

Mapping the non-structural transmembrane proteins of SARS-CoV-2

Sunil Thomas

Lankenau Institute for Medical Research, Wynnewood, PA-19096, USA.

E-mail: suntom2@gmail.com

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) responsible for the disease COVID-19 has wreaked havoc on the health and economy of humanity. In addition, the disease is observed in domestic and wild animals. The disease has impacted directly and indirectly every corner of the planet. Currently, there are no effective therapies for the treatment of COVID-19. Vaccination to protect against COVID-19 started in December 2020. SARS-CoV-2 is an enveloped virus with a single-stranded RNA genome of 29.8 kb. More than two-thirds of the genome comprises *Orf1ab* encoding 16 non-structural proteins (nsps) followed by mRNAs encoding structural proteins, spike (S), envelop (E), membrane (M), and nucleocapsid (N). These genes are interspaced with several accessory genes (open reading frames [Orf] 3a, 3b, 6, 7a, 7b, 8, 9b, 9c and 10). The functions of these proteins are of particular interest for understanding the pathogenesis of SARS-CoV-2. Several of the nsps (nsp3, nsp4, nsp6) and Orf3a are transmembrane proteins involved in regulating the host immunity, modifying host cell organelles for viral replication and escape and hence considered drug targets. In this paper we report mapping the transmembrane structure of the non-structural proteins of SARS-CoV-2.

Keywords: SARS-CoV-2, COVID-19, coronavirus, non-structural proteins, nsps, nsp3, nsp4, nsp6, Orf1ab, Orf3a, endoplasmic reticulum, double membrane vesicle.

INTRODUCTION

The coronavirus-19 disease (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in late 2019 is responsible for the pandemic and is a major public health concern. The disease has resulted in mortality, massive hospitalizations, lockouts, economic loss, employment layoffs, school and college closings, and have impacted all the countries. The SARS-CoV-2 virus not only infects humans, but also companion and wild animals. Even after a year, there are no effective therapies for the treatment of the disease (Thomas, 2020). Current effective strategies to decrease the incidence of the disease is social distancing, use of masks and lockdowns in seriously affected areas. The first vaccines to protect against COVID-19 were rolled out in December 2020. All the vaccines currently available are based on the spike protein (S). However, there is a risk that the vaccines may not be effective against mutant strains of the virus. Recent reports suggest that the AstraZeneca vaccine do not provide complete protection against the South African strain (501Y.V2) of SARS-CoV-2.

People with diabetes are at risk of the disease. As yet we do not know why the virus is highly successful in causing the pandemic within 3 months of its first report (Thomas, 2020). Understanding the structure and function of the proteins of SARS-CoV-2 will lead to the development of effective vaccines and drugs to protect against the virus.

The structural proteins of SARS-CoV-2 include, membrane glycoprotein (M), envelope protein (E), nucleocapsid protein (N) and the spike protein (S). SARS-CoV-2 contains a 29.8-kb single-stranded RNA genome wrapped in a helical nucleocapsid composed of multiple copies of N protein, which in turn is surrounded by an envelope containing S glycoprotein, M glycoprotein, and a small E protein. The viral gene order is similar to that in other known coronaviruses, with the first two open reading frames (1a and 1b) encoding the viral replicase and the downstream mRNAs encoding the structural proteins. Using bioinformatics, we previously showed that the M proteins of SARS-CoV-2 resembles the sugar transporters that may be involved in functions favorable to the virus (Thomas, 2020). The non-structural proteins (nsps) of SARS-CoV-2 are involved in inhibiting innate immunity and also induce virus replication. Two overlapping Orfs, Orf1a and Orf1b, are

translated from the positive-strand genomic RNA and generate continuous polypeptides, which are cleaved into a total of 16 nsps. The genes coding for the structural proteins are interspaced with several accessory genes including the open reading frames (Orf) 3a, 3b, 6, 7a, 7b, 8, 9b, 9c, 10). The functions of these proteins are of particular interest for understanding the pathogenesis of SARS-CoV-2 (Yoshimoto, 2020; Gordon et al. 2020).

To evade detection by host innate immune sensors, viruses that replicate in the cytoplasm compartmentalize their genome transcription in organelle-like structures, thereby protecting the virus against host cell defenses and increasing the replication efficiency (Belov and van Kuppeveld, 2012). The transmembrane nsp3, nsp4 and nsp6 are known to rearrange endoplasmic reticulum (ER) membranes thereby inducing curvature of the ER membrane, essential for virus replication. Lack of antiviral therapies is the paucity of knowledge regarding the β -coronavirus-host cell interface (Ghosh et al. 2020). In this paper we provide information on the transmembrane and lumen domains of the nsps of SARS-CoV-2.

MATERIALS AND METHODS

SARS-CoV-2 transmembrane non-structural protein structure

The transmembrane non-structural protein sequences of the SARS-CoV-2 were downloaded from the NCBI (<https://www.ncbi.nlm.nih.gov/protein/>) protein database. The transmembrane non-structural proteins of SARS-COV-2 include nsp3 (Accession No. YP_009725299), nsp4 (Accession No. YP_009725300), nsp6 (Accession No. YP_009725302) and Orf3a (Accession No. BCI50534).

Protein modeling

For a comprehensive understanding of biological systems of how proteins complexes and networks operate a detailed description of the interactions of proteins and the overall quaternary structure is essential. Residue-based diagrams of proteins, also called snake diagrams or protein plots, are 2-D representations of a protein sequence that contain information about properties such as secondary structure (Skrabanek et al. 2003). To

determine a snake diagram model of a protein we used Protter (<http://wlab.ethz.ch/protter>). Protter is an interactive and customizable web-based application that enables the integration and visualization of both annotated and predicted protein sequence features together with experimental proteomic evidence for peptides and posttranslational modifications onto the transmembrane topology of a protein. It allows users to choose from numerous annotation sources, integrate their own proteomics data files, select the best-suited peptides for targeted quantitative proteomics applications, and export publication-quality illustrations (Omasits et al. 2014).

For the three dimensional homology modeling we employed the iterative threading assembly refinement (I-TASSER) (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) with default settings. The protein sequences of SARS-CoV-2 was entered in FASTA format.

We used multiple bioinformatics software for the prediction of transmembrane regions of membrane proteins. SOSUI discriminates membrane and soluble proteins and predicts transmembrane proteins based on the nature of the amino acids (Hirakawa et al. 1998) (https://harrier.nagahama-i-bio.ac.jp/sosui/sosui_submit.html). TMHMM (transmembrane hidden Markov model) is a method for prediction of transmembrane helices based on a hidden Markov model (Krogh et al. 2001) (<http://www.cbs.dtu.dk/services/TMHMM/>). Subcellular Localization Predictive System (CELLO) is a multi-class support vector machines (SVM) classification system. CELLO uses 4 types of sequence coding schemes: the amino acid composition, the di-peptide composition, the partitioned amino acid composition and the sequence composition based on the physico-chemical properties of amino acids. CELLO is a simple, straight forward implementation of a single module (SVM) based on multiple *n*-peptide composition to predict subcellular localization (Yu et al. 2004) (<http://cello.life.nctu.edu.tw/>).

RESULTS

As yet there are no effective therapies and vaccines for COVID-19. To meet the increasing demand for the treatment of COVID-19 there is a need to accelerate novel antiviral drug development as quickly as possible. Target-based drug development may be a promising approach to achieve this goal (Liu et al. 2021). Identifying molecular targets could lead to development of medications that protect against SARS-CoV-2 virus.

Coronaviruses, including SARS-CoV-2 that replicate in the cytoplasm compartmentalize their genome transcription in organelle-like structures thereby protecting the virus against host cell defenses and increasing the replication efficiency (Santerre et al. 2020).

The nsps are critical elements of the replication and transcription complex (RTC), as well as immune system evasion. Through hijacking the endoplasmic reticulum (ER) membrane, nsps help the virus establish the RTC (Santerre et al. 2020). The structure and function of nsps of SARS-CoV-2 is similar to SARS-CoV. In SARS-CoV, the primary structures of the three nsps- nsp3, nsp4, and nsp6, contain hydrophobic stretches, and these proteins are predicted to be integral membrane proteins. Hence, they are likely to function in anchoring the replication complexes to the lipid bilayer (Oostra et al. 2007).

Using bioinformatics, we mapped the transmembrane nsps of SARS-CoV-2. We initially mapped the Orf1ab. The Orf1ab codes for 16 nsps. The Orf1ab is cleaved into 16 nsps. Analysis of the Orf1ab by Protter showed the three transmembrane nsps- nsp3, nsp4, nsp6 (Fig. 1).

We modeled each of the transmembrane nsps. Luminal loops of the transmembrane nsps are essential for rearranging ER membranes thereby inducing curvature of the ER membrane. The nsp3 has two transmembrane domains and a luminal domain in the ER lumen. The nsp3 snake diagram, 3D model and the various domains (cytoplasmic, transmembrane, luminal) are shown in Fig. 2. We used SOSUI, TMHMM, and CELLO bioinformatics software to determine the accuracy of the predicted transmembrane

regions (Fig. 2D-F). The transmembrane regions predicted by all the bioinformatic software were similar confirming the accuracy.

The nsp4 has the largest extra transmembrane domain (loop) among the nsp's and the second largest extra transmembrane domain of SARS-CoV-2 after the spike protein. The nsp4 protein has four transmembrane domains and a large luminal domain in the ER lumen between the first two transmembrane domains. Nsp4 also has a smaller luminal domain in the ER lumen between the third and four transmembrane domains (Fig. 3).

The nsp6 has six transmembrane domains and two small luminal domains between the third and fourth as well as fifth and sixth transmembrane in the ER lumen (Fig. 4).

The ORF3a of SARS-CoV-2 is also a transmembrane protein. The Orf3a has three transmembrane domains and a long and short luminal domain jutting into the ER lumen (Fig. 5).

The transmembrane domains of the nsps of SARS-CoV-2 is responsible for inhibiting the host immunity as well as increasing the replication efficiency of the virus. In this paper we have mapped the transmembrane nsps of SARS-CoV-2 and they could be used as a target to inhibit virus replication.

DISCUSSION

The COVID-19 pandemic caused by the SARS-CoV-2 virus has immobilized the world. It is the most severe pandemic of the twenty first century. As of the first week of December 2020, the virus has infected 63 million people, with 1.5 million deaths world-wide. The disease is more severe in old people compared to children and young adults. Currently there are no therapies and vaccines for the disease; hence, there is an urgent need to develop effective therapies and vaccines for the deadly disease.

The four structural proteins of SARS-CoV-2 are spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins. In addition, the virus consists of a large polyprotein, Orf1ab that proteolytically cleave to form 16 non-structural proteins (nsps). There are also accessory proteins: Orf3a, Orf3b, Orf6, Orf7a, Orf7b, Orf8, Orf9b, Orf9c and Orf10. Although accessory proteins have been viewed as dispensable for viral replication *in vitro*, some have been shown to play an important role in virus-host interactions *in vivo* (Yoshimoto, 2020; Gordon et al. 2020).

The entry of the SARS-CoV-2 into cells starts when the spike glycoprotein expressed on the viral envelope binds to ACE2 receptor of the host cell. The virus enters the cells through endocytosis process, which is possibly facilitated, via a pH dependent endosomal cysteine protease cathepsins. Once inside the cells, SARS-CoV-2 exploits the endogenous transcriptional machinery of host cells to replicate and spread inside the cell. The virus activates or hijack the intracellular pathways of the host in favor of its replication (Sureda et al. 2020).

Knowledge on the structure of the structural proteins of SARS-CoV-2 is essential to the development of vaccines. Most of the vaccines that protect against SARS-CoV-2 developed in the laboratory are based on the S protein (Gu et al. 2020; vanDoremalan et al. 2020; McKay et al. 2020). Several commercial entities are also developing the vaccines based on the Spike mRNA, adenovector or recombinant protein.

In a previous paper we showed the structures of the structural proteins of SARS-CoV-2. *In silico* analysis showed that the M protein of SARS-CoV-2 resembled the prokaryotic sugar transporter SemiSWEET (Thomas, 2020). The nsps are involved in inhibiting host innate immunity, inducing RNA replication and virus exit. They are potential drug targets. However, the transmembrane domains of these nsps are not clearly documented. In this paper we report the domains of the nsps that may be targets for potential drug candidates.

The first nsps encoded by *Orf1a/Orf1ab* are papain-like proteinase (PL proteinase, nsp3) and 3-chymotrypsin-like proteinase (3CLPro protease). The PL proteinase nsp3 cleaves

nsp3, nsp4, and nsp6 form a complex to induce double-membrane vesicles (DMV) (Rohaim et al. 2020).

Replication and transcription of the virus happen within a replication/transcription complex (RTC) encoded by the virus with nsps as primary constituents. Nearly all RTC elements are encoded by the large replicase gene that consists of Orf1a and Orf1ab. The early formation of the RTC is an essential step in the SARS-CoV-2 life cycle to safeguard viral genome replication and to synthesize subgenomic mRNA (Santerre et al. 2020).

SARS-CoV-2 is a positive stranded RNA virus. The virus replication is supported by the transformed endoplasmic reticulum (ER) membranes of the host cell. The virus RNA synthesis is mediated in the double membrane vesicles (DMV) that acts as central hubs, offer a favorable microenvironment and my protect against innate immune sensors activated by dsRNA. The newly synthesized viral RNA is exported to the cytosol through the molecular pore spanning the DMVs. The components of the pore include nsp3, nsp4 and nsp6. Finally, the nucleocapsid (N) protein package the genome of the virus (Wolff et al. 2020). In this paper we map the ER luminal domain of nsp3, nsp4 and nsp6 that may be involved in DMV formation.

Interferons (IFNs) are cytokines with strong antiviral activities and is the first line of defense against invading pathogens. Multiple nsps are involved in inhibiting IFN-I production. The nsp6 binds TANK binding kinase 1 (TBK1) to suppress interferon regulatory factor 3 (IRF3) phosphorylation, nsp13 binds and blocks TBK1 phosphorylation, and Orf6 binds importin Karyopherin α 2 (KPNA2) to inhibit IRF3 nuclear translocation. SARS-CoV-2 nsp1 and nsp6 suppress IFN-I signaling more efficiently than SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV) (Xia et al. 2020).

The Orf3a protein is expressed abundantly in infected and transfected cells, which localizes to intracellular and plasma membranes. (Hassan et al. 2020). ORF3a induce apoptosis of cells mediated through caspase 3 (Ren et al. 2020). Issa et al. (2020) identified six functional domains (I to VI) in the SARS-CoV-2 Orf3a protein. The functional domains were linked to virulence, infectivity, ion channel formation, and virus release. Orf 3a may also be involved in vesicle trafficking (Gordon et al. 2020). Based on our analysis, the Orf3a has three transmembrane domains and a long and short luminal domain jutting into the ER lumen.

Overall, this paper maps the structure of the nsps that modify the ER to DMVs so as to induce replication and further exit the host cell. Targeting the nsp transmembrane domains will lead to the development of therapies that treat COVID-19.

REFERENCES

- Belov GA, van Kuppeveld FJ. (+)RNA viruses rewire cellular pathways to build replication organelles. *Curr Opin Virol.* 2012 Dec;2(6):740-7. doi: 10.1016/j.coviro.2012.09.006.
- Ghosh S, Dellibovi-Ragheb TA, Kerviel A, Pak E, Qiu Q, Fisher M, Takvorian PM, Bleck C, Hsu VW, Fehr AR, Perlman S, Achar SR, Straus MR, Whittaker GR, de Haan CAM, Kehrl J, Altan-Bonnet G, Altan-Bonnet N. β -Coronaviruses use lysosomes for egress instead of the biosynthetic secretory pathway. *Cell.* 2020 Oct 27:S0092-8674(20)31446-X. doi: 10.1016/j.cell.2020.10.039.
- Gordon DE, Jang GM, Bouhaddou M, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature.* 2020 Jul;583(7816):459-468. doi: 10.1038/s41586-020-2286-9.
- Gu H, Chen Q, Yang G, He L, Fan H, Deng YQ, Wang Y, Teng Y, Zhao Z, Cui Y, Li Y, Li XF, Li J, Zhang NN, Yang X, Chen S, Guo Y, Zhao G, Wang X, Luo DY, Wang H, Yang X, Li Y, Han G, He Y, Zhou X, Geng S, Sheng X, Jiang S, Sun S, Qin CF, Zhou Y. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science.* 2020 Sep 25;369(6511):1603-1607. doi: 10.1126/science.abc4730.
- Hassan SS, Choudhury PP, Basu P, Jana SS. Molecular conservation and differential mutation on ORF3a gene in Indian SARS-CoV2 genomes. *Genomics.* 2020 Sep;112(5):3226-3237. doi: 10.1016/j.ygeno.2020.06.016.

Hirokawa T, Boon-Chieng S, Mitaku S. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*. 1998;14(4):378-9. doi: 10.1093/bioinformatics/14.4.378.

Issa E, Merhi G, Panossian B, Salloum T, Tokajian S. SARS-CoV-2 and ORF3a: Nonsynonymous Mutations, Functional Domains, and Viral Pathogenesis. *mSystems*. 2020 May 5;5(3):e00266-20. doi: 10.1128/mSystems.00266-20.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*. 2001 Jan 19;305(3):567-80. doi: 10.1006/jmbi.2000.4315.

Liu XH, Zhang X, Lu ZH, Zhu YS, Wang T. Potential molecular targets of nonstructural proteins for the development of antiviral drugs against SARS-CoV-2 infection. *Biomed Pharmacother*. 2020 Nov 17;133:111035. doi: 10.1016/j.biopha.2020.111035.

McKay PF, Hu K, Blakney AK, Samnuan K, Brown JC, Penn R, Zhou J, Bouton CR, Rogers P, Polra K, Lin PJC, Barbosa C, Tam YK, Barclay WS, Shattock RJ. Self-amplifying RNA SARS-CoV-2 lipid nanoparticle vaccine candidate induces high neutralizing antibody titers in mice. *Nat Commun*. 2020 Jul 9;11(1):3523. doi: 10.1038/s41467-020-17409-9.

Omasits U, Ahrens CH, Müller S, Wollscheid B. Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics* 2014; 30:884-6. doi: 10.1093/bioinformatics/btt607.

Ren Y, Shu T, Wu D, Mu J, Wang C, Huang M, Han Y, Zhang XY, Zhou W, Qiu Y, Zhou X. The ORF3a protein of SARS-CoV-2 induces apoptosis in cells. *Cell Mol Immunol*. 2020 Aug;17(8):881-883. doi: 10.1038/s41423-020-0485-9.

Rohaim MA, El Naggat RF, Clayton E, Munir M. Structural and functional insights into non-structural proteins of coronaviruses. *Microb Pathog*. 2020 Nov 23;104641. doi: 10.1016/j.micpath.2020.104641.

Santerre M, Arjona SP, Allen CN, Shcherbik N, Sawaya BE. Why do SARS-CoV-2 NSPs rush to the ER? *J Neurol*. 2020 Sep 1:1–10. doi: 10.1007/s00415-020-10197-8.

Skrabanek L, Campagne F, Weinstein H. Building protein diagrams on the web with the residue-based diagram editor RbDe. *Nucleic Acids Res* 2003; 31: 3856-3858. doi: 10.1093/nar/gkg552.

Sureda A, Alizadeh J, Nabavi SF, Berindan-Neagoe I, Cismaru CA, Jeandet P, Łos MJ, Clementi E, Nabavi SM, Ghavami S. Endoplasmic reticulum as a potential therapeutic target for covid-19 infection management? *Eur J Pharmacol*. 2020 Sep 5;882:173288. doi: 10.1016/j.ejphar.2020.173288.

Thomas S. The Structure of the Membrane Protein of SARS-CoV-2 resembles the sugar transporter SemiSWEET. *Pathog Immun.* 2020 Oct 19;5(1):342-363. doi: 10.20411/pai.v5i1.377.

van Doremalen N, Lambe T, Spencer A, Belij-Rammerstorfer S, Purushotham JN, et al. ChAdOx1 nCoV-19 vaccine prevents SARS-CoV-2 pneumonia in rhesus macaques. *Nature.* 2020 Oct;586(7830):578-582. doi: 10.1038/s41586-020-2608-y.

Wolff G, Limpens RWAL, Zevenhoven-Dobbe JC, Laugks U, Zheng S, de Jong AWM, Koning RI, Agard DA, Grünwald K, Koster AJ, Snijder EJ, Bárcena M. A molecular pore spans the double membrane of the coronavirus replication organelle. *Science.* 2020 Sep 11;369(6509):1395-1398. doi: 10.1126/science.abd3629.

Xia H, Cao Z, Xie X, Zhang X, Chen JY, Wang H, Menachery VD, Rajsbaum R, Shi PY. Evasion of Type I Interferon by SARS-CoV-2. *Cell Rep.* 2020 Oct 6;33(1):108234. doi: 10.1016/j.celrep.2020.108234. Epub 2020 Sep 19.

Yoshimoto FK. The proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19. *Protein J.* 2020 Jun;39(3):198-216. doi: 10.1007/s10930-020-09901-4.

Yu CS, Lin CJ, Hwang JK: Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Science* 2004, 13:1402-1406. doi: 10.1110/ps.03479604.

ACKNOWLEDGEMENTS

The author acknowledges Abraham Thomas Foundation for providing the resources for this work.

CONFLICT OF INTEREST

The author declares no conflict of interest.

FIGURES

Fig.1. The topology of Orf1ab (snake diagram) determined using Protter. The Orf1ab is cleaved into 16 nsps. Analysis of the Orf1ab shows the three transmembrane non-structural proteins-nsp3, nsp4, nsp6. The 12 transmembrane domains of the nsps are shown in figures 2-4.

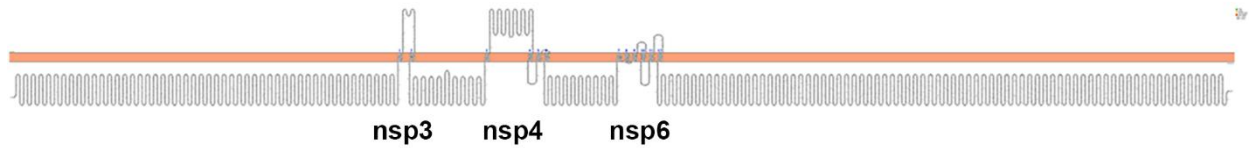
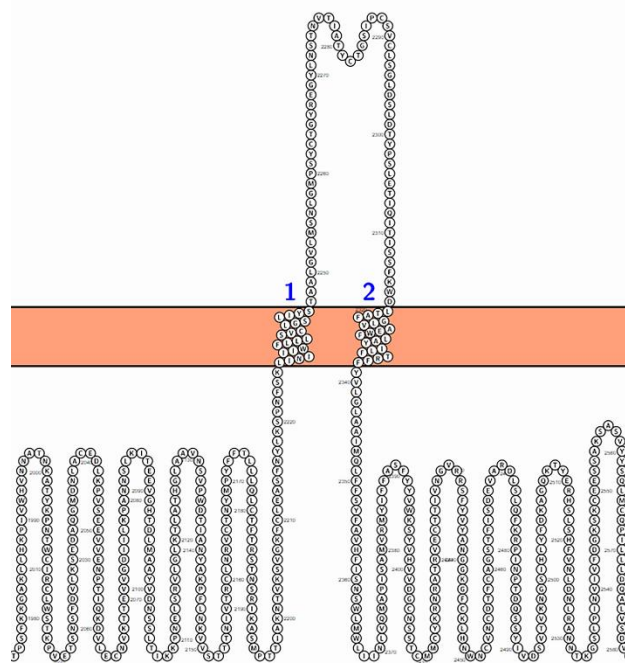


Fig.2. The topology of nsp3 of SARS-CoV-2. (A) Membrane topology of nsp3 determined using Protter. (B) The domains (cytoplasmic, transmembrane, luminal) of nsp3. (C) The predicted nsp3 protein structure of SARS-CoV-2 (ribbon diagram) determined using the software I-TASSER. The amino acid sequence from 1201-1945 of nsp3 is used for modeling the protein. The Orf1ab is cleaved into 16 nsps. The domains (1, 2) of nsp3 denotes the first two transmembrane domain of Orf1ab. (D) Transmembrane helix prediction (TMHMM) of nsp3. (E) SOSUI predicted the membrane has 4 transmembrane helices. (F) CELLO prediction of the nature of the nsp3 protein.

A



B

```

1  aptkvtfgdd tvievqgyks vnitfelder idkvlnekcs aytvelgtev nefacvvada
61 viktqpvse lltpigidld ewsmatyylf desgefklas hmycsfyppd edeeegdcee
121 eefepstqye ygteddyqgk plefgatsaa lqpeeeqeed wldddssqqtv gqddgsednq
181 ttiqtivev qpqlmeltp vvqtievnsf sgylkltdnv yiknadivee akkvkptvvv
241 naanvylkhg ggvagalnka tnnamqvesd dyiatngplk vggscvlsgh nlakhclhvv
301 gpnvnkgedi qlksayenf nqhevllapl lsagifgadp ihslrvcvdt vrtnvylavf
361 dknlydklvs sflemksekk veqkiaepk eevkpfites kpsveqrkqd dkkikacvee
421 vtttleetkf ltenllyid ingnlhpdsa tlvsdiditf lkkdapyivg dvvqegvltv
481 vviptkkagg ttemlakalr kvptdnyitt ypgqglngt veeaktvlkk cksafyilps
541 iisnekqeil gtvswnlrem lahaetrkl mpvcvetkai vstiqrkykg ikiqegvvdv
601 garfyfysk ttvaslintl ndlnetlvtm plgyvthgl leaarymrs lkvpatsvsv

```

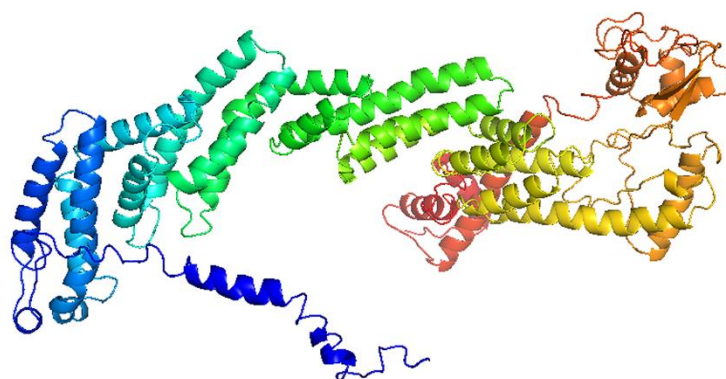
661 spdavtayng yltsssktpe ehfietisla gsykdwsysg qstqlgiefl krgdksvyyt
 721 snpttfhldg evitfdnlkt llsrevrti kvftvdnin lhtqvvdmsm tygqqfgpty
 781 ldgadvtkik phnshegktf yvlpnddlr veafeyyhtt dpsflgryms alnhtkkwky
 841 pqvngltsik wadnncylat alltlqqiel kfnppalqda yyrarageaa nfcailayc
 901 nktvgelgdv retmsylfqh andsckrvl nvvcktcgqq qttlkgveav mymgtlsyeq
 961 fkkgvqipct cgkqatkylv qqespfvmms appaqyelkh gtftcaseyt gnyqcghykh
 1021 itsketlyci dgalltksse ykgpitudfy kensytttik pvtlykldgvv cteidpkldn
 1081 yykkdnsyft eqpidlvpnq pypnasfdnf kfvcndnikfa ddlnqltgyk kpasrelkvt
 1141 ffpdlnqdvv aidykhytps fkkgakllhk pivwhvnnat nkatykpntw circlwstkp
 1201 vetsnsfdvl ksedaqgmdn lacedlkpvs eevenptiq kdvlecnvkt tevgdiiik
 1261 pannslkite evghdlmaa yvdnssltik kpnelsrvlg lktlathgla avnsvpwdti
 1321 anyakpflnk vvstttnivt rclnrvcny mpyffllllq lctftrstns rikasmptti
 1381 akntvksvgk fcleasfnyl kspnfsk lin iiiwflllsv clgsllys ta algvlmsnlg
 1441 mpsyctgyre gylnstnvti atyctgsipc svclsgldsl dtypsletiq itissfkwd l
 1501 tafglvaewf layilftrff yvlglaaimq lffsyfavhf isnswlmwli invlqmapis
 1561 amvrmyiffa sfyyvwksyv hvvdgcnst cmmcyknra trvectivn gvrrsfyvyv
 1621 nggkgfcklh nwnvcndtf cagstfide vardlsqfk rpinptdqs yivdsvtvkn
 1681 gshlyfdka gqktyerhsl shfvnldnr anntkgsipi nvivfdgksc ceessaksas
 1741 vyysqlmcqp illldqalvs dvgsaevav kmfdavvntf sstfnvpmek lktlvataea
 1801 elaknvslnd vlstfisaar qgfvdsdvet kdvveclks hqsdievtgd scnnymlytn
 1861 kvenmtprdl gacidesarh inaqvakshn ialiwvkvdf mslseqlrkq irsaakknnl
 1921 pfkltcattr qvvnvvttki alkkg

Yellow: Luminal domain

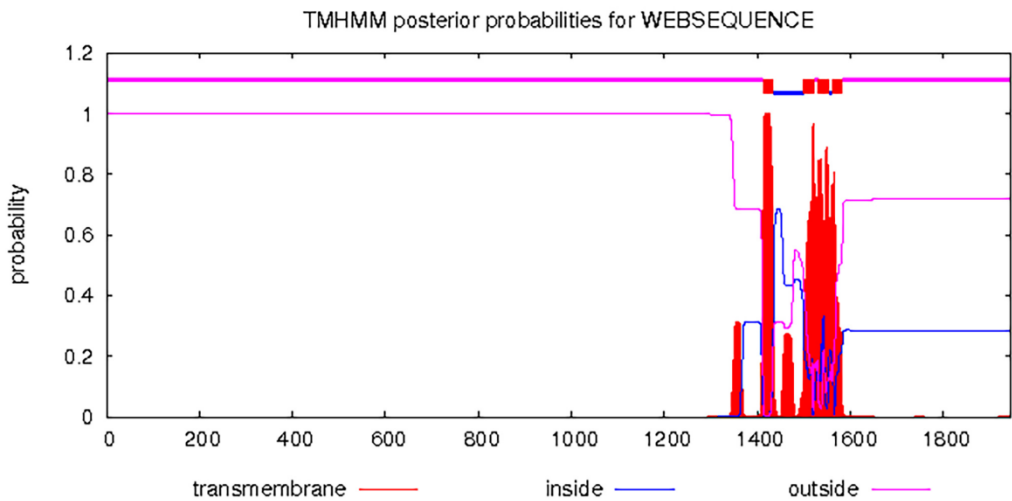
Blue: Transmembrane domain

Green: Cytoplasmic domain

C



D



E. SOSUI prediction: This amino acid sequence is of a MEMBRANE PROTEIN which have 4 transmembrane helices.

No.	N terminal	transmembrane region	C terminal	type	length
1	1414	WFLLLSVCLGSLIYSTAALGVLM	1436	PRIMARY	23
2	1457	NVTIATYCTGSIPCSVCLSG LDS	1479	SECONDARY	23
3	1509	WFLAYILFTRFFYVLGLAAIMQL	1531	PRIMARY	23
4	1551	INLVQMAPISAMVRMYIFFASFY	1573	PRIMARY	23

F. CELLO Prediction

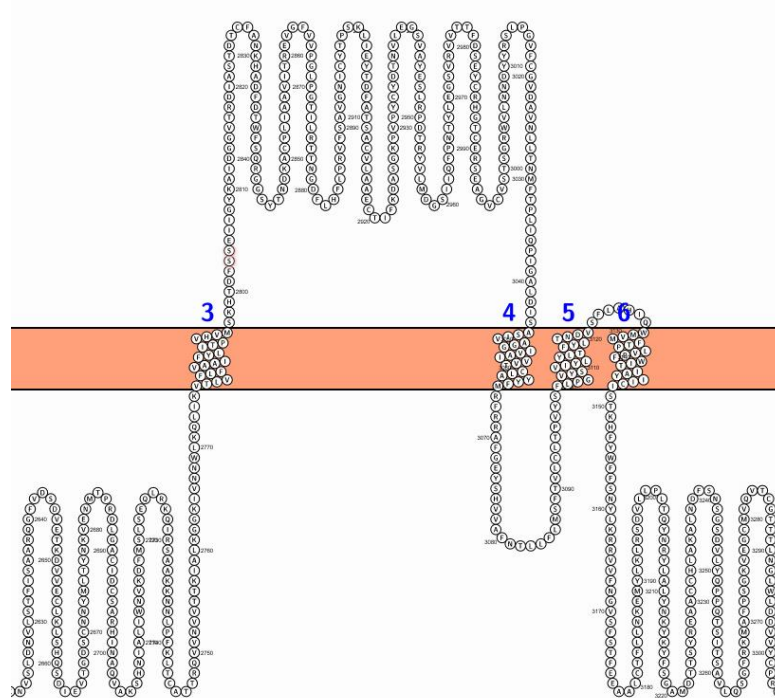
SeqID: YP_009725299.1 nsp3 [Severe acute respiratory syndrome coronavirus 2]

Analysis Report:

Support Vector Machine (SVM)	LOCALIZATION	RELIABILITY
Amino Acid Comp.	OuterMembrane	0.647
N-peptide Comp.	Extracellular	0.411
Partitioned seq. Comp.	OuterMembrane	0.881
Physico-chemical Comp.	InnerMembrane	0.374
Neighboring seq. Comp.	OuterMembrane	0.772
CELLO Prediction:		
	OuterMembrane	2.926 *
	Extracellular	0.889
	InnerMembrane	0.612
	Periplasmic	0.306
	Cytoplasmic	0.265

Fig.3. The topology of nsp4 of SARS-CoV-2. (A) Topology of nsp4 is determined using Protter. (B) The domains (cytoplasmic, transmembrane, luminal) of nsp4. (C) The predicted nsp4 structure of SARS-CoV-2 (ribbon diagram) determined using the software I-TASSER. The Orf1ab is cleaved into 16 nsps. The domains (3-6) of nsp4 denotes the transmembrane domain 3-6 of Orf1ab. (D) Transmembrane helix prediction (TMHMM) of nsp4. (E) SOSUI predicted the membrane has 4 transmembrane helices. (F) CELLO prediction of the nature of the nsp4 protein.

A



B

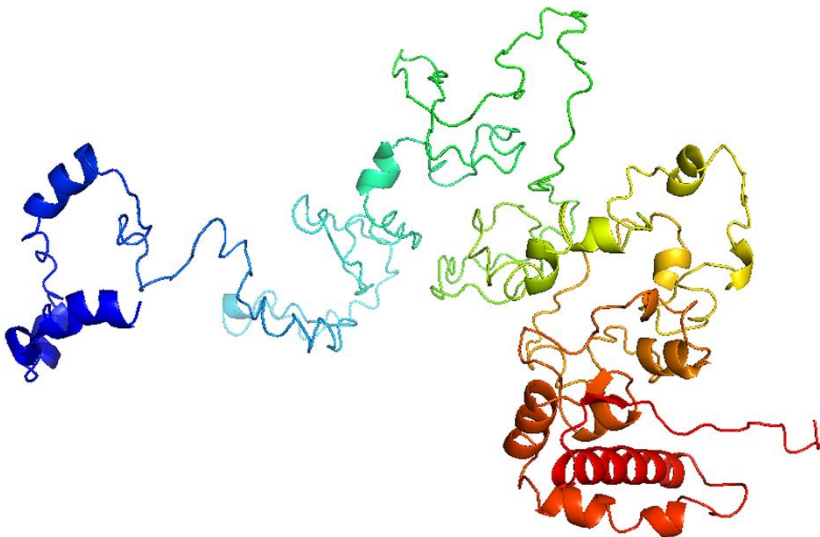
```

1  kivnnwlkql ik vtlvflfv aaifylitpv hvm skhtdfs seiigykaid ggvtrdiast
61 dtcfankhad fdtwfsqrgg syndkacpl iaavitrevg fvpvglpgti lrtngdflh
121 flprvfsavg nicytpskli eytdfatsac vlaaectifk dasgkpvpyc ydtnvlegsv
181 ayeslrpdtr yvlmdgsiiq fpntylegsv rvvttfdsey crhgtcerse agvcvstsgv
241 wvlnndyyrs lpgvfegvda vnlltnmftp liqipigaldi s asivaggiv aivvtclayy
301 fm rfrrafge yshvafntl lfmsftvlc ltpvys flpg vysviylt fyltndv sfl
361 ahicq wvmvmt plvpfwitia yiici stkhf ywffsnylkr rrvfngvsfs tfeaaactf
421 llnkemykl rsdvllpltq ynrylalnky kyfsgamdt tsyreaacch lakalndfsn
481 sgsvdlyqpp qtsitsavlq

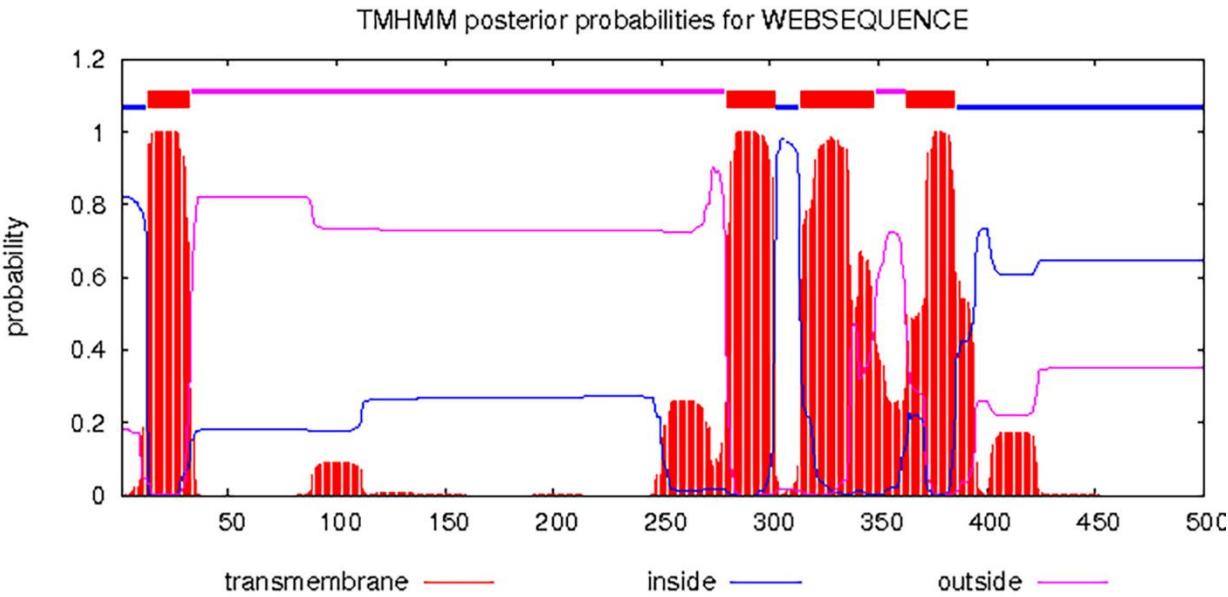
```


Yellow: Luminal domain
Blue: Transmembrane domain
Green: Cytoplasmic domain

C



D



E. SOSUI prediction: This amino acid sequence is of a MEMBRANE PROTEIN which have 4 transmembrane helices.

No.	N terminal	transmembrane region	C terminal	type	length
1	10	LIKVTLVFLFVAAIFYLITPVHV	32	PRIMARY	23
2	280	ISASIVAGGIVAIVVTCLAYYFM	302	PRIMARY	23
3	319	TLLFLMSFTVLCLTPVYSFLPGV	341	PRIMARY	23
4	364	QWMVMFTPLVPFWITIAIICIS	386	PRIMARY	23

F. CELLO Prediction

SeqID: YP_009725300.1 nsp4 [Severe acute respiratory syndrome coronavirus 2]

Analysis Report:

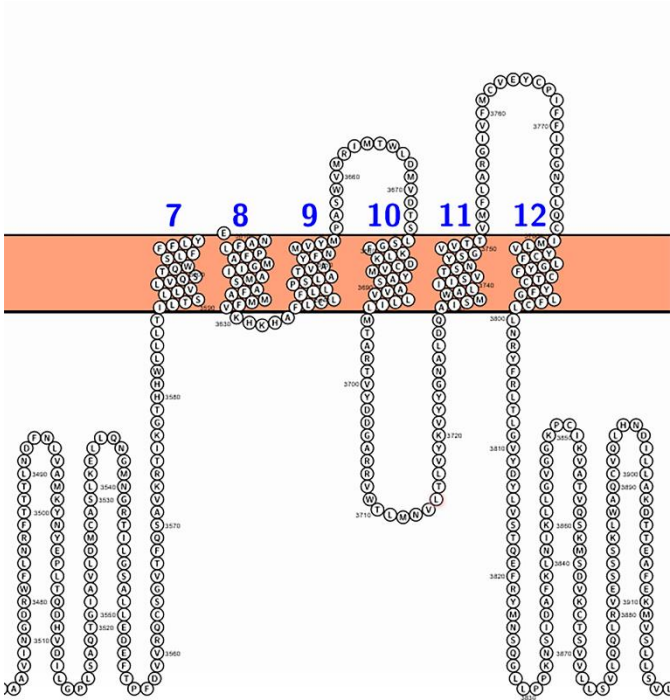
Support Vector Machine (SVM)	LOCALIZATION	RELIABILITY
Amino Acid Comp.	InnerMembrane	0.447
N-peptide Comp.	InnerMembrane	0.521
Partitioned seq. Comp.	InnerMembrane	0.492
Physico-chemical Comp.	OuterMembrane	0.539
Neighboring seq. Comp.	InnerMembrane	0.661

CELLO Prediction:

InnerMembrane	2.421 *
OuterMembrane	1.426
Extracellular	0.565
Cytoplasmic	0.415
Periplasmic	0.173

Fig.4. The topology of nsp6 of SARS-CoV-2. (A) Topology of nsp6 is determined using Protter. (B) The domains (cytoplasmic, transmembrane, ER lumen) of nsp6. (C) The predicted nsp6 protein structure of SARS-CoV-2 (ribbon diagram) determined using the software I-TASSER. The Orf1ab is cleaved into 16 nsps. The domains (7-12) of nsp6 denotes the transmembrane domain 7-12 of Orf1ab. (D) Transmembrane helix prediction (TMHMM) of nsp6. (E) SOSUI predicted the membrane has 7 transmembrane helices. (F) CELLO prediction of the nature of the nsp6 protein.

A

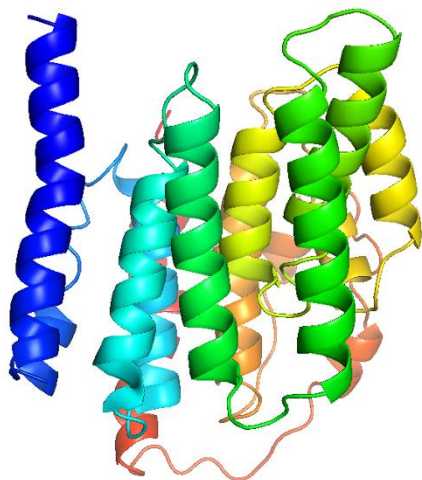


B

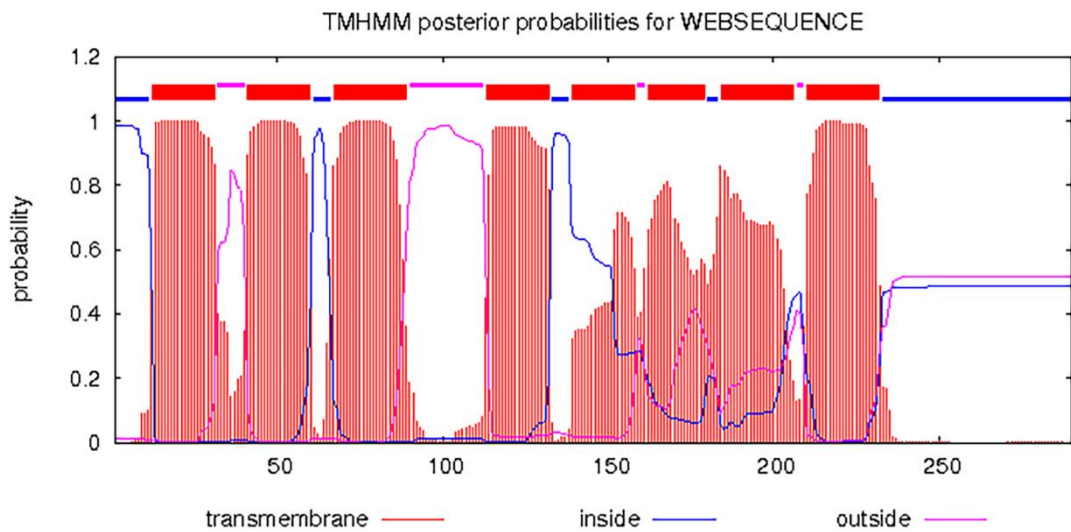
```
1 savkrtikgt hhwlllt ilt sllvlvqstq wslffflyen aflpfamgii amsafammfv
61 khkha flclf llpslatvay fnmvym pasw vmrimtwldm vdtls lsgfkl kdcvmyasav
121 vllil mtart vyddgarrvw tlmnvlty kvyynaldq aismwaliis vtsnysgvvt
181 t vmflargiv fmcveycpif fitgntlqc i mlvycflgyf ctcyfglfc lnrlyfrltlg
241 vydyylvstqe frymnsqgll ppknsidafk lnklkgvgg kpcikvatvq
```

Yellow: Luminal domain
Blue: Transmembrane domain
Green: Cytoplasmic domain

C



D



E. SOSUI prediction: This amino acid sequence is of a MEMBRANE PROTEIN which have 7 transmembrane helices.

No.	N terminal	transmembrane region	C terminal	type	length
1	12	HWLLLTILTSLLVLVQSTQWSLF	34	PRIMARY	23
2	40	NAFLPFAMGIIAMSAFAMMFVK	61	PRIMARY	22
3	65	AFLCLFLLPSLATVAYFNMVYMP	87	PRIMARY	23
4	112	DCVMYASAVVLLILMTARTVYDD	134	PRIMARY	23
5	155	GNALDQAISMWALIISVTSNYSG	177	SECONDARY	23
6	179	VTTVMFLARGIVFMCVEYCPIFF	201	PRIMARY	23
7	213	VYCFLGYFCTCYFGLFCLLNRYF	235	PRIMARY	23

F. CELLO Prediction

SeqID: YP_009725302.1 nsp6 [Severe acute respiratory syndrome coronavirus 2]

Analysis Report:

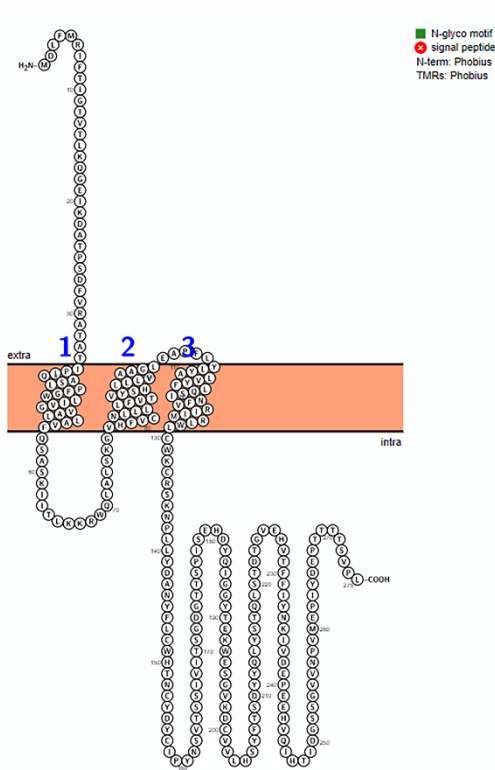
Support Vector Machine (SVM)	LOCALIZATION	RELIABILITY
Amino Acid Comp.	InnerMembrane	0.883
N-peptide Comp.	InnerMembrane	0.990
Partitioned seq. Comp.	InnerMembrane	0.890
Physico-chemical Comp.	InnerMembrane	0.974
Neighboring seq. Comp.	InnerMembrane	0.992

CELLO Prediction:

InnerMembrane	4.729 *
Periplasmic	0.103
Extracellular	0.069
Cytoplasmic	0.060
OuterMembrane	0.040

Fig.5. The topology of Orf3a of SARS-CoV-2. (A) The topology of Orf3a of SARS-CoV determined using Protter. (B) The domains (cytoplasmic, transmembrane, luminal) of Orf3a of SARS-CoV-2. (C) The predicted Orf3a protein structure of SARS-CoV-2 (ribbon diagram) determined using the software I-TASSER. (D) Transmembrane helix prediction (TMHMM) of Orf3a. (E) SOSUI predicted the membrane has 4 transmembrane helices. (F) CELLO prediction of the nature of the Orf3a protein.

A

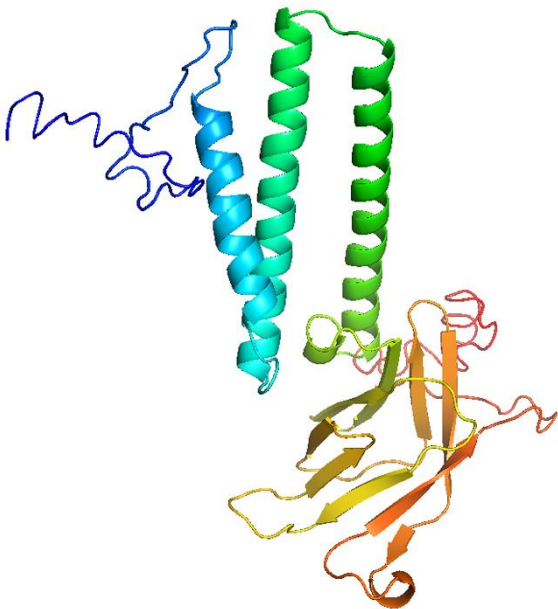


B

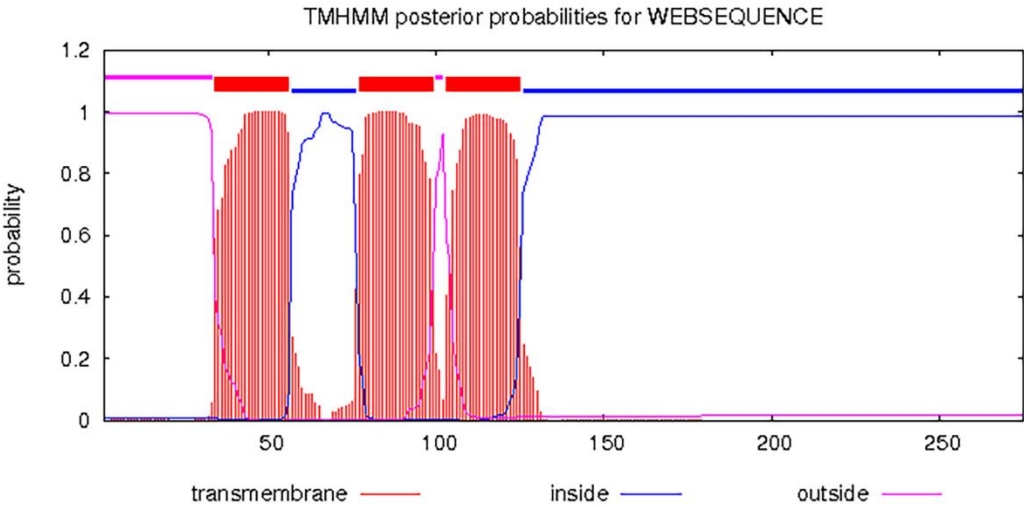
1 mdlfmrifti gvtlqkgei kdatpsdfvr atat ipiqas lpfgwlvigv allavf qsas
61 kiitlkrwq lalskg vhfv cnllllfvtv yshlllvaagl eapfl ylyalvyflqsinf
121 vriimrlwl cwkersknpll ydanyflcwh tncydycipy nsvtssivit sgdgttspis
181 ehdyqiggyt ekcesgvkdc vvlhsyftsd yyqlystqls tdtgvehvtf fiynkivdep
241 eehvqihtid gssgvnpvm epiydepttt tsvpl

Yellow: Luminal domain
Blue: Transmembrane domain
Green: Cytoplasmic domain

C



D



E. SOSUI prediction: This amino acid sequence is of a MEMBRANE PROTEIN which have 4 transmembrane helices.

No.	N terminal	transmembrane region	C terminal	type	length
1	1	MDLFMRIFTIGTVTLKQGEI	20	SECONDARY	20
2	40	SLPFGWLIVGVALLAVFQSASKI	62	PRIMARY	23
3	74	SKGVHFVCNLLLLFVTVYSHLLL	96	PRIMARY	23
4	108	LYALVYFLQSINFVRIIMRLWLC	130	PRIMARY	23

F. CELLO Prediction

SeqID: BCI50534.1 ORF3a [Severe acute respiratory syndrome coronavirus 2]

Analysis Report:

Support Vector Machine (SVM)	LOCALIZATION	RELIABILITY
Amino Acid Comp.	InnerMembrane	0.349
N-peptide Comp.	InnerMembrane	0.361
Partitioned seq. Comp.	InnerMembrane	0.312
Physico-chemical Comp.	Extracellular	0.552
Neighboring seq. Comp.	InnerMembrane	0.467

CELLO Prediction:

InnerMembrane	1.576 *
Extracellular	1.161 *
Cytoplasmic	0.889
OuterMembrane	0.716
Periplasmic	0.657