

Protein Structure Analysis Method of SARS-COV-2 M Protein for Possible Clues Regarding Virion Stability, Longevity and

Saam Hasan^{1, 2, #} and Muhammad Maqsud Hossain^{1, 2, #, *}

¹Department of Biochemistry and Microbiology, North South University, Dhaka, Bangladesh; ²NSU Genome Research Institute-NGRI, North South University, Dhaka, Bangladesh

*For correspondence: Muhammad.maqsud@northsouth.edu

#Contributed equally to this work

[Abstract] The Severe Acute Respiratory Syndrome Coronavirus 2 or SARS-COV-2 has been the cause of a global pandemic in 2020. With the numbers infected rising well above a 1.9 million and confirmed deaths above 122,000 as of 15th April 2020, it has become the paramount health concern for the global community at present. The SARS-COV-2 genome has since been sequenced and its predicted proteome identified. In this study, we looked at the expected SARS-COV-2 proteins and compare them to its close relative, the Severe Acute Respiratory Syndrome-Related Coronavirus. In particular we focussed on the M protein which is known to play a significant role in the virion structure of Coronaviruses. The rationale here was that since the major risk factor associated with SARS-COV-2 was its ease of spread, we wished to focus on the viral structure and architecture to look for clues that may indicate structural stability, thus prolonging the time span for which it can survive free of a host. As a result of the study, we found some rather interesting differences between the M protein for SARS-COV-2 and the SARS-CoV virus M protein. This included amino acid changes from non-polar to polar residues in regions important for anchoring the protein in the envelope membrane.

Keywords: Proteomics, Systems Biology, M protein, SARS-CoV-2, SARS-CoV

[Background] The SARS-COV-2 Novel Corona virus or Severe Acute Respiratory Syndrome Coronavirus 2 has been the cause of a worldwide pandemic in 2020. The novel strain, originating from Wuhan, China has as of 25th April World Health Organization report, infected over 2.7 million people and led to the deaths of over 187,000 individuals worldwide (Coronavirus disease (COVID-2019) situation report-96, 2020). It has led to considerable disruptions in the economic and psychological welfare of populations, on top of the continuously increasing loss of life (Chen *et al.*, 2020). Efforts at identifying possible solutions to combating the virus have thus far ranged from basic precautionary awareness campaigns to the suggestions of nucleotide based drugs currently being tested for other diseases as possible therapeutic agents (Martinez, 2020). The genome of the SARS-COV-2 novel coronavirus has been decoded and its expected proteome already established (RefSeq ID: [NC_045512.2](https://www.ncbi.nlm.nih.gov/RefSeq/record/NC_045512.2)). However up until this point, to the best of our knowledge not much progress has been made in terms of identifying unique molecular features that are exclusive to this strain of the Coronavirus. Hence, in this study we made an attempt at studying the protein sequences of a relevant SARS-COV-2 architectural protein in order to identify unique advantageous features. We chose to focus our efforts on the M Protein as it is

the most abundant protein in the viral nucleocapsid and is believed to be responsible for maintaining the virion in its characteristic shape (Nal, 2005). This coincides with our investigative rationale; is there any particular feature that aids the spread of this strain. The percentage of infected patients who have died from this virus has been estimated to be approximately 6.9% percent as of this writing (WHO, 2020b). Furthermore the primary susceptible groups are the immunocompromised. This leads us to believe that the primary focus needs to be on the mechanism of spreading, as the rapid rates at which this strain has infected communities across the world has arguably been its most dangerous aspect. The coronaviruses are known to contain four main structural proteins, the S, M, E, and N proteins (Fehr and Perlman, 2015). All of whom have been identified in the predicted proteome of SARS-COV-2. The S protein is the spike protein, responsible for gaining entry into the ER following infection (Delmas and Hurst, 2009; Beniac *et al.*, 2006). The N protein resides inside the nucleocapsid and is an RNA binding protein (Chang *et al.*, 2005; Hurst *et al.*, 2009). The M and E proteins are both transmembrane proteins. The M protein, a 3 pass transmembrane protein is particularly significant as it is believed to be responsible for maintaining and giving the virion its shape (Armstrong *et al.*, 1984).

Software

1. NCBI Blast (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)
2. SWISS-MODEL (<https://swissmodel.expasy.org/interactive>)

Procedure

A. Selecting Target Protein

Target protein was selected based on its importance to virion structure. M Protein is the most abundant protein in viral nucleocapsid. It plays a key role in maintenance of virion shape and architecture. Hence it was chosen as the target of analysis based on its unique features may play a crucial role in facilitating the survival and transmission of the virion.

B. Sequence Alignment

1. Refseq M protein amino acid sequences were obtained from NCBI Protein.
2. SARS-CoV-2 and SARS-CoV M protein sequences were aligned using Blastp. All parameters were kept as default.
3. Amino acid substitutions between the two proteins were checked and correlated with their functional significance to the protein.

C. Protein Modelling

3D structures for both proteins were predicted using SWISS-MODEL. It should be noted that reference structures exist for the SARS-CoV M protein but we chose to put it through SWISS-MODEL so we could compare the two proteins after they had been modelled by the same tool, so

as to eliminate any discrepancy resulting from use of different software.

Data analysis

1. We obtained the protein sequences for the SARS Related Coronavirus as well as SARS-COV-2 and did a multiple sequence alignment using NCBI Blast (Altschul *et al.*, 1990). Figure 1 shows the results of the alignment. Afterwards we took the sequences from the both proteins and attempted to predict their structures using SWISS-MODEL (Biasini *et al.*, 2014).

As Figure 1 shows, there were a total of 20 mismatches and 1 gap between the two sequences. What caught our eye most was the fact that there were multiple amino acid substitutions in the transmembrane domains where bulkier non-polar amino acids in the SARS-CoV gene were replaced by more polar or less bulky ones in the SARS-COV-2 protein. In addition, we found a serine inserted into position 4 of the SARS-COV-2 M protein which its SARS Related Coronavirus counterpart did not have.

```

Query: YP_009724393.1 membrane glycoprotein [Severe acute respiratory syndrome coronavirus 2] Query ID: lcl|Query_21953 Length: 222

>NP_828855.1 matrix protein [Severe acute respiratory syndrome-related coronavirus]
Sequence ID: Query_21955 Length: 221
Range 1: 1 to 221

Score:416 bits(1070), Expect:3e-155,
Method:Compositional matrix adjust.,
Identities:201/222(91%), Positives:214/222(96%), Gaps:1/222(0%)

Query 1   MADSNGTITVEELKLLLEQWNLVIGFLFTWICLLQFAYANRRNFLYIIKLI FLWLLWPV 60
          MAD NGTITVEELK+LLEQWNLVIGFLFL WI LLQFAY+NRNRFLYIIKL+FLWLLWPV
Sbjct 1   MAD-NGTITVEELKQLLEQWNLVIGFLFLAWIMLLQFAYSNRNFLYIIKLVFLWLLWPV 59

Query 61  TLACFVLAAYVRINWITGGIAIAMACLVGLMWLSYFIASFRLFARTRSMWSFNPETNILL 120
          TLACFVLAAYVRINW+TGGIAIAMAC+VGLMWLSYF+ASFRLFARTRSMWSFNPETNILL
Sbjct 60  TLACFVLAAYVRINWVTGGIAIAMACIVGLMWLSYFVASRLFARTRSMWSFNPETNILL 119

Query 121 NVPLHGTILTRP LLESELVIGAVILRGHLRIAGHHLGRCDIKDLPKEITVATSRTL SYYK 180
          NVPL GTI+TRPL+ESELVIGAVI+RGHLR+AGH LGRCDIKDLPKEITVATSRTL SYYK
Sbjct 120 NVPLRGTIVTRPLMESELVIGAVIIRGHLRMAGHSLGRCDIKDLPKEITVATSRTL SYYK 179

Query 181 LGASQRVAGDSGFAAYSRYRIGNYKLNTDHSSSDNIALLVQ 222
          LGASQRV DSGFAAY+RYRIGNYKLNTDH+ S+DNIALLVQ
Sbjct 180 LGASQRVGTD SGFAAYNRYRIGNYKLNTDHAGSDNIALLVQ 221
    
```

Figure 1. BLAST Multiple Sequence Protein alignment of SARS-COV-2 M Protein with SARS-CoV M Protein

Serine provides an extra OH group near the N terminal end of the protein which we know makes up its ectodomain that is glycosylated (Nal *et al.*, 2005). Coming over to the transmembrane regions, at position 33 the SARS-COV-2 has a Cysteine, the equivalent position on SARS-CoV is occupied by a Methionine. At position 30, SARS-COV-2 contains a Threonine, the corresponding position on SARS-CoV has an Alanine. Then at position 76 on the second

transmembrane domain, the SARS-CoV protein contained a valine, compared to an Isoleucine on the SARS-COV-2 protein. Starting from the latter, Isoleucine has a bulkier sidechain compared to valine, which in the hydrophobic interior of an envelope membrane could produce more hydrophobic interactions with the fatty acid chains of the lipid bilayer, possibly stabilizing the membrane to a greater degree. Furthermore, this substitution occurs within the Mn2 epitope that binds to cytotoxic T lymphocytes (Liu *et al.*, 2010). The previous couple of changes, the Met to Cys and Ala to The substitutions would seem more curious. Both would replace non-polar sidechains with polar groups. This may seem counterintuitive for stabilizing a membrane structure, however crucially perhaps, this could theoretically allow for inter-chain bonds between protein molecules.

The OH group introduced by Threonine could contribute to hydrogen bonding, while the thiol group introduced by Cysteine could give rise to a disulphide bond. We believe this may hold special importance, especially since our modelling run predicted the SARS-COV-2 M Protein to form homo-trimers, whereas the SARS-CoV M protein is known to form homo-dimers (Godet *et al.*, 1992). The latter was backed up when ran the SWISS-MODEL tool on the amino acid sequence from the SARS-CoV protein for which the first predicted structure was a homo-dimer. Figures 2A and 2B show the predicted structures for SARS-COV-2 and SARS-CoV M Protein respectively.

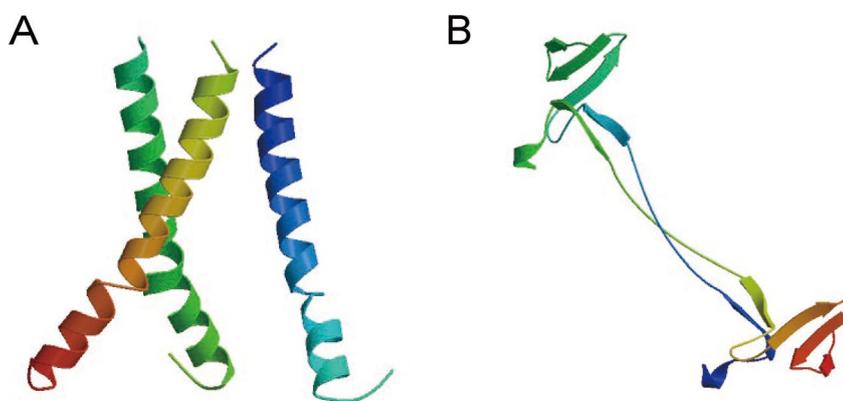


Figure 2A: SARS-COV-2 M Protein predicted 3D structure. **2B:** SARS-CoV M Protein 3D structure.

2. Lastly we looked up known protein domains and epitopes within the SARS Related Coronavirus M Protein and scrutinized for any changes in those regions for the SARS-COV-2 M Protein. The most significant epitope we found and considered was the Mn2 epitope that binds to human Cytotoxic T Lymphocytes (Liu *et al.*, 2010).

Discussion

The goal of this study was to design a simple method for analysing a viral structural protein and understanding the role it may play in the survival and stability of the virion. The two main requirements for using this protocol is A) a structural protein that is known to play a significant role in viral architecture and B) a close relative of the virus that expresses the same protein but is known to be less easily transmittable. If both of these are available, using this method can provide a simple framework for making an initial deduction on the importance of said structural protein. As for the results of this particular study, they lead us to a possible answer to the question we posed at the very start. We believe there might be a possibility of additional bonding interactions in the SARS-COV-2 M protein that allows its structure to remain more stable and survive for longer. In general, it is what one would expect to happen with the introduction of additional protein-protein bonding, as per the usual biochemical make-up of lipid bilayer membranes and transmembrane proteins that pass through them (Alberts, 2008). The Serine at the N terminus is another interesting observation. The N terminal of this protein is expected to be on the outside of the envelope (Nal *et al.*, 2005) and possibly be exposed to the air and outside interactions.

Something that has come up in recent news reports is the ability of the SARS-COV-2 viral particles to survive for longer on metal surfaces, although the actual study has not yet been published (Kumar and Salzman, 2020). Metals have lattice structures where the positive ions are in the middle and the electrons delocalised and free to move about (Atkins and Shriver, n.d.). A polar group such as an –OH where the hydrogen atom would likely have a partial positive charge, may open the possibility for some kind of bio-electrostatic association that could help the virion to adhere to these surfaces. Although we are not sure as to how or if it could preserve the architectural integrity of the virion.

The major difference between the SARS-CoV and the SARS-COV-2 outbreaks is the spread of the latter. SARS-Cov only effected 26 counties, with around 8000 confirmed cases (WHO | SARS (Severe Acute Respiratory Syndrome), 2020). Granted, this needs to factor in the increased air travel in recent times, which, as per Statista is almost double it was in 2003 when the SARS epidemic happened. The information for that can be found here; <https://www.statista.com/statistics/564769/airline-industry-number-of-flights>. Nonetheless, based on evidence and our molecular level analysis, we believe the M protein should be a candidate into future investigations that could shed more light on how the SARS-COV-2 virus survives and functions.

Competing interests

The authors declare no competing interests.

References

1. Chen, Q., Quan, B., Li, X., Gao, G., Zheng, W., Zhang, J., Zhang, Z., Liu, C., Li, L., Wang, C., Zhang, G., Li, J., Dai, Y., Yang, J. and Han, W. (2020). [A report of clinical diagnosis and treatment of nine cases of coronavirus disease 2019](#). *J Med Virol* 92(6): 683-687.
2. Fehr, A. R. and Perlman, S. (2015). [Coronaviruses: An Overview of Their Replication and Pathogenesis](#). In Maier, E., Bickerton, E. and Britton, P. (eds). *Coronaviruses: Methods and Protocols*. vol 1282. Humana Press, New York, NY
3. Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T., Bertoni, M., Bordoli, L. and Schwede, T. (2014). [SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information](#). *Nucleic Acids Res* 42(W1):W252-W258.
4. Armstrong, J., Niemann, H., Smeekens, S., Rottier, P. and Warren, G. (1984). [Sequence and topology of a model intracellular membrane protein, E1 glycoprotein, from a coronavirus](#). *Nature* 308(5961): 751-752.
5. Liu, W., Sun, Y., Qi, J., Chu, F., Wu, H., Gao, F., Li, T., Yan, J. and Gao, G. (2010). [The membrane protein of severe acute respiratory syndrome coronavirus acts as a dominant immunogen revealed by a clustering region of novel functionally and structurally defined cytotoxic t-lymphocyte epitopes](#). *J Infect Dis* 202: 1171-1180.
6. Chang, C.-k., Sue, S.-C., Yu, T.-h., Hsieh, C.-M., Tsai, C.-K., Chiang, Y.-C., Lee, S.-j., Hsiao, H.-h., Wu, W.-J., Chang, W.-L., Lin, C.-H. and Huang, T.-h. (2006). [Modular organization of SARS coronavirus nucleocapsid protein](#). *J Biomed Sci* 13(1): 59-72.
7. Nal, B., Chan, C.-M., Kien, F., Siu, L., Tse, J., Chu, K., Kam, Y.-W., Staropoli, I., Crescenzo-Chaigne, B., Escriou, N., Werf, S., Yuen, K.-Y. and Altmeyer, R. (2005). [Differential maturation and subcellular localization of severe acute respiratory syndrome coronavirus surface proteins S, M and E](#). *J Gen Virol* 86: 1423-1434.
8. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). [Basic local alignment search tool](#). *J Mol Biol* 215(3): 403-410.
9. Beniac, D. R., Andonov, A., Grudeski, E. and Booth, T. F. (2006). [Architecture of the SARS coronavirus prefusion spike](#). *Nat Struct Mol Biol* 13(8): 751-752.
10. Delmas, B. and Laude, H. (1990). [Assembly of coronavirus spike protein into trimers and its role in epitope expression](#). *J Virol* 64(11): 5367-5375.
11. Hurst, K. R., Koetzner, C. A. and Masters, P. S. (2009). [Identification of *in vivo*-interacting domains of the murine coronavirus nucleocapsid protein](#). *Journal of Virology* 83(14): 7221-7234.
12. Martinez, M. A. (2020). [Compounds with therapeutic potential against novel respiratory 2019 coronavirus](#). *Antimicrob Agents Chemother*: AAC.00399-00320.
13. Who.int. 2020a. *WHO | SARS (Severe Acute Respiratory Syndrome)*. [online] Available at: [<https://www.who.int/ith/diseases/sars/en/>](https://www.who.int/ith/diseases/sars/en/) [Accessed 26 April 2020].

14. Who.int. 2020b. *Coronavirus Disease (COVID-2019) Situation Report-96*. [online] Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200425-sitrep-96-covid-19.pdf?sfvrsn=a33836bb_2 [Accessed 26 April 2020].
15. Godet, M., L'Haridon, R., Vautherot, J.-F. and Laude, H. (1992). [TGEV corona virus ORF4 encodes a membrane protein that is incorporated into virions](#). *Virology* 188(2): 666-675.
16. Alberts, B. (2008). [Molecular biology of the cell](#). New York: Garland Science Taylor & Francis.
17. Atkins, P. and Shriver, D., n.d. *Shriver & Atkins Inorganic Chemistry*. p.74.
18. Kumar, V. and Salzman, S. (2020). [SARS-COV-2 can last a few days on surfaces, according to new experiment findings](#). *ABC News*.