARS-CoV-2 genome (>200 kDa) and plays important roles in the virus life cycle, starting from the early stage of infection. Being a multi-domain protein with RNA-binding activity and a papain-like protease activity (releasing itself together with nsp1 and nsp2), the nsp3 protein is a molecular hub for replication and nucleocapsid assembly (20). It interacts with itself and with several different viral and host proteins (21, 22). In particular, a virus-specific interaction between nsp3 and nsp4 is critical for membrane rearrangement and DMV formation for viral genome replication, as described for MHV and SARS-CoV (16, 17). Harboring transmembrane segments in its C-terminus, nsp3 spans the membrane (Fig.

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

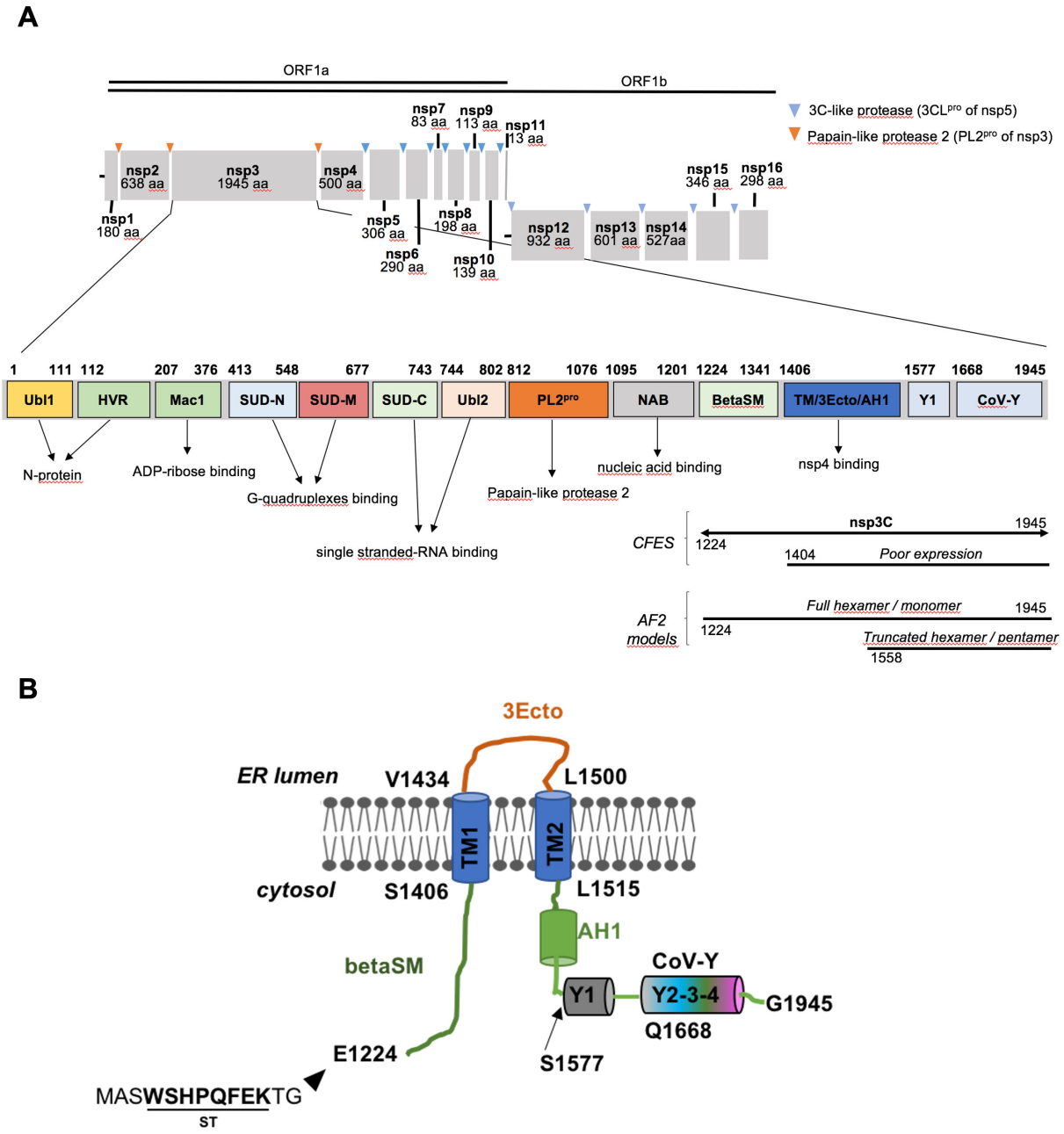

FIG 1 Organization of SARS-CoV-2 non-structural nsp3 protein. (A) Two open reading frames (ORF1a and ORF1b), constituting two-thirds of the SARS-CoV-2 genome, encode two viral polyproteins, pp1a and pp1ab. Production of the longest (pp1ab) or shortest protein (pp1a) depends on a −1 ribosomal frameshift event. Pp1a and pp1ab are the precursors of 16 so-called non-structural proteins (nsps), which are produced by self-cleavage of the two precursors. The cleavage by papain-like protease 2 ($PL2^{pro}$) activity of nsp3 releases nsp1, nsp2, and nsp3 proteins, while the other nsps are released by chymotrypsine-like cysteine-dependent protease ($3CL^{pro}$) of nsp5. Bottom is shown the domain organization of nsp3 in SARS-CoV-2 as well as associated functions. The nsp3C construct (E1224-G1945) used in the present work contains a strep-tag in its N-terminus part and includes the betaSM (betacoronavirus specific marker) superfamily domain and the transmembrane region and the CoV-Y domain. *Abbreviations*: Ubl1, ubiquitin-like domain 1; HVR, hyper variable region; Mac1, Macrodomain 1; SUD, SARS-unique domain; Ubl2, ubiquitin-like domain 2; $PL2^{pro}$, Papain-like protease 2; NAB, nucleic acid-binding domain; BetaSM, betacoronavirus-specific-marker; TM/3Ecto/AH1, transmembrane region; Y1 domain, domain with unknown function; CoV-Y, coronavirus specific unknown function. (B) Topological organization of nsp3C. The Strep-tag (ST) is located at the N-terminus of the nsp3C construct (E1224-G1945). Both N- and C-terminal point to the cytosol. The N-terminus starts with the betaSM region (green) followed by the TM/3Ecto/AH1 region. The two TM helices (TM1 and TM2, blue cylinders) are separated by a luminal loop (3Ecto, orange). The colors are based on previous work reporting the mapping of transmembrane domains of SARS-CoV-2 nsps (6). Next to TM2, a helical region (AH1) precedes the C-terminus comprising the Y domains, as reported by reference (7).

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

1B). Co-expression of nsp4 and a C-terminal part of nsp3 including the transmembrane region affects their sub-cellular localization and leads to the appearance of perinuclear foci in which both of them concentrate (23). On the other hand, the deletion of the transmembrane segments of nsp3 prevents the formation of DMVs (16).

The membrane topology of nsp3, previously reported in SARS-CoV and MHV, showed that out of three predicted hydrophobic helices, only the first two would cross the lipid bilayer (10). The so-called TM1 and TM2 helices are linked by a loop (ectodomain or 3Ecto) that extends into the luminal space of the ER, while both N- and C-terminal parts of nsp3 are exposed on the cytoplasmic face (Fig. 1B). Other studies conducted in MHV have concluded that the transmembrane region of nsp3 would correspond to four membrane-spanning domains (24). The luminal loop (3Ecto domain) is important for membrane rearrangement in SARS coronavirus (16). Interaction of nsp3 and nsp4 *via* their luminal loops drives the formation of cellular membrane rearrangements onto which the RTC assembles (23). The presence of conserved cysteine residues in luminal loops of both proteins would allow intramolecular disulfide bonds to occur and would play an important role in bringing the opposing ER membranes together by mediating correct folding of the luminal loops and allowing appropriate nsp3-nsp4 interactions (16, 25). Next to TM2, a helical region (AH1) is found followed by a C-terminal cytosolic region which encompasses two Y domains, Y1 and CoV-Y (10), Fig. 1B. The crystal structure of the CoV-Y domain of SARS-CoV-2 was recently solved revealing a V-shaped fold with three distinct sub-domains Y2, Y3, and Y4 (7). Upstream of the TM region, there is the betaSM region for which the function is still unknown.

While the predominant involvement of nsp3 and nsp4 in inducing MHV and SARS-CoV membrane remodeling is now well established, it appears that other factors would also participate (26). The nsp6 protein is the third non-structural and transmembrane protein. It is a key determinant in autophagy functions and viral pathogenicity (27). From a recent study, nsp6 would act as an organizer of the DMVs induced by the nsp3-nsp4 pair, by forming (*via* a homo-dimerization process) a type of zipper joining the replicative vesicles with the ER (26). The simultaneous presence of nsp6 with nsp3 and nsp4 allows the formation of organized and mature DMVs, which look very similar to those observed in coronavirus-infected cells (14). The newly formed viral RNA is believed to be exported from the inside of the DMVs into the cytosol through molecular pores, which were revealed using *in situ* cryo-electron microscopy and tomography in MHV-infected cells (28). The very elegant work by Wolff et al. concluded that nsp3 is an essential component of the pore spanning the double-membrane of DMVs. A hexameric shape was observed with six prongs pointing from the membrane on the cytosolic side of DMVs and attributed to the N-terminal ends of six nsp3 molecules (28). This finding was confirmed and extended very recently by electron tomograms of the SARS-CoV-2 pore, showing that the expression of nsp3-nsp4 is sufficient for DMV and pore formation, that the pore itself, indeed, exhibits six-fold symmetry, and that the cytosolic prongs are formed by the N-terminus of nsp3 upstream of the TM1 helix (29).

While experimental structures are now available for most soluble domains of nsp3 [summarized in references (30) and (7)], the structural organization of its C-terminus including the transmembrane region remains to be elucidated. In this work, we have applied an *in vitro* approach using a wheat-germ cell-free expression system to produce the C-terminus part of nsp3 (residues E1224-G1945), starting at the betaSM region and including all membrane-spanning and membrane-interacting elements (hereafter nsp3C), Fig. 1A. With its high translation efficiency, the eukaryotic wheat germ cell-free method (hereafter, CFES for cell-free expression system) offers an interesting alternative to other commonly used cell-free protein expression methods (31). Furthermore, this system has proven to be successful for the production of transmembrane proteins that are typically difficult to express (32) and also for the production of viral polyproteins of +RNA viruses (33). While the CFES has been used to produce a variety of accessory and structural proteins of SARS-CoV-2 (34), it has not yet been reported for nsp3. In this work, we used the CFES to express nsp3C and successfully reconstituted detergent-solubilized

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

nsp3C into phospholipid nanodiscs for structural characterization, using negative-stain transmission electron microscopy (TEM).

We obtained a reconstructed 3D volume consistent with the expected location of nsp3C in the membrane on the cytosolic side of DMVs in the subtomogram average of the SARS-CoV-2 nsp3-4 pore (29). Predicted structural models using AlphaFold2 fit particularly well with our experimental data and further support hexameric assembly forming a pore. We performed molecular dynamic simulations of nsp3C that showed significant stability of the pore forming region, while the rest of the protein remained quite flexible. Our modeling suggests the presence of four putative TM helices and conformational changes that would be required to allow them to get closer together. We propose here the first structural view of the hexameric organization of the TM-3Ecto-AH1-Y1-CoV-Y region of the SARS-CoV-2 nsp3 protein.

RESULTS

Protein constructs and topology

The plasmid encoding the full-length and multi-domain nsp3 protein (1,945 residues, 220 kDa) appeared to be difficult to express using the CFES approach (data not shown). Based on previous work (see Introduction) and secondary structure predictions (not shown), we considered two C-terminal constructs encompassing all the necessary elements for DMV formation in cells (Fig. 1A). The genes encoding these constructs were first cloned into the pEU expression vector optimized to maximize synthetic efficiency and translation through CFES (see Materials and Methods). The shorter construct (residues N1404-G1945, nsp3C^{5}) without the whole betaSM domain was poorly expressed in the CFES (Fig. S1). The longer construct (E1224-G1945, hereafter nsp3C) gave higher expression levels compatible with biophysical studies (see below). It comprises the full-length betaSM domain, the TM/3Ecto/AH1 region, and the C-terminal Y domains (Fig. 1B). The betaSM superfamily is a specific marker of nsp3 in betacoronaviruses, including SARS-CoV and SARS-CoV-2, and is predicted to be non-enzymatic with poor structure propensity. Next to the betaSM domain, the TM/3Ecto/AH1 region includes two transmembrane helices separated by a luminal loop (3Ecto) and followed by a helical region (AH1). These domains are poorly conserved in coronaviruses compared to the C-terminal half of the protein that encompasses the Y1 and the CoV-Y domains (Fig. S2). The latter is made of three sub-domains Y2, Y3, and Y4 whose exact function in the virus replication cycle remains unclear (7).

Detergent screening for the cell-free expression and production of detergent-solubilized form of nsp3C

Our protocol relies on a two-step assay where the first reaction allows RNA transcription. The produced RNA is then used as a template for protein production (31). Detergent solubilization is the critical step in the purification of membrane proteins (35). The CFES is particularly well adapted for the use of detergents without damaging the wheat germ translational machinery (36). To easily and rapidly define which detergent is best suited for nsp3C solubilization during translation, we used a rapid microscale solubilization in 96 well plates (Fig. 2A). We analyzed expression levels in the presence of several detergents and the amount of nsp3C expressed in a soluble state. We considered a reasonable number of non-ionic detergents with relatively low CMC values and rather mild denaturation properties (Table 1). They proved to be good choices for successful expression yield and solubilization of membrane proteins in the CFES (37). Comparison between the bands obtained for the CFS, the pellet (P), and the supernatant (Sn) fractions with different detergents allowed us to choose the best-suited detergent for nsp3C stabilization. GDN was very effective in improving protein yield and maintaining the solubility of the protein, with no insoluble protein detected (Fig. 2B). Therefore, this steroid-based amphiphile detergent was kept for subsequent purification steps (see below). The addition of detergent after translation did not improve the yield of soluble protein recovered (Fig. 2B).

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

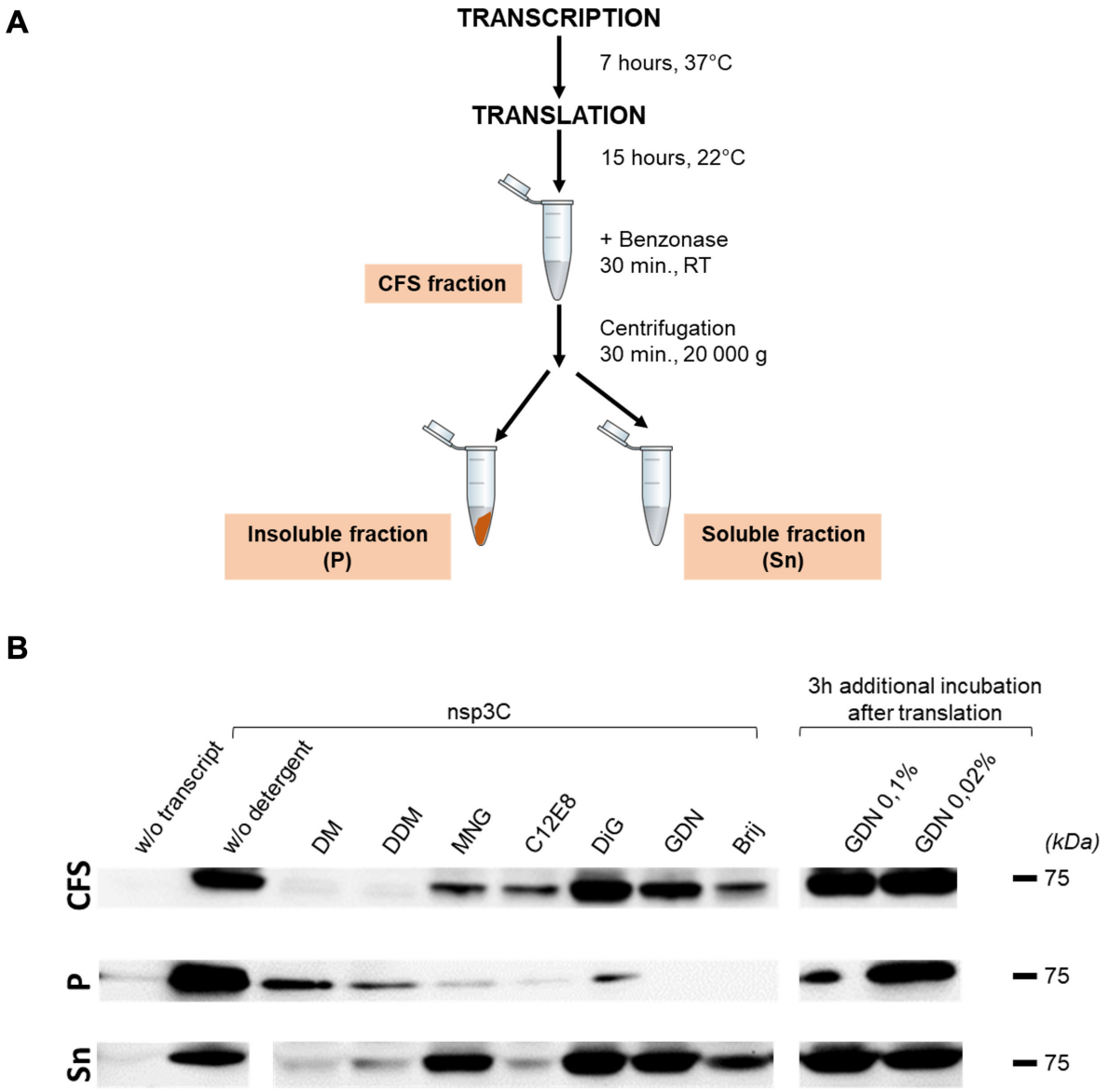

FIG 2 Expression of nsp3C using the wheat germ extract CFES and detergent screening. (A) After transcription of pEU-nsp3C gene and translation using the CFES, the translation reaction was stopped with benzonase prior to centrifugation to collect the soluble fraction (Sn) and the insoluble one (pellet, P). (B) Detergent screening. The amounts of recovered nsp3C in the different fractions (CFS, P, SN) in the absence or presence of detergent were estimated by western-blotting using anti-Strep-Tag antibodies. The translation reaction mixture was supplemented with the following detergents: DM, DDM, LMNG, C12E8, DiG (digitonin), GDN, and Brij (Brij58), see Table 1. The first two lanes correspond to negative control (w/o transcript) and nsp3C expression without detergent (w/o detergent). Following translation, the CFS extract with GDN was further incubated for 3 h at room temperature with additional GDN at final concentration of 0.02% or 0.1% to try to increase the amount of soluble nsp3C. For clarity, only the band corresponding to STnsp3C, whose expected size is 82 kDa, is shown at around 75 kDa and blots were cut as lanes were not always in the same order but the CFS, P and Sn fractions were analyzed on one membrane per fraction.

The C-terminal region of nsp3C encompassing the two transmembrane domains is oligomeric in solution

For structural studies, the production of nsp3C was adapted to large-scale expression in 6-well plates. The soluble fraction of nsp3C was incubated with Strep-Tactin resin, and elution was performed using 2.5 mM desthiobiotine in a small volume of elution buffer to avoid dilution of the protein (Fig. 3A and B). The E1 and E2 elution fractions were further purified using size-exclusion chromatography on a superdex 200 increase 10/300 GL column (Fig. 3C and D). The GDN detergent concentration was lowered

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

TABLE 1 Name and properties of the detergents used in the screen[a]

Detergents		MW (g mol^{-1})	CMC mM (wt%)
DM (38)	*n*-Decyl-β-D-maltoside	488	~1.8 (0.087%)
DDM	*n*-Dodecyl-β-D-maltopyranoside	511	~0.17 (0.0087%)
MNG-3 (LMNG) (39)	Maltose neopentyl glycol-3	1,005	~0.01 (0.001%)
C12E8	Octaethylene glycol monododecyl ether	539	~0.09 (0.0048%)
Digitonin	Steroid derivative	1,229	<0.5
GDN (40)	Glyco-diosgenin	473	~0.018 (0.002%)
Brij58	Polyethylene glycol hexadecyl ether	1,124	~0.08 (0.0048%)

[a]To allow comparison, the concentration used in the translation buffer was set at 0.1% for all detergents, which is higher than the recommended value (2 × CMC) for complete solubilization of the target membrane protein.

throughout the purification process (from 0.02% to 0.01%). The SEC profile obtained for nsp3C indicated the presence of an elution peak centered at 8.7 mL (Fig. 3C). From a SEC run using a mixture of water-soluble calibration proteins, such a volume suggests objects with molecular weight (MW) between 440 and 669 kDa and stokes radii (Rs) around 8 nm. Knowing that a globular protein with a MW of 82 kDa would have an approximate stoke radius of 4 nm (41), our SEC data indicate that nsp3C is unlikely monomeric in solution, even accounting for the associated GDN, but rather forms detergent-soluble oligomers. We then collected the SEC fractions (F13-14) corresponding to the peak and concentrated the protein pool about 10 times. A small aliquot was used for SDS-PAGE and Western Blot analysis (Fig. 3D). On SDS-PAGE, a band is detected around 75 kDa, close to the expected 82 kDa molecular weight of our nsp3C construct, keeping in mind that the migration pattern of membrane proteins does not often match the theoretical MW. Nsp3C was confirmed to be the major component of this band by protein identification using LC-MS/MS (data not shown). To further analyze the size and the oligomeric state of the protein, we performed a native polyacrylamide gel electrophoresis on a concentrated aliquot of nsp3C solubilized in GDN. Two protein standards with known molecular weights and diameters, ferritin (443 kDa/Rs 6.1 nm) and ovalbumin (43 kDa/Rs 2.8 nm), were used (Fig. 3E). For nsp3C sample, a band was detected on the top of the gel, slightly above that of soluble ferritin, reinforcing the idea that the protein would adopt an oligomeric conformation when solubilized in detergent.

To further analyze the sample quality, we next visualized the nsp3C sample using negative-stain TEM. In the concentrated sample, the final detergent concentration was estimated to be around 0.1%, in the range of the optimal concentration of detergent recommended for electron microscopy studies (42). TEM images revealed the presence of particles some 17 nm across, that form ring structures with an inner opening in their center and a width of ~5 nm (Fig. 3F). However, at this point, although no significant aggregation into larger particles was observed, we did not attempt 3D reconstruction due to sample heterogeneity.

In vitro assembly of nsp3C-nanodiscs

In order to increase the homogeneity of our samples and to stabilize the proteins, we then turned to the use of lipid nanodiscs that have become ideal systems for EM studies of membrane proteins (43). Lipid nanodiscs (NDs) offer a good alternative to liposomes that appear more often instable and difficult to prepare with defined size. Reconstitution into lipid NDs of membrane proteins transiently solubilized with detergent relies on the use of an amphipathic alpha-helical membrane scaffold protein (MSP) and additional detergent-solubilized phospholipids forming the lipid bilayer (44, 45). The general workflow we used is summarized in Fig. 4A. Following detergent removal with adsorbant bio-beads, the self-assembly of two MSP proteins and a fixed number of phospholipids results in the formation of NDs whose diameter is dictated by the length of the MSP variant used at the optimum lipid content. Indeed, the MSP needs to be large enough in order to accommodate both the membrane protein and sufficient phospholipids within the nanodisc (46). A key to success lies in the phospholipid:MSP ratio during

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

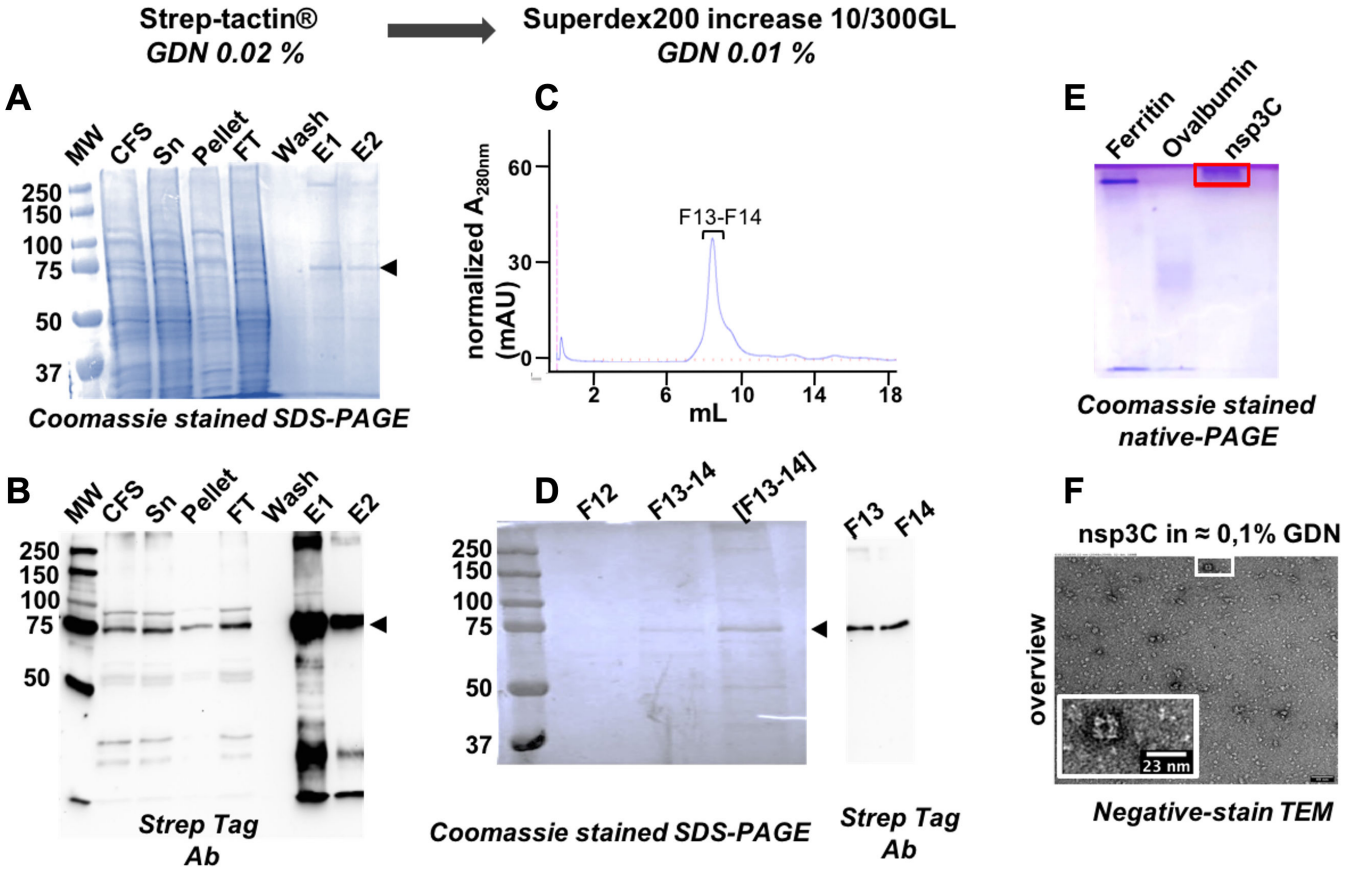

FIG 3 Purification of nsp3C for structural studies. (A) The GDN-solubilized fraction of nsp3C (CFS extract) was first submitted to an affinity chromatography step (Strep-Tactin in 0.02% GDN) and the detergent-solubilized protein was collected in the elution fractions (E1 and E2). The corresponding fractions were analyzed by SDS-PAGE and either stained with Coomassie blue (A) or after western-blotting using an anti-Strep-Tag antibody (B). Fractions E1 and E2 were then submitted to a size exclusion chromatography (Superdex200 increase 10/300 GL column) in a buffer composed of 100 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH 8 supplemented with GDN 0.01%. The corresponding fractions were analyzed by SDS-PAGE stained with Coomassie blue (D, left panel) and, after western-blot, with an anti-Strep-Tag antibody (D, right panel). Fractions F13 and F14 corresponding to the nsp3C protein were pooled and concentrated. Coomassie-stained native PAGE was performed to observe the oligomeric state of nsp3C. Soluble ferritin and ovalbumin were used as calibration proteins in the absence of alternative membrane protein marker (E). An aliquot was deposited onto an airglow-discharged carbon-coated grid to be analyzed using negative-stain TEM. Representative micrograph image is shown with a zoomed view illustrating the presence of spherical particles and oligomeric assemblies with a pore at their center (F).

the self-assembly in order to obtain nanodiscs of uniform size. We used the MSP1E3D1 scaffold protein which is known to form nanodiscs of ~12–14 nm in diameter. MSP1E3D1 bearing a N-terminal 6His-tag was expressed and purified as described in reference (47). After affinity purification, fractions were analyzed onto coomassie-stained SDS-PAGE (data not shown).

Empty nanodiscs were prepared using theoretical MSP1E3D1:POPC molar ratio of 2:260. Nsp3C-containing nanodiscs were first prepared with a theoretical nsp3C:MSP1E3D1:POPC ratio of 1:2:256, albeit the nsp3C and MSP1E3D1protein concentrations were most likely overestimated. Nanodisc samples with (ND-nsp3C) and without (empty ND) nsp3C were injected onto a Superose 6 Increase 5/150 GL gel filtration column pre-equilibrated in 150 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH 8.0. The highest peak of UV absorbance at 280 nm obtained with empty nanodiscs (i.e., without nsp3C) corresponded to an elution volume of 1.7 mL (Fig. 4B). For nsp3C reconstitution in lipid nanodiscs, we considered a molar ratio of nsp3C that gave a minimal content of empty NDs within the fraction. The conserved ratio allowed us to obtain a ND-nsp3C sample which displayed a maximal peak at an elution volume of

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

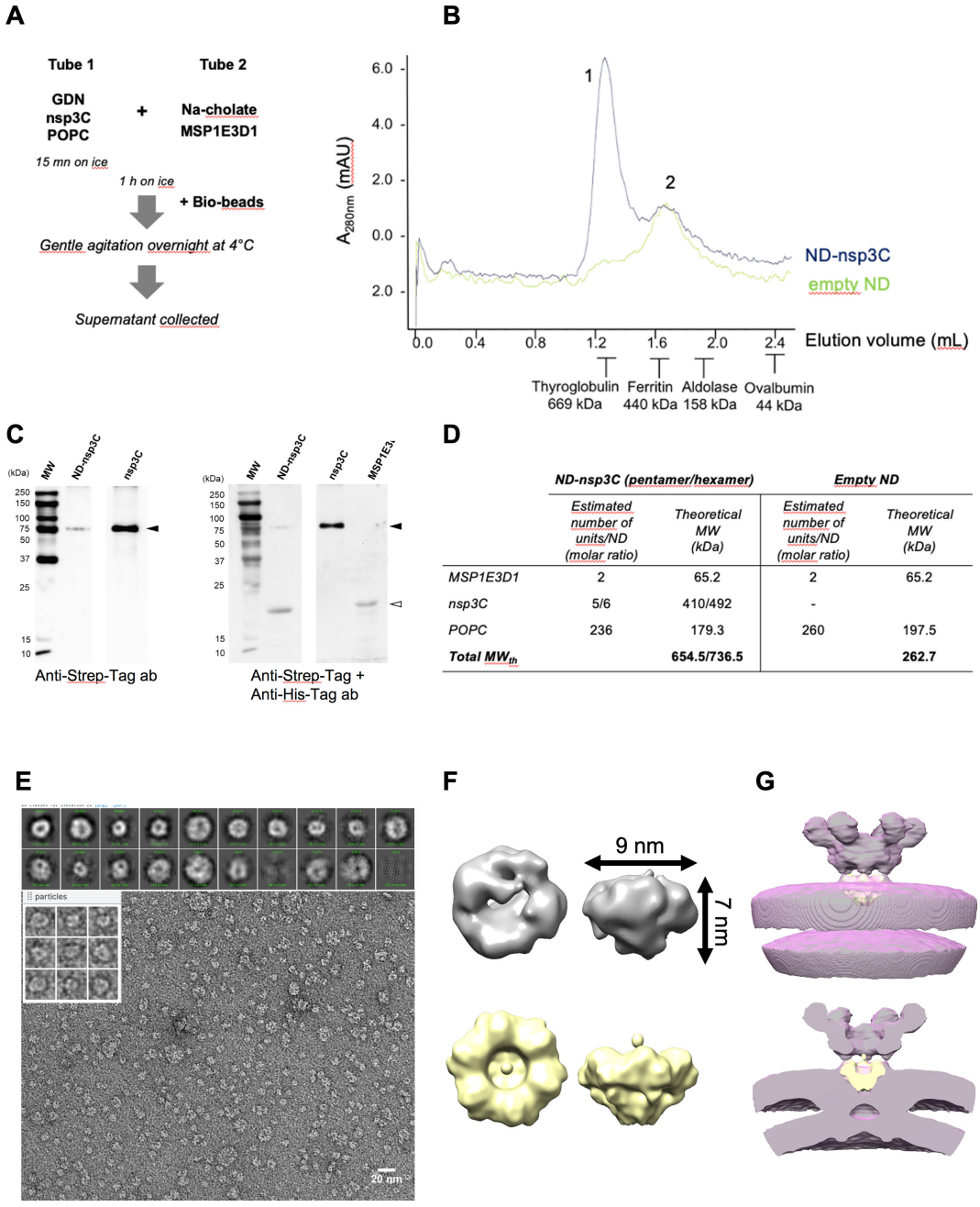

FIG 4 Reconstitution of nsp3C into nanodiscs. The overall procedure used for ND-nsp3C preparation is shown in (A). For empty NDs, nsp3C was omitted. After reconstitution into lipid ND of the concentrated nsp3C elution fractions solubilized in GDN, the supernatant was submitted to SEC chromatography on a superose 6 Increase 10/300 GL column to separate assembled nanodiscs incorporating nsp3C (ND-nsp3C, peak 1) from empty nanodiscs (peak 2) (B). After reconstitution, the ND-nsp3C sample was further concentrated and applied onto a 10% SDS-PAGE for protein separation. Proteins were sequentially visualized using antibodies directed against both Strep-Tag (for nsp3C) and His-Tag (for MSP1E3D1). Purified and concentrated nsp3C, as well as purified MSP1E3D1, were used as control (C). Expected sizes for ND-nsp3C (pentameric or hexameric forms) and for empty nanodiscs are indicated (D). TEM micrograph of ND-nsp3C (scale bar, 20 nm) and representative particles extracted from the images. 2D class averages obtained for 446 picked particles are shown on top (E). 3D reconstructions obtained from 8,300 particles corresponding to the most well-defined, medium-sized object are displayed in "top" view (left) and "side" view (right), filtered to 15 Å, with symmetries imposed as C1 (no symmetry, gray volume) or C6 (yellow volume) (F). (G) Side view of the C6 symmetrized volume (yellow) positioned in the subtomogram average of the C6-symmetric nsp3-4 pore (purple) (29) (top, complete view and bottom, slice view with half of the volumes removed to display the centers of the reconstructions).

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

1.2 mL corresponding to bigger objects. Indeed, the use of calibration proteins indicated that the elution peak at 1.2 mL (peak 1) corresponded to objects with estimated MW of approximately 670 KDa. The ND-nsp3C fraction was analyzed onto SDS-PAGE, and nsp3C and MSP1E3D1 were visualized after western-blotting using anti-Strep tag and anti-His tag antibodies. Both proteins were confirmed to be present in the ND-nsp3C sample (Fig. 4C). Taken together, these results suggest successful incorporation of nsp3C into POPC nanodiscs, and the MW value is consistent with an hexameric assembly (Fig. 4D).

A membrane environment supports a pore organization for the C-terminal region of nsp3

We then attempted negative-stain TEM again, this time on nanodisc-reconstituted nsp3C. Micrographs of ND-reconstituted nsp3C negatively stained with uranyl acetate showed no aggregation but heterogeneity in the sizes of particles on grids (Fig. 4E). *Ab initio* volume generation followed by sorting of 38,807 initially picked particles into 3–6 3D classes showed that particles consistently partitioned into volumes of three different sizes. The smaller-sized volumes are featureless and noisy and comprise most particles (about 20,000). They likely correspond to empty nanodiscs. The larger-sized volume comprises about 9,600 particles. The medium-sized volume is the most well-defined and arranged as a two-layered ring of 9 nm diameter and 7 nm thickness with a central pore (Fig. 4F). This ring reconstructed from some 8,300 particles showed features that could indicate 6- or 5-fold rotational symmetry, or possibly a mixture thereof. The 3D reconstruction with sixfold rotational symmetry is also shown in Fig. 4F. The volumes match very well the expected location of nsp3C in the very recent subtomogram average of the nsp3-nsp4 pore (29) (Fig. 4G). Particularly, the wider layer of the ring would comprise the betaSM domains as placed by Zimmermann et al., while the narrower layer matches a more constricted region of the nsp3-nsp4 pore in the membrane on the cytosolic side of DMVs. We used the C6 reconstruction in further analysis combining this single-particle image analysis with molecular modeling.

Predicted structural models using AlphaFold2 software

In the early 2020s, DeepMind revolutionized the world of structural biology by announcing that its AlphaFold machine-learning method system had significantly outperformed all other methods in the biennial Critical Assessment of Protein Structure Prediction (CASP) and by making an unprecedented number of human and SARS-CoV-2 non-membrane protein structure models available to the community (48, 49).

In order to investigate the oligomeric properties of nsp3C, we used the latest AlphaFold2 (AF2) prediction tool to compute models of nsp3C in pentameric and hexameric oligomeric states. To do that, we first generated models of monomeric nsp3C (E1224-G1945, 722 residues). As shown in Fig. 5A, the protein is mainly composed of alpha-helices and some β-strand regions that are linked by disordered parts. With the exception of a triple-stranded β-sheet predicted in the betaSM region, β-strands were found exclusively in the Y domains. The AF2 model shows two structural blocks predicted with two confidence levels. In fact, pLDDT (predicted Local Distance Difference Test) values above 90 were found for the last 300 residues, indicating high prediction confidence (50). They correspond to residues in the Y1 and CoV-Y domains. By contrast, between positions 100 and 350, low pLDDT values (below 50) suggested potential intrinsically disordered regions (Fig. 5B). The boundaries for the TM domains considered here are slightly different from those proposed in previous works (6, 10, 24). Particularly, two long helices containing TM1 and TM2 are now S1406-V1434 and L1500-L1515, respectively (Fig. 1B).

Bridging TM1 and TM2, the 3Ecto region is exposed toward what would be the ER lumen and structured as a small alpha-helical domain followed by a flexible loop. The four cysteine residues are close in space and potentially form intramolecular disulfide bonds. Next to TM2, AH1 is composed of three hydrophobic helices.

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

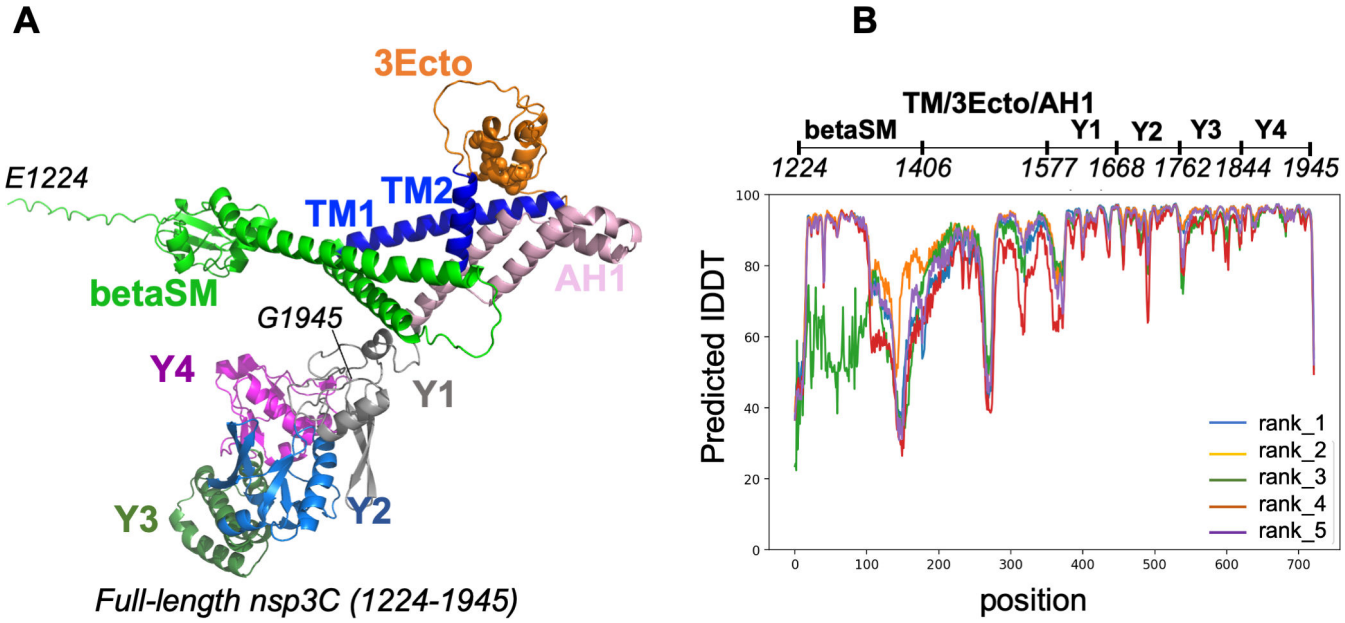

FIG 5 Computed structure of nsp3C (full sequence, 722 residues, E1224-G1945). (A) Model of nsp3C monomer computed using AlphaFold2 machine learning. The four conserved cysteine residues present in the 3Ecto domain of nsp3C are depicted as orange spheres. The Y domains (Y2–Y4) forming the Y-CoV shape are indicated, following boundaries recently published (7). (B) Confidence in the prediction of five models is given by plDDT values per residue position. Values above 90 indicate high accuracy, while pLDDT scores below 50 indicate low confidence (51).

With a model of monomeric nsp3C in hands, we then generated 3D models of the protein with different oligomeric shapes using the AlphaFold2-Multimer software specifically built to predict protein-protein complexes (52). The strategy was the following: a dimer was produced first, based on the model of the monomer. Models of pentamers and hexamers were then computed using nsp3C sequences of different lengths. However, at this stage, the 3D models showed great structural variability in the N-terminal region of nsp3C (E1224-A1557) due to the presence of several unstructured parts, generating, therefore, multiple possible foldings. Thus, we first considered input a truncated sequence of nsp3C from residues P1558 to G1945 that includes part of the AH1 helix and the Y1–Y4 domains in the C-terminus of nsp3.

The computed structure of this short construct with a hexameric symmetry revealed an assembly that forms a channel with a central pore of ~20 Å (Fig. 6A). The height of the pore-like assembly is ~131 Å, and the outer diameter is ~135 Å. Electrostatic potential was mapped to the surface of the hexamer model (Fig. 6B). The front view shows that the positively charged side chains are concentrated around the pore, while negatively charged residues are mostly distributed at the bottom of the assembly. A 90° rotated view clearly shows a crown-shaped arrangement of positive residues with six positive patches along each unit encircling the assembly. The positive character of the core is compatible with the passage of RNA molecules. Several arginines present in the Y1 domain are responsible for the positive enrichment of the core and two of them (R1613 and R1614) largely contribute to the constriction of the pore (Fig. S3). Interestingly, the arginine residue at position 1614 is well conserved in coronaviruses, while the arginine residue at position 1613 is replaced by a hydrophobic residue in MHV and SARS-CoV (Fig. S2). This potentially modifies the size and/or the charge of the pore and may affect the RNA export efficiency among coronaviruses. The protein displays a hydrophobic distribution with a particularly hydrophobic N-terminal region (Fig. 6C). The latter corresponds to the helical part at the end of the AH1 domain, which is exposed in the truncated hexamer but should be internalized in a longer nsp3C construct. In addition to this helix, each protomer of the truncated assembly encompasses the Y1 domain and the three subdomains Y2–Y4 of the CoV-Y domain (Fig. 6D, left panel).

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

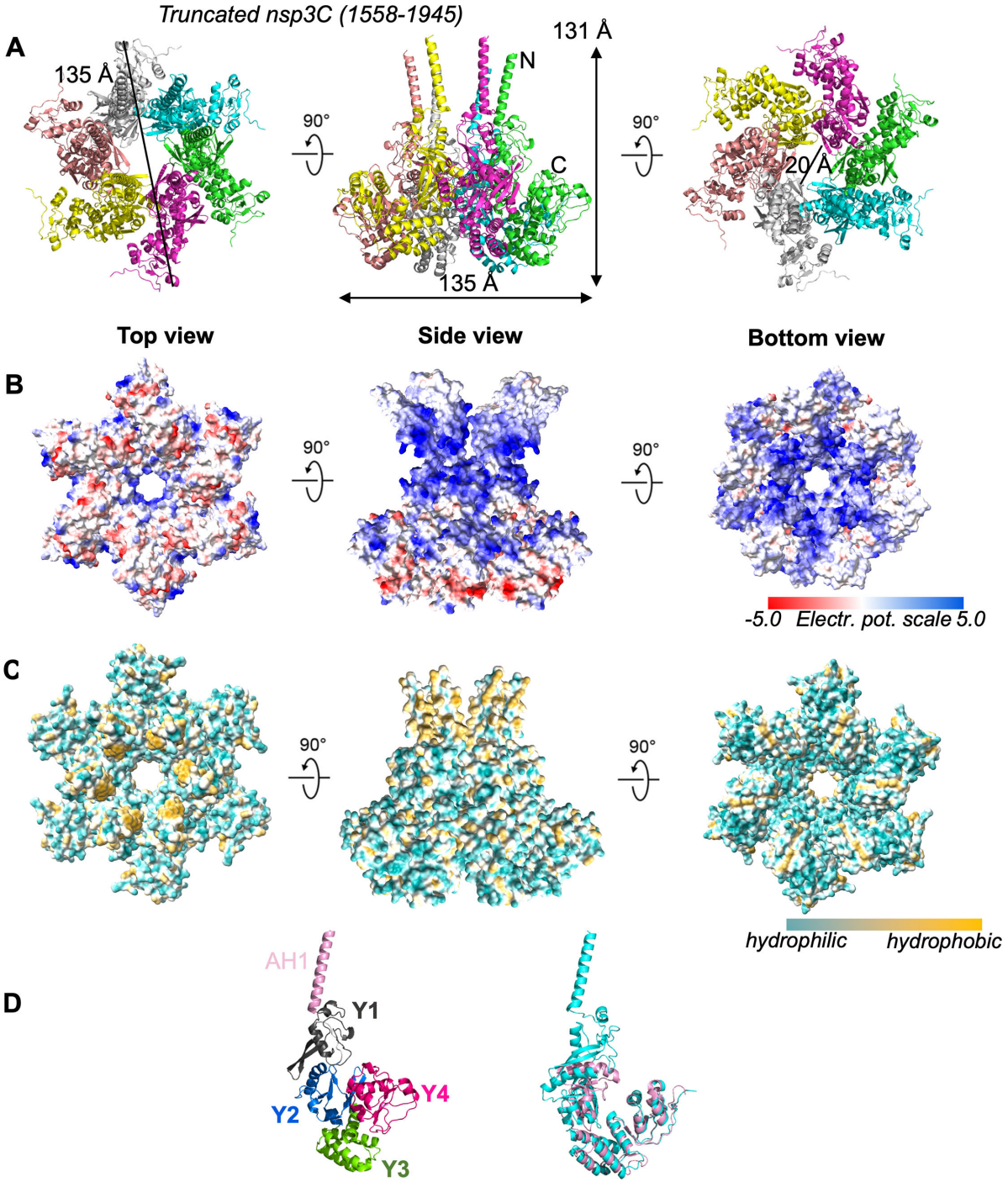

FIG 6 Predicted hexameric structure of nsp3C (truncated version, P1558-G1945). (A) AlphaFold2 model of the hexameric assembly. The six protomers are highlighted with different colors. Three views (top, side, bottom) indicate dimensions of the nsp3C hexamer. (B) Surface electrostatic representation of the hexamer model. Same orientation as in (A). Negative and positive charges are colored red and blue, respectively. (C) Hydrophobic representation of the hexamer model. (D) Left, model of a protomer with the Y1–Y4 domains highlighted in different colors. The helical part that precedes the Y1 domain is the C-terminal segment of the AH1 domain (pink). Right, the model (cyan) is superimposed onto the recently solved crystal structure of the CoV-Y domain (pink) (7).

The predicted structure of this region, indeed, matches the recently reported crystal structure of the CoV-Y domain with its V-shaped fold (7), with a RMSD of 2.1 Å (213 atoms superimposed). Noteworthy, the Y3 and Y4 domains superimpose onto the crystal structure with an excellent RMSD score (1.2 Å for 158 atoms), Fig. 6D, right panel. A

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

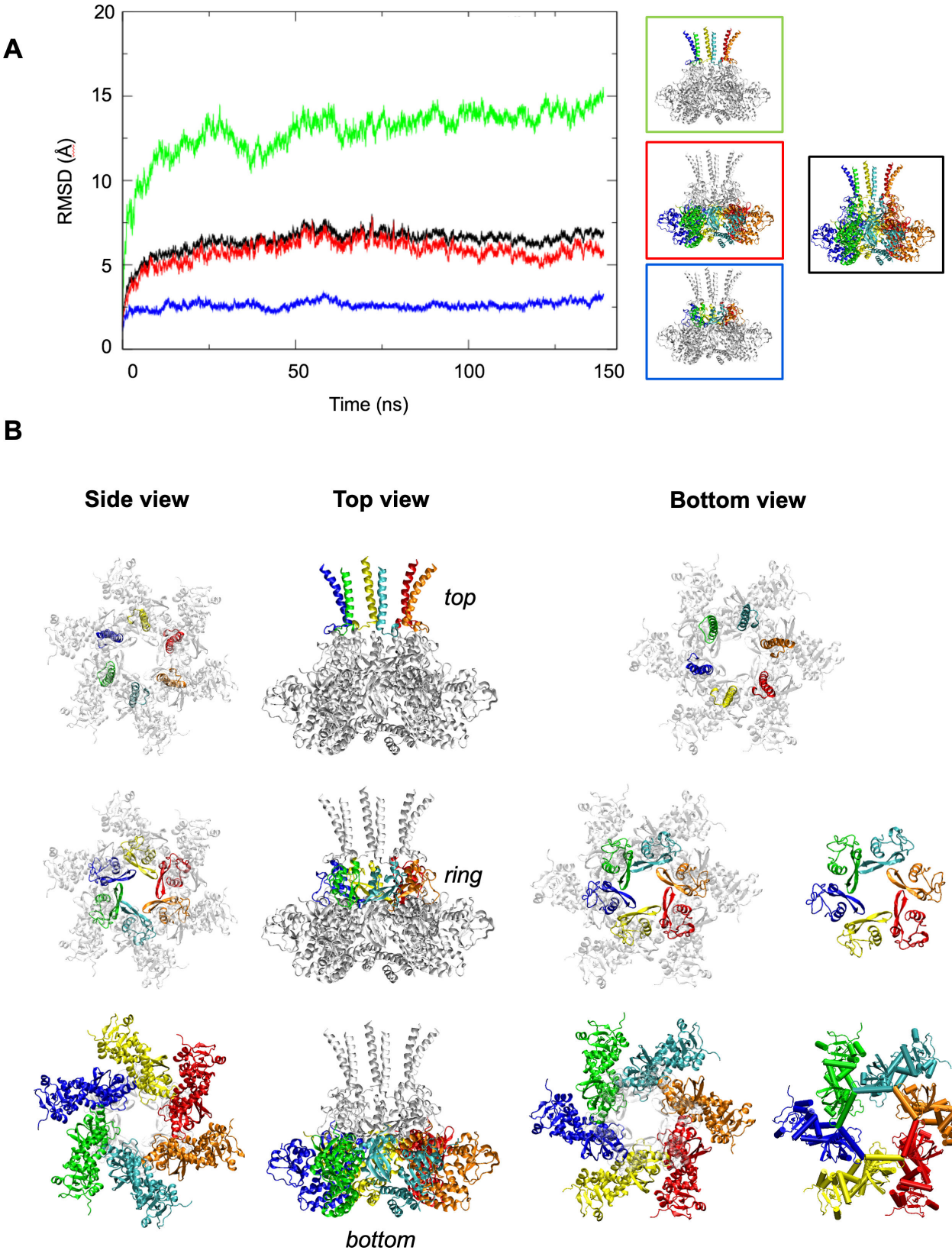

FIG 7 Molecular dynamics in explicit solvent of the hexamer model of nsp3C (truncated version, P1558-G1945). (A) RMSD plot of backbone atoms throughout the MD simulations (black, full hexameric assembly; green, top region; red, bottom region; blue, ring region). (B) Three views (side, top, and bottom) are shown for three regions of the assembly that experience distinct dynamic behaviors (top, ring, bottom). The right panels give a view of the ring region organized around β-strands and a view of the bottom region that relies on six horizontal helices forming a hexagonal platform.

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

model of pentameric assembly was also generated (Fig. S4), displaying a pore size half that of the hexameric model.

Molecular dynamic simulations

Molecular dynamics (MD) simulations (150 ns) in explicit solvent were performed to examine the structural stability of the truncated hexameric assembly (residues P1558 to G1945). Starting from a minimized structure, we could observe that the global RMSD of the hexameric structure for the backbone atoms stabilizes after about 15 ns, reaching a plateau close to 7 Å (black curve, Fig. 7A). Further analysis revealed that three regions of the hexamer displayed distinct dynamic behaviors during MD simulation. The N-terminus showed strong fluctuations and greatest instability, reaching final high RMSD values of 13.2 Å (green curve, Fig. 7A). This N-terminal region of the truncated hexamer contains six exposed and flexible helices (residues I1559-V1583) with no contact between them (Fig. 7B). By contrast, the central part (83 residues, amino acids G1585-D1667) is particularly stable during the simulation. The RMSD plot indicates that the structure stabilizes shortly after a few ns of simulation reaching a plateau with a RMSD value of 2.5 Å (blue curve, Fig. 7A). This region corresponds to the narrowest part of the hexamer assembly, which forms a ring of ~2 nm with little fluctuation during molecular dynamics. Interestingly, this "ring" forming region predicted as α + β structure corresponds to the Y1 domain whose function was previously unknown. The fold of this "ring" region in the hexamer would be similar to that found in pseudouridine synthase, as assessed by the DALI server (53) and consistent with a RNA chaperone function. Several inter-chain contacts contribute to the stability of this region (Fig. S5A). In particular, two salt bridges are found between R1613 and E1604, and between R1602 and D1654. Noteworthy, this central region of the hexamer model falls within the highest accuracy score of AF2 (Fig. S5B), in connection with the high conservation of the Y1 domain among members *of the Nidovirales*.

Finally, the C-terminal part of the structure from residues L1709 to G1945 is quite flexible with a plateau achieved at 15 ns after which backbone RMSD fluctuations remain around 6.3 Å (red curve, Fig. 7A) a value close to that obtained for the global structure. This "bottom" region explores a fairly wide amplitude range and exhibits conformational fluctuations of opening and closing, like movements of flower petals. It rests on a hexagonal platform made up of six practically horizontal helices present in the Y2 domain (residues K1733-M1750) (Fig. 7B). In each protomer, the platform forming helix [H23 in the crystal structure of Y2, (7)] slides along the beta hairpin of the Y1 domain. This leads to a polar contact observed between two helices coming from Y1 and Y2 oriented perpendicularly (Q1748 in Y2 and S1648 in Y1).

We then considered the full sequence of nsp3C (722 residues, amino acids E1224-G1945) to compute structural models (Fig. 8A). To do this, we first used alpha-Fold2 (AF2) software to predict the structure adopted by the first 333 amino acids (residues E1224 to A1557) that include the betaSM and 3Ecto domains and transmembrane helices. Here again, several predictions were done considering various lengths and various structural blocks. We then added this model to the structure of the truncated nsp3C described above (P1558-G1945). The models indicated the presence of numerous helices predicted with varying degrees of accuracy (Fig. 8A). Interestingly, the two TM helices, TM1 (S1406-V1434) and TM2 (L1500-L1515), described in the AF2 model of the nsp3C monomer are located in two separate blocks that are connected by a poorly structured sequence which corresponds to the 3Ecto domain (Fig. 8B and 9). The computed full-length model of nsp3C hexamer model showed several long helical structures in these two blocks (Fig. S6). We, therefore, sought to reassess the prediction of transmembrane helices and their position in the model using the DeepTMHMM server based on the latest deep neural networks (54). Four TM domains were predicted with high probability (Fig. 8C; Fig. S7), with the extra TM3 and TM4 in formerly annotated AH1. TM3 (L1523-F1540) lies immediately downstream TM2, while TM4 (residues I1559-K1577) is a segment of a longer helical structure at the C-terminus of AH1 (I1559-V1583) which

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

was found to be exposed on the top of each protomer of the truncated hexamer model (Fig. 8D).

DISCUSSION

The nsp3 protein was proposed as the major component of the molecular pore complex spanning DMVs in MHV-infected cells (28). Six nsp3 molecules could contribute to the crown-shaped pore architecture, with the N-terminal part of nsp3 residing in the six branches. Whether other components are required, and which of these are essential and sufficient to form a pore complex in coronaviruses remained to be elucidated. During the reviewing process for the present work, a study conducted in transfected cells expressing SARS-CoV-2 nsp3-4 proteins brought to light that nsp3 and nsp4 are the two minimal components required to form the molecular pore spanning the DMVs (29). The $PL2^{pro}$ cleavage of the nsp3-4 fusion protein is not only required to observe formation of SARS-CoV-2 DMVs as previously proposed in MERS coronavirus (18) but also allows assembly of the pore (29). The integrity of the pore's crown-like structure relies on the presence of nsp3 N-terminal domains. However, how the C-terminal region of nsp3 -encompassing the transmembrane segments- contributes to the pore organization remains to be elucidated. Except for the 3Ecto domain that connects the TM segments and faces the lumen of the DMVs, all nsp3 regions are expected to be on the cytoplasmic side (20).

Our experimental data indicate that nsp3C is able to oligomerize in solution and after reconstitution in nanodiscs our negative-stain TEM images revealed a two-layered ring in the sample. To successfully express and reconstitute nsp3C into POPC nanodiscs, it was necessary to consider a construct starting from the betaSM domain and extending to the CoV-Y domain. We could readily place our EM volume of nsp3C at the bottom part of the prongs in the subtomogram average of the nsp3-nsp4 pore published by Zimmermann et al. (29), see Fig. 4G. The wider part of the two-layered ring would thus, correspond to the betaSM domains.

Taking into account the presence of six nsp3 molecules in the molecular pore complex (28, 29), we set out to model an hexameric assembly of nsp3C. To do that, we had to play with several sequence lengths and types of oligomers to improve the accuracy of the 3D predicted structures. The AF2 model showed a ring-like organization, consistent with our experimental data and our data confirm the ability of nsp3C to form oligomers with a crown-like organization. The constriction of the pore is provided by a central α + β region (ring), which corresponds to the Y1 domain. Its great stability during MD simulation together with its high conservation among the *Nidovirales* members suggests a key role to maintain and stabilize the hexameric assembly, particularly at the level of pore constriction. While our EM data cannot presently establish that this organization is correct, they do show a constriction of the pore in the nsp3C oligomer consistent with both the AlphaFold hexamer model and the experimental data of Zimmermann et al. (29).

The membrane disposition of SARS-CoV-2 nsp3 has not been fully elucidated yet. This task is made all the more difficult by the fact that nsp3 is involved in membrane rearrangement with nsp4 and nsp6, and helical segments in the TM region may be associated with membranes without spanning them. Our AF2 models of nsp3C display two structural blocks where the two transmembrane segments, TM1 and TM2, unanimously proposed to span the membrane bilayer are distant and separated by the luminal 3Ecto domain. TM2, TM3, and TM4 helices would be in the same structural block (Fig. 8). A more precise position of the four predicted TM helices in the sequence of nsp3C (Fig. 9) was deduced from analysis of several prediction programs.

While AlphaFold has shown impressive capabilities in predicting the overall fold and structure of proteins, the accurate prediction of transmembrane segments still remains challenging. This difficulty arises from AlphaFold's tendency to cluster transmembrane helices with other hydrophobic elements. Thus, the nsp3C hexamer model must be taken with caution as the TM1 segment is wrongly positioned (Fig. S6). The assignment of TM3

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

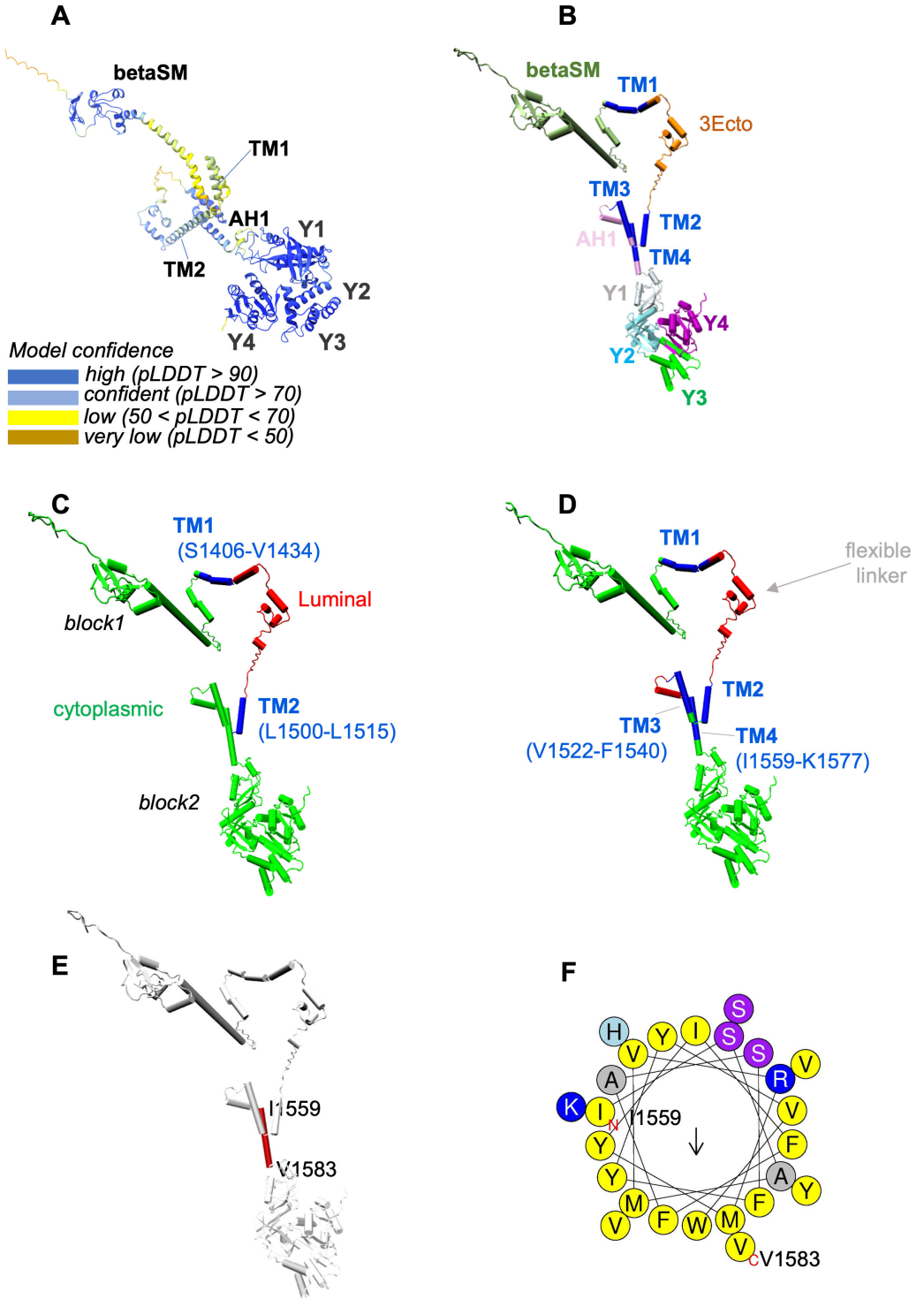

FIG 8 Mapping of the TM segments in the AF2 model (full nsp3C sequence, E1224-G1945). (A) The model of the monomer is shown with a color-based code according to per-residue confidence score (pLDDT). (B) View of a protomer showing the different domains of nsp3C, including the putative transmembrane helices TM3 and TM4. (C) View of a protomer with the location of the two transmembrane helices (TM1 and TM2). The inter-domain linker is shown in black-orange and corresponds to the 3Ecto domain that is known to point to the luminal side. Blocks 1 and 2 should be both oriented towards the cytoplasmic side. (D) Same view with four TM helices (TM1–TM4) predicted using the deepTMHMM algorithm (54). (E, F) The predicted TM4 is found at the N-terminus of a long alpha-helix (I1559-V1583) rich in hydrophobic residues, according to Heliquest program (55). The flexible nature of the linker sequence (3Ecto domain) could allow the helices to be rearranged.

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

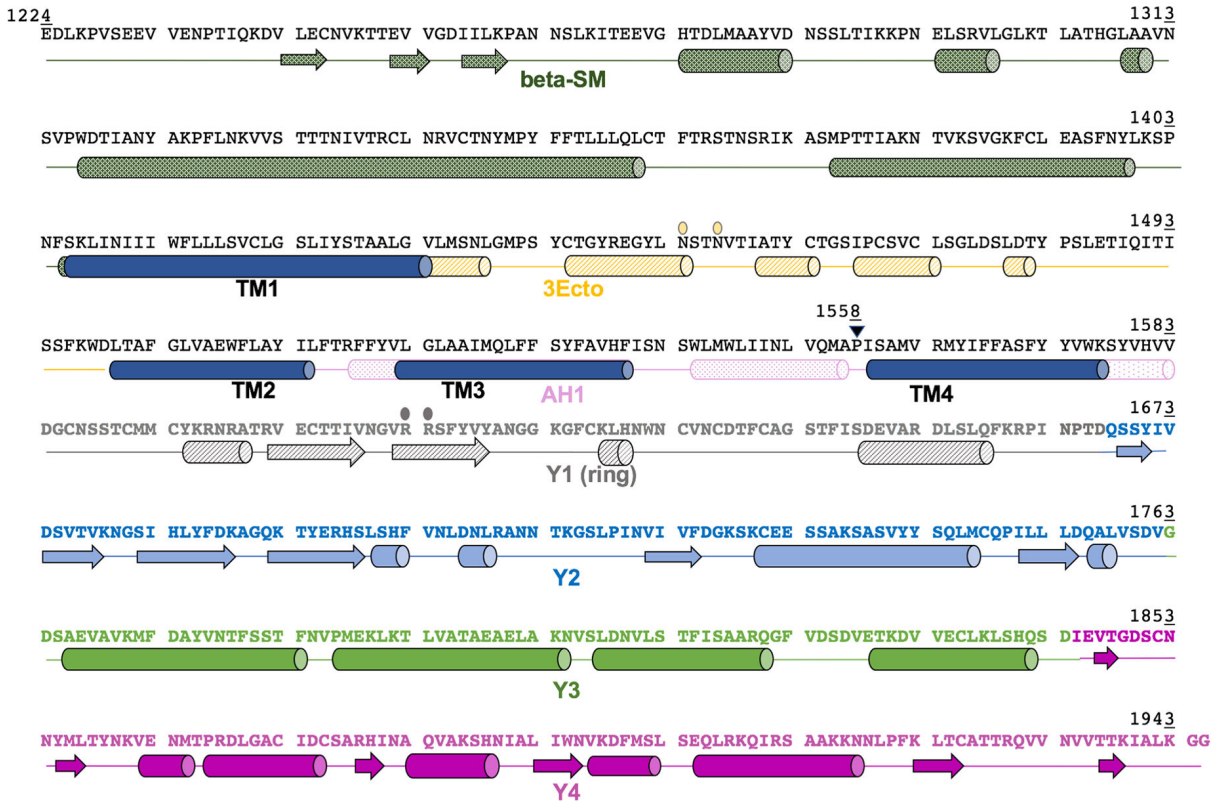

FIG 9 Sequence of nsp3C with a revised domain organization (full sequence, E1224-G1945). The nsp3C sequence used in this study starts at residue E1224 (numbering corresponds to the full-length nsp3 protein). The streptavidin tag present in the N-terminal part is not taken into account. The secondary structure elements (cylinders for helices and arrows for β strands) using AF2 program are depicted with a grid pattern. They span the betaSM (green) and AH1 domains (light pink), the 3Ecto domain (orange), and the Y1 domain (gray). The four predicted helices TM1–TM4 according to DeepTMHMM algorithm are shown in blue. Dark colors are chosen for α-helices and β-strands that were reported in the crystal structure of the CoV-Y domain (7). The sub-domains Y2, Y3, and Y4 are shown in blue, green, and magenta, respectively. Small gray spheres indicate the two arginine residues present in the ring that lead to the constriction of the pore. Small yellow spheres indicate the asparagine residues that act as putative N-linked glycosylation sites. The vertical arrow (position P1558) indicates the sequence start for AF2 modeling of the truncated version of nsp3C.

and TM4 in AH1 must remain tentative, as their packing against the transmembrane part in our models could either reflect an alternate conformation of nsp3C with four TMs, as proposed below, or be a result of the aforementioned known behavior of AlphaFold2. Supporting the former hypothesis, the most up-to-date software for transmembrane helix prediction does assign TM3 and TM4 as transmembrane segments. A long helical structure (I1559-V1583) which was found to be exposed on the top of each protomer of the truncated hexamer model encompasses the predicted TM4 segment. This helix is made up of 25 residues, 19 of which are hydrophobic (Fig. 8E). Its size and hydrophobic nature would be sufficient to cover the 30 Å hydrophobic core of a lipid bilayer (56). There is a conserved histidine residue at the end of the helix that could play a role in membrane insertion (57). On the basis of our models, it is very tempting to consider a rotation around the connecting element, which would bring the four TM segments closer together. TM helices can undergo concerted motions when a membrane protein is at work, and they are key determinants for oligomerization and proper folding of membrane proteins (58, 59). We suggest that nsp3 could be inserted into the lipid bilayer by its TM1 and TM2 upon synthesis, with AH1 remaining on top of the cytosolic leaflet as previously reported, and that nsp3 could contain four transmembrane helices when the protein is oligomeric, with TM3 and TM4 in AH1 then spanning the membrane. The membrane insertion of TM3 and TM4 would require reorientation of TM helices and large conformational changes made possible by the flexible character of the 3Ecto domain.

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

The betaSM domain upstream of the TM region comprises a N-terminal disordered tail that could establish transient interactions for proper position of TM helices. Note that in our model, the N- and C-terminus are not correctly oriented to be both exposed towards the cytosol. Transmembrane topological changes and reorientation of TM domains at successive steps in the formation of viral complexes have been reported in previous studies on hepatitis B and hepatitis C virus proteins (60, 61). Concerning nsp3C, bringing blocks 1 and 2 together would allow them to face the cytoplasmic side while exposing the 3Ecto domain to the luminal face (Fig. 8). Putative sites for *N*-linked glycosylations in the 3Ecto domain were proposed in SARS-CoV and MHV (24). Two asparagines (N1454 and N1457) are present in the 3Ecto domain of nsp3 and could constitute signals for conformational changes (Fig. 9). At this stage, we must keep in mind that our experiments were conducted with a truncated protein and without its nsp4 partner. It is not excluded that rearrangement of the TM helical segments could occur later during DMV formation upon interaction with other components such as nsp4. In the recent paper conducted by Zimmermann et al., nsp4 and nsp3 were shown to be the two minimal constituents of the SARS-CoV-2 pore (29). Therefore, implementing our experimental approach to the study of the nsp3-nsp4 complex should provide a more comprehensive view.

MATERIALS AND METHODS

Cell-free protein expression

Synthetic genes of, respectively, 2,208 and 1,674 bp for the long and short versions of nsp3C were designed based on the nsp3 coding sequence from ORF1ab polyprotein (Seq ID : PP043139.1). The nsp3 template coding sequence ranging from nucleotides 6,351 to 8,516 (long version) and 6,891–8,516 of the ORF1ab was used for synthesis to obtain the coding region of the long and short versions of nsp3C. Start and stop codons were added in 5′ and 3′ of both sequences. A strep-tag II coding sequence (TGGTCTCATC CTCAATTTGAAAAA, for amino-acids sequence: WSHPQFEK) was inserted upstream both genes surrounded by 6–12 nucleotides linkers. Both genes were subcloned (*Spe*I-*BamH*I) into a pEU-E01 vector. Plasmids were amplified in *E. coli* (DH5alpha) and purified using a maxiprep kit (Millipore) followed by a phenol/chloroform extraction. Transcription was performed during 7 h at 37°C, and the transcription product was mixed with wheat germ extract for cell-free translation. Translation was performed overnight at 22°C in 96-well microplates (detergent screening) or in 6-well plates (large-scale production). We used bilayer-based cell-free protein synthesis system (62), and GDN detergent (final concentration 0.1%) was added to both layers for direct solubilization of the protein. After protein synthesis, the reaction was stopped by incubating the mixture with home-made purified benzonase during 1 h at room temperature prior to centrifugation at 20,000 *g* during 30 min (4°C).

Detergent screening experiment

Cell free expression was tested in different detergents that were directly included to the reaction mixture at a final concentration of 0.1%.

After centrifugation of CFS, the efficacy of protein solubilization was determined by comparing the amount of protein in the pellet (insoluble fraction) to that of the soluble fraction in the supernatant (Sn fraction). Analysis of protein samples was done by coomassie brilliant blue-stained SDS-PAGE and by western blotting (see below). Comparison between the bands obtained for the CFS, the pellet (P), the supernatant (Sn) with the different detergents allowed to select the most favorable detergent for protein solubilization.

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

Protein purification

After incubation with benzonase and centrifugation, the obtained supernatant was mixed with Strep-tag affinity resin (StrepTactin superflow high capacity, IBA Lifesciences) on a wheel for 3 h. The resin was washed 5 times with 1 CV of washing buffer composed of 100 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM EDTA (W1X buffer, iba Lifesciences), supplemented with 0.02% GDN. Elution of the strep-tagged proteins was done using 5 CV of the same buffer supplemented with 2.5 mM desthiobiotin. Elution fractions were pooled and further purified using SEC chromatography (Superdex 200 Increase, Cytiva) in W1X buffer with 0.01% GDN. The protein was detected on Western Blot using anti-Strep-tag antibodies (IBA Lifesciences). Briefly, proteins were separated by electrophoresis on 10% Tris-glycine gels and transferred onto nitrocellulose membrane (GE Healthcare Life Sciences). Membrane blocking was performed in 10% nonfat milk powder in PBS-T (150 mm NaCl, 50 mM Tris-HCl, pH 7.9, 0.1% Tween 20) and probed with monoclonal anti-Strep-tag antibodies (IBA-Lifesciences). All primary antibodies were diluted to 1:1,000 in PBS-T + 2% milk. Blots were washed three times in PBS-T and incubated with anti-mouse IgG-HRP conjugate secondary antibodies (Promega) (1:10,000 in PBS-T + 2% milk). Revelation was achieved using ECL Prime Western Blotting Detection Reagent (GE Healthcare Life Sciences), and images were taken with a CCD camera (Gbox, Syngene).

Reconstitution into nanodisc

Membrane scaffold protein MSP1E3D1 was expressed in *E. coli* and purified on a Ni-NTA agarose resin according to reference (63). Lipid POPC (Avanti) was dissolved in chloroform and dried under steady flow of Argon followed by exposure to vacuum for 1 h and stored at −20°C. The lipid film was resuspended in a buffer containing 10 mM Tris-HCl, 100 mM NaCl, pH 8.0. POPC solubilization was achieved by a 5-min incubation with GDN at Rsol on ice. To our knowledge, GDN has never been fully characterized biophysically, and the Rsol concentration was unknown. Therefore, we used Rsol value from cholate, which is a compound chemically close to GDN. This approximation proved to work well to fully solubilize POPC. Solubilized lipids were then added to the purified nsp3C protein solubilized in GDN and incubated on ice for another 15 min (tube 1). MSP1E3D1 was mixed with 15 mM sodium cholate in the same buffer containing 10 mM Tris-HCl, 100 mM NaCl, pH 8.0 (tube 2). The two tubes were mixed and incubated an hour on ice. To estimate the molar ratio, nsp3C concentration was estimated by measuring the absorbance at 280 nm on a NanoDrop spectrophotometer (epsilon 99,710 M^{-1} cm^{-1}). Nanodisc self-assembly was performed upon the addition of hydrophobic Bio-beads (Biorad) and incubation at 4°C overnight under constant stirring. Bio-beads were then removed after a short centrifugation, and protein-containing nanodiscs were further analyzed on a Superose6 chromatography (Superose 6 Increase 10/300 GL; Cytiva) in a running buffer containing 25 mM Tris-HCl and 150 mM NaCl, pH 8.0.

Negative-stain transmission electron microscopy

Four microliters of a tenfold diluted suspension of freshly reconstituted nsp3c into lipid nanodiscs was deposited onto an airglow-discharged carbon-coated grid. Excess liquid was removed, and the grids were negatively stained with 2% (wt/vol) aqueous uranyl acetate. Micrographs were taken on a Tecnai Spirit 12G transmission electron microscope (ThermoFisher–former FEI) operating at 100 kV and equipped with a K2 Base 4k × 4k camera (GATAN, Pleasanton CA, USA). EM micrographs were recorded at a magnification of 27,500×, resulting in a pixel size of 1.8 Å/pixel. CryoSPARC was used for image analysis. Four hundred forty-six particles were manually picked on five images, and a first set of 2D classes was produced (Fig. 4E, top). From these templates, 77,037 picks were obtained and inspected, resulting in 38,807 initial particles. Subsequent *ab initio* 3D volume generation and heterogeneous 3D refinement was performed to sort particles. Eight thousand three hundred forty-two particles from the most homogeneous 3D class

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

were used to reconstruct a final low-resolution volume with C1 (no symmetry imposed), C5 (assuming a symmetric pentameric object), or C6 (assuming symmetric hexameric) symmetry. We could not further classify these 8,342 particles into subclasses with clear symmetry.

Structural prediction using AlphaFold2

At the beginning of this work, we only had access to Alphafold2 (48) predictions with the free ColabFold web interface, so our requests were limited to sequences of less than 2,000 amino acids. We first used ColabFold to build a 3D model of the C-terminal part of nsp3 corresponding to residues ranging from E1224 to G1945. We then considered the AlphaFold2-Multimer program to generate a dimer of the C-terminus of the protein. From the dimer model, it was possible to build by symmetry a pentamer model. Many requests were then performed (trimer, hexamer of full or incomplete sequence) to obtain different 3D models with different topologies. Using our queries, we were able to identify common surfaces of interactions between structured domains predicted with high pLDDT values (51). Then, as the AlphaFold2-Multimer program was installed locally, we could query for hexamer model with the last 389 amino acids corresponding to the C-terminal sequence from residue A1557 to residues G1945. When this work was initiated, it was not possible to obtain a good model when we requested for a hexamer model, even with the AF2-Multimer program, corresponding to the full 722 residues of nsp3C. Since the N-terminus of the sequence has many unstructured parts, the global fold is not easy to predict. Indeed, the structured domains were well predicted while the global topology was not. Thus, changing the sequence length and the number of oligomers for a request significantly improved the accuracy of the 3D predicted models. Our strategy was to compare many requests to define the common sequences that were identified in the docked parts of the different models.

Manual fitting of the hexamer model to the EM density map was performed using UCSF Chimera (64). Hovewer, while this article is being submitted, the most up-to-date version of AlphaFold2 (2.3) was used and was able to generate a similar hexamer model when fed with the whole nsp3C sequence.

Molecular dynamics simulations

Atom molecular dynamic was performed with the AMBER 20 package (65) and the ff99SB force field. The hexamer model was immersed in a cubic box with boundaries extending from the protein periphery for at least 14 Å in all directions. The box was filled with TIP3P water molecules (163,541 water molecules), and 48 counter ions Cl^- was added to neutralize the total charge of the system. The total size of the system was 531,291 atoms. The system was then minimized using a steepest descent algorithm and used to initiate molecular dynamic simulation. Molecular dynamic simulation was carried out at 300 K. The SHAKE algorithm was used to constrain the motion of hydrogen-containing bonds within a 2fs time step to integrate the equations of motions. The cut off distance for van der Waals interactions was set to 10 Å, and the Ewald particle mesh was used for long-range electrostatic interactions. Production runs of 150 ns was performed on the equilibrated structure with the NPT ensemble using a 2fs time step. The trajectory was analyzed with CPPTRAJ module, VMD 1.9.4 (66) and UCSF ChimeraX (67) programs.

ACKNOWLEDGMENTS

The authors are grateful to M. Ouldali (I2BC) for her help in collecting EM images and L. Daury and O. Lambert (University of Bordeaux) for their valuable input in ND reconstitution. The authors are grateful to G. Lenoir and C. Montigny for the generous gift of MSP plasmids and M. Paternostre for fruitful discussion with lipid preparation.

This work benefited from the CryoEM platform of I2BC, supported by the French Infrastructure for Integrated Structural Biology (FRISBI) [ANR-10-INSB-05-05] and member

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

of IBISA, and from the PAPPSO platform for the LC-MS/MS analysis and was granted access to the CCRT High-Performance Computing (HPC) facility.

AUTHOR AFFILIATION

[1]Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), Gif-sur-Yvette, France

PRESENT ADDRESS

Samira Agouda, UCB Pharma S.A. Belgium, Brussels, Belgium

AUTHOR ORCIDs

Marion Babot http://orcid.org/0009-0009-8777-1986
Virginie Gervais http://orcid.org/0000-0002-1144-6185

AUTHOR CONTRIBUTIONS

Marion Babot, Conceptualization, Formal analysis, Methodology, Writing – original draft, Writing – review and editing | Yves Boulard, Formal analysis, Software | Samira Agouda, Formal analysis | Laura Pieri, Resources | Sonia Fieulaine, Methodology | Stéphane Bressanelli, Funding acquisition | Virginie Gervais, Conceptualization, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

Files for the hexamer model of truncated nsp3C (P1558-G1945) obtained with AlphaFold2 and for the 3D map model (C6 symmetry) generated using cryosparc are available at Zenodo.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Fig.S1 (JVI01575-23-s0001.tiff). Expression assays for the two SARS-CoV-2 nsp3C constructs.
Fig. S2 (JVI01575-23-s0002.tiff). nsp3 C-terminal part sequence alignment.
Fig. S3 (JVI01575-23-s0003.tiff). Pore constriction.
Fig. S4 (JVI01575-23-s0004.tiff). Pentamer model of nsp3C (full sequence, E1224-G1945).
Fig. S5 (JVI01575-23-s0005.tiff). nsp3C hexamer model (truncated version, P1558-G1945) is stabilized by several inter-chain contacts.
Fig. S6 (JVI01575-23-s0006.tiff). Predicted hexameric structure of nsp3C (full sequence, E1224-G1945).
Fig. S7 (JVI01575-23-s0007.tiff). nsp3C topology prediction of the TM domains.
Supplemental legends (JVI01575-23-s0008.pdf). Legends for Fig. S1 to S7.

REFERENCES

1. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, Si H-R, Zhu Y, Li B, Huang C-L, et al. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270–273. https://doi.org/10.1038/s41586-020-2012-7
2. Wang R, Chen J, Wei GW. 2021. Mechanisms of SARS-CoV-2 evolution revealing vaccine-resistant mutations in Europe and America. J Phys Chem Lett 12:11850–11857. https://doi.org/10.1021/acs.jpclett.1c03380
3. Cox RM, Lieber CM, Wolf JD, Karimi A, Lieberman NAP, Sticher ZM, Roychoudhury P, Andrews MK, Krueger RE, Natchus MG, Painter GR, Kolykhalov AA, Greninger AL, Plemper RK. 2023. Comparing molnupiravir and nirmatrelvir/ritonavir efficacy and the effects on SARS-CoV-2 transmission in animal models. Nat Commun 14:4731. https://doi.org/10.1038/s41467-023-40556-8

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

4. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, et al. 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 395:565–574. https://doi.org/10.1016/S0140-6736(20)30251-8
5. Narayanan K, Huang C, Makino S. 2008. SARS coronavirus accessory proteins. Virus Res 133:113–121. https://doi.org/10.1016/j.virusres.2007.10.009
6. Thomas S. 2021. Mapping the nonstructural transmembrane proteins of severe acute respiratory syndrome coronavirus 2. J Comput Biol 28:909–921. https://doi.org/10.1089/cmb.2020.0627
7. Li Y, Pustovalova Y, Shi W, Gorbatyuk O, Sreeramulu S, Schwalbe H, Hoch JC, Hao B. 2023. Crystal structure of the CoV-Y domain of SARS-CoV-2 nonstructural protein 3. Sci Rep 13:2890. https://doi.org/10.1038/s41598-023-30045-9
8. Rohaim MA, El Naggar RF, Clayton E, Munir M. 2021. Structural and functional insights into non-structural proteins of coronaviruses. Microb Pathog 150:104641. https://doi.org/10.1016/j.micpath.2020.104641
9. Oostra M, te Lintelo EG, Deijs M, Verheije MH, Rottier PJM, de Haan CAM. 2007. Localization and membrane topology of coronavirus nonstructural protein 4: involvement of the early secretory pathway in replication. J Virol 81:12323–12336. https://doi.org/10.1128/JVI.01506-07
10. Oostra M, Hagemeijer MC, van Gent M, Bekker CPJ, te Lintelo EG, Rottier PJM, de Haan CAM. 2008. Topology and membrane anchoring of the coronavirus replication complex: not all hydrophobic domains of nsp3 and nsp6 are membrane spanning. J Virol 82:12392–12405. https://doi.org/10.1128/JVI.01219-08
11. Nagy PD, Strating J, van Kuppeveld FJM. 2016. Building viral replication organelles: close encounters of the membrane types. PLoS Pathog 12:e1005912. https://doi.org/10.1371/journal.ppat.1005912
12. Zhang J, Lan Y, Sanyal S. 2020. Membrane heist: coronavirus host membrane remodeling during replication. Biochimie 179:229–236. https://doi.org/10.1016/j.biochi.2020.10.010
13. van der Hoeven B, Oudshoorn D, Koster AJ, Snijder EJ, Kikkert M, Bárcena M. 2016. Biogenesis and architecture of arterivirus replication organelles. Virus Res 220:70–90. https://doi.org/10.1016/j.virusres.2016.04.001
14. Angelini MM, Akhlaghpour M, Neuman BW, Buchmeier MJ. 2013. Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. mBio 4:e00524-13. https://doi.org/10.1128/mBio.00524-13
15. Snijder EJ, Limpens R, de Wilde AH, de Jong AWM, Zevenhoven-Dobbe JC, Maier HJ, Faas F, Koster AJ, Bárcena M. 2020. A unifying structural and functional model of the coronavirus replication organelle: tracking down RNA synthesis. PLoS Biol 18:e3000715. https://doi.org/10.1371/journal.pbio.3000715
16. Hagemeijer MC, Monastyrska I, Griffith J, van der Sluijs P, Voortman J, van Bergen en Henegouwen PM, Vonk AM, Rottier PJM, Reggiori F, de Haan CAM. 2014. Membrane rearrangements mediated by coronavirus nonstructural proteins 3 and 4. Virology 458–459:125–135. https://doi.org/10.1016/j.virol.2014.04.027
17. Sakai Y, Kawachi K, Terada Y, Omori H, Matsuura Y, Kamitani W. 2017. Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication. Virology 510:165–174. https://doi.org/10.1016/j.virol.2017.07.019
18. Oudshoorn D, Rijs K, Limpens R, Groen K, Koster AJ, Snijder EJ, Kikkert M, Bárcena M. 2017. 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 8:e01658-17. https://doi.org/10.1128/mBio.01658-17
19. Posthuma CC, Pedersen KW, Lu Z, Joosten RG, Roos N, Zevenhoven-Dobbe JC, Snijder EJ. 2008. Formation of the arterivirus replication/transcription complex: a key role for nonstructural protein 3 in the remodeling of intracellular membranes. J Virol 82:4480–4491. https://doi.org/10.1128/JVI.02756-07
20. Lei J, Kusov Y, Hilgenfeld R. 2018. Nsp3 of coronaviruses: structures and functions of a large multi-domain protein. Antiviral Res 149:58–74. https://doi.org/10.1016/j.antiviral.2017.11.001
21. Hurst KR, Koetzner CA, Masters PS. 2013. Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural protein 3 of the viral replicase-transcriptase complex. J Virol 87:9159–9172. https://doi.org/10.1128/JVI.01275-13
22. Koetzner CA, Hurst-Hess KR, Kuo L, Masters PS. 2022. Analysis of a crucial interaction between the coronavirus nucleocapsid protein and the major membrane-bound subunit of the viral replicase-transcriptase complex. Virology 567:1–14. https://doi.org/10.1016/j.virol.2021.12.004
23. Hagemeijer MC, Ulasli M, Vonk AM, Reggiori F, Rottier PJM, de Haan CAM. 2011. Mobility and interactions of coronavirus nonstructural protein 4. J Virol 85:4572–4577. https://doi.org/10.1128/JVI.00042-11
24. Kanjanahaluethai A, Chen Z, Jukneliene D, Baker SC. 2007. Membrane topology of murine coronavirus replicase nonstructural protein 3. Virology 361:391–401. https://doi.org/10.1016/j.virol.2006.12.009
25. Klatte N, Shields DC, Agoni C. 2022. Modelling the transitioning of SARS-CoV-2 nsp3 and nsp4 lumenal regions towards a more stable state on complex formation. Int J Mol Sci 24:720. https://doi.org/10.3390/ijms24010720
26. Ricciardi S, Guarino AM, Giaquinto L, Polishchuk EV, Santoro M, Di Tullio G, Wilson C, Panariello F, Soares VC, Dias SSG, Santos JC, Souza TML, Fusco G, Viscardi M, Brandi S, Bozza PT, Polishchuk RS, Venditti R, De Matteis MA. 2022. The role of NSP6 in the biogenesis of the SARS-CoV-2 replication organelle. Nature 606:761–768. https://doi.org/10.1038/s41586-022-04835-6
27. Sun X, Yu J, Wong SH, Chan MTV, Zhang L, Wu WKK. 2022. SARS-CoV-2 targets the lysosome to mediate airway inflammatory cell death. Autophagy 18:2246–2248. https://doi.org/10.1080/15548627.2021.2021496
28. Wolff G, Limpens R, Zevenhoven-Dobbe JC, Laugks U, Zheng S, de Jong AWM, Koning RI, Agard DA, Grünewald K, Koster AJ, Snijder EJ, Bárcena M. 2020. A molecular pore spans the double membrane of the coronavirus replication organelle. Science 369:1395–1398. https://doi.org/10.1126/science.abd3629
29. Zimmermann L, Zhao X, Makroczyova J, Wachsmuth-Melm M, Prasad V, Hensel Z, Bartenschlager R, Chlanda P. 2023. SARS-CoV-2 nsp3 and nsp4 are minimal constituents of a pore spanning replication organelle. Nat Commun 14:7894. https://doi.org/10.1038/s41467-023-43666-5
30. Yan W, Zheng Y, Zeng X, He B, Cheng W. 2022. Structural biology of SARS-CoV-2: open the door for novel therapies. Signal Transduct Target Ther 7:26. https://doi.org/10.1038/s41392-022-00884-5
31. Harbers M. 2014. Wheat germ systems for cell-free protein expression. FEBS Lett 588:2762–2773. https://doi.org/10.1016/j.febslet.2014.05.061
32. Fogeron M-L, Lecoq L, Cole L, Harbers M, Böckmann A. 2021. Easy synthesis of complex biomolecular assemblies: wheat germ cell-free protein expression in structural biology. Front Mol Biosci 8:639587. https://doi.org/10.3389/fmolb.2021.639587
33. Habersetzer J, Debbah M, Fogeron M-L, Böckmann A, Bressanelli S, Fieulaine S. 2020. *In vitro* translation of virally-encoded replication polyproteins to recapitulate polyprotein maturation processes. Protein Expr Purif 175:105694. https://doi.org/10.1016/j.pep.2020.105694
34. Altincekic N, Korn SM, Qureshi NS, Dujardin M, Ninot-Pedrosa M, Abele R, Abi Saad MJ, Alfano C, Almeida FCL, Alshamleh I, et al. 2021. Large-scale recombinant production of the SARS-CoV-2 proteome for high-throughput and structural biology applications. Front Mol Biosci 8:653148. https://doi.org/10.3389/fmolb.2021.653148
35. Orwick-Rydmark M, Arnold T, Linke D. 2016. The use of detergents to purify membrane proteins. Curr Protoc Protein Sci 84:4. https://doi.org/10.1002/0471140864.ps0408s84
36. Rajesh S, Knowles T, Overduin M. 2011. Production of membrane proteins without cells or detergents. N Biotechnol 28:250–254. https://doi.org/10.1016/j.nbt.2010.07.011
37. Fogeron M-L, Badillo A, Jirasko V, Gouttenoire J, Paul D, Lancien L, Moradpour D, Bartenschlager R, Meier BH, Penin F, Böckmann A. 2015. Wheat germ cell-free expression: two detergents with a low critical micelle concentration allow for production of soluble HCV membrane proteins. Protein Expr Purif 105:39–46. https://doi.org/10.1016/j.pep.2014.10.003
38. VanAken T, Foxall-VanAken S, Castleman S, Ferguson-Miller S. 1986. Alkyl glycoside detergents: synthesis and applications to the study of membrane proteins. Methods Enzymol 125:27–35. https://doi.org/10.1016/s0076-6879(86)25005-3
39. Cho KH, Husri M, Amin A, Gotfryd K, Lee HJ, Go J, Kim JW, Loland CJ, Guan L, Byrne B, Chae PS. 2015. Maltose neopentyl glycol-3 (MNG-3)

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.

analogues for membrane protein study. Analyst 140:3157–3163. https://doi.org/10.1039/c5an00240k
40. Chae PS, Rasmussen SGF, Rana RR, Gotfryd K, Kruse AC, Manglik A, Cho KH, Nurva S, Gether U, Guan L, Loland CJ, Byrne B, Kobilka BK, Gellman SH. 2012. A new class of amphiphiles bearing rigid hydrophobic groups for solubilization and stabilization of membrane proteins. Chemistry 18:9485–9490. https://doi.org/10.1002/chem.201200069
41. La Verde V, Dominici P, Astegno A. 2017. Determination of hydrodynamic radius of proteins by size exclusion chromatography. Bio Protoc 7:e2230. https://doi.org/10.21769/BioProtoc.2230
42. Kampjut D, Steiner J, Sazanov LA. 2021. Cryo-EM grid optimization for membrane proteins. iScience 24:102139. https://doi.org/10.1016/j.isci.2021.102139
43. Denisov IG, Sligar SG. 2016. Nanodiscs for structural and functional studies of membrane proteins. Nat Struct Mol Biol 23:481–486. https://doi.org/10.1038/nsmb.3195
44. Salvador D, Glavier M, Schoehn G, Phan G, Taveau JC, Decossas M, Lecomte S, Mongrand S, Garnier C, Broutin I, Daury L, Lambert O. 2019. Minimal nanodisc without exogenous lipids for stabilizing membrane proteins in detergent-free buffer. Biochim Biophys Acta Biomembr 1861:852–860. https://doi.org/10.1016/j.bbamem.2019.01.013
45. Sligar SG, Denisov IG. 2021. Nanodiscs: a toolkit for membrane protein science. Protein Sci 30:297–315. https://doi.org/10.1002/pro.3994
46. Bayburt TH, Grinkova YV, Sligar SG. 2006. Assembly of single bacteriorhodopsin trimers in bilayer nanodiscs. Arch Biochem Biophys 450:215–222. https://doi.org/10.1016/j.abb.2006.03.013
47. Ritchie TK, Grinkova YV, Bayburt TH, Denisov IG, Zolnerciks JK, Atkins WM, Sligar SG. 2009. Chapter 11 - Reconstitution of membrane proteins in phospholipid bilayer nanodiscs. Methods Enzymol 464:211–231. https://doi.org/10.1016/S0076-6879(09)64011-8
48. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589. https://doi.org/10.1038/s41586-021-03819-2
49. Laskowski RA, Thornton JM. 2022. PDBsum extras: SARS-CoV-2 and AlphaFold models. Protein Sci 31:283–289. https://doi.org/10.1002/pro.4238
50. Mariani V, Biasini M, Barbato A, Schwede T. 2013. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics 29:2722–2728. https://doi.org/10.1093/bioinformatics/btt473
51. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, et al. 2022. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res 50:D439–D444. https://doi.org/10.1093/nar/gkab1061
52. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Žídek A, Bates R, Blackwell S, Yim J, Ronneberger O, Bodenstein S, Zielinski M, Bridgland A, Potapenko A, Cowie A, Tunyasuvunakool K, Jain R, Clancy E, Kohli P, Jumper J, Hassabis D. 2021. Protein complex prediction with AlphaFold-Multimer. Bioinformatics. https://doi.org/10.1101/2021.10.04.463034
53. Holm L, Laiho A, Törönen P, Salgado M. 2023. DALI shines a light on remote homologs: one hundred discoveries. Protein Sci 32:e4519. https://doi.org/10.1002/pro.4519
54. Hallgren J, Tsirigos KD, Pedersen MD, Almagro Armenteros JJ, Marcatili P, Nielsen H, Krogh A, Winther O. 2022. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. Bioinformatics. https://doi.org/10.1101/2022.04.08.487609
55. Gautier R, Douguet D, Antonny B, Drin G. 2008. HELIQUEST: a web server to screen sequences with specific α-helical properties. Bioinformatics 24:2101–2102. https://doi.org/10.1093/bioinformatics/btn392
56. Whitley P, Grau B, Gumbart JC, Martínez-Gil L, Mingarro I. 2021. Folding and insertion of transmembrane helices at the ER. Int J Mol Sci 22:12778. https://doi.org/10.3390/ijms222312778
57. Wang Y, Park SH, Tian Y, Opella SJ. 2013. Impact of histidine residues on the transmembrane helices of viroporins. Mol Membr Biol 30:360–369. https://doi.org/10.3109/09687688.2013.842657
58. Ren Z, Ren PX, Balusu R, Yang X. 2016. Transmembrane helices tilt, bend, slide, torque, and unwind between functional states of rhodopsin. Sci Rep 6:34129. https://doi.org/10.1038/srep34129
59. Shelar A, Bansal M. 2016. Helix perturbations in membrane proteins assist in inter-helical interactions and optimal helix positioning in the bilayer. Biochim Biophys Acta 1858:2804–2817. https://doi.org/10.1016/j.bbamem.2016.08.003
60. Bruss V. 2007. Hepatitis B virus morphogenesis. World J Gastroenterol 13:65–73. https://doi.org/10.3748/wjg.v13.i1.65
61. Cocquerel L, Op de Beeck A, Lambot M, Roussel J, Delgrange D, Pillez A, Wychowski C, Penin F, Dubuisson J. 2002. Topological changes in the transmembrane domains of hepatitis C virus envelope glycoproteins. EMBO J 21:2893–2902. https://doi.org/10.1093/emboj/cdf295
62. Sawasaki T, Hasegawa Y, Tsuchimochi M, Kamura N, Ogasawara T, Kuroita T, Endo Y. 2002. A bilayer cell-free protein synthesis system for high-throughput screening of gene products. FEBS Lett 514:102–105. https://doi.org/10.1016/s0014-5793(02)02329-3
63. Bayburt TH, Sligar SG. 2003. Self-assembly of single integral membrane proteins into soluble nanoscale phospholipid bilayers. Protein Sci 12:2476–2481. https://doi.org/10.1110/ps.03267503
64. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612. https://doi.org/10.1002/jcc.20084
65. Case DA, Belfon K, Ben-Shalom IY, BrozellSR, Cheatham DS, Cruzeiro TE, Darden VWD, Duke TA, Giambasu RE, Gilson G, et al. 2020. Amber 2020. University of California, San Francisco.
66. Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. J Mol Graph 14:33–38. https://doi.org/10.1016/0263-7855(96)00018-5
67. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. 2021. UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci 30:70–82. https://doi.org/10.1002/pro.3943

Downloaded from https://journals.asm.org/journal/jvi on 26 March 2024 by 132.166.53.231.