

SARS-CoV2 envelope protein: non-synonymous mutations and its consequences

Sk. Sarif Hassan^{a,*}, Pabitra Pal Choudhury^b, Bidyut Roy^c

^aDepartment of Mathematics, Pingla Thana Mahavidyalaya, Maligram 721140, India

^bApplied Statistics Unit, Indian Statistical Institute, Kolkata 700108, West Bengal, India

^cHuman Genetics Unit, Indian Statistical Institute, Kolkata 700108, West Bengal, India

Abstract

In the NCBI database, as on June 6, 2020, total number of available complete genome sequences of SARS-CoV2 across the world is 3617. The envelope protein of SARS-CoV2 possesses several non-synonymous mutations over the transmembrane domain and (C)-terminus in 0.414% of these 3617 genomes. The C-terminus motif DLLV has been changed to DFLV and YLLV in the proteins QJR88103 (Australia: Victoria) and QKI36831 (China: Guangzhou) respectively, which might affect the binding of this motif with the host protein PALS1.

Keywords: Envelope protein; SARS-CoV2; COVID-19; Non-synonymous mutations;

1. Introduction

The present pandemic of the Severe Acute Respiratory Syndrome (COVID-19) is caused by the RNA virus SARS-CoV2 which is characterized by its rapid mutations up to a million times higher than that of their hosts [1]. There are several mutations happened in various proteins of the SARS-CoV2 over a short period of time, which are recently reported in various articles [2, 3, 4]. Genomic variations and evolution enabled the virus to escape host immunity [5, 6]. So, such variability and evolution are enabling the virus to develop drug resistance [1]. Among various proteins of SARS-CoV2, spike(S), envelope (E), membrane(M) and nucleocapsid (N) are the four structural proteins which form the outer layer of the coronavirus and they help in assembling and releasing new copies of the virus within human cell [7].

The CoV envelope (E) protein is the smallest among the four structural proteins involved in several aspects of the virus life cycle, such as assembly, budding, envelope formation, and pathogenesis [7]. However, the molecular mechanism involving E in pathogenesis is not yet clearly understood. Notably, this protein interacts with other structural proteins such as membrane(M), nucleocapsid (N) and other accessory proteins viz. ORF3a, ORF7a and host cell proteins [8]. Envelope protein of SARS-CoV2 is 76 amino acids long and it possesses three important domains viz. (N)-terminus, transmembrane domain (TMD) and (C)-terminus (Fig.1). The (C)-terminal domain of envelope protein in SARS-CoV2 binds to human PALS1, a tight junction-associated protein, which is essential for the establishment and maintenance of epithelial polarity in mammals

*Corresponding author

Email addresses: sarimif@gmail.com (Sk. Sarif Hassan), pabitrupalchoudhury@gmail.com (Pabitra Pal Choudhury), broy@isical.ac.in (Bidyut Roy)

[9].

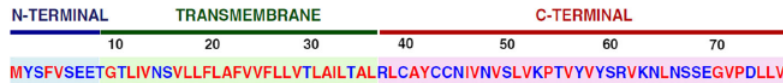


Figure 1: Amino acid sequence and domains of the envelope protein of SARS-CoV2 [7]

Figure 1 Legend: Red and blue colors are representing hydrophobic and hydrophilic amino acid, respectively.

There are four mutations including one deletion have been found in the envelope protein of SARS-CoV2 with reference to that of the SARS-CoV1. The alignment is given in Fig.2.

```

CLUSTAL O(1.2.4) multiple sequence alignment

SARS-CoV1-Envelope      MYSFVSEETGLIVNSVLLFLAFVFLVTLAILTALRLCAYCCNIVNLSLVKPTVYVYS      60
SARS-CoV2-Envelope      MYSFVSEETGLIVNSVLLFLAFVFLVTLAILTALRLCAYCCNIVNLSLVKPSFYVYS      60
                        ***** . ****

SARS-CoV1-Envelope      RVKNLNSSEGVPDLLV          76
SARS-CoV2-Envelope      RVKNLNSR-VPDLLV          75
                        ***** . *****

```

Figure 2: Clustal alignment of the envelope protein of SARS-CoV1 and SARS-CoV2

The Fig.2 legends: Mutations in (C)-terminus domain in the envelope protein of SARS-CoV2 are T55S, V56F, E69R (the mutation of an amino acid A_1 to an amino acid A_2 is denoted by A_1pA_2 where p denotes location in the reference amino acid sequence) and the deletion mutation G at the 70th position with respect to the reference envelope protein of SARS-CoV1.

The envelope protein of SARS-CoV possesses three important domains viz. (N)-terminus, transmembrane domain (TMD) and (C)-terminus. It is reported that the C-terminus domain of the envelope protein contains the motif DLLV (i.e. PDZ domain of the virus E-protein) which binds to the host cell PALS1 protein to facilitate infection [9, 10]. In this present study, non-synonymous mutations in the envelope protein of SARS-CoV2 across 3617 SARS-CoV2 genomes (as on 6th June, 2020), have been identified and accordingly their probable consequences are presented.

In this present study, non-synonymous mutations over the envelope protein of SARS-CoV2 across the available 3617 SARS-CoV2 genomes (as on 6th June, 2020), have been found and accordingly their probable consequences are presented.

2. Methods

From the NCBI virus database, all the protein sequences of 3617 SARS-CoV2 genomes are fetched. Then the amino acid sequences of envelope protein of SARS-CoV2 are exported in fasta format using file oper-

ations through Matlab. These sequences (fasta formatted) are blasted using Clustal-Omega and found the
 40 mismatched and from there mutations and their associated positions are accounted [11].

3. Results

Among these virus genomes from 3617 patients; there are 2588 were from USA, 372 were from Asia, 287
 were from Europe, 365 were from Oceania and 5 were from Africa. Here, we present the non-synonymous
 mutations of the envelope protein over the available 3617 SARS-CoV2 genomes (Table 1). Note that the
 45 mutation of an amino acid A_1 to an amino acid A_2 is denoted by A_1pA_2 where p denotes location in the
 reference amino acid sequence. In the Table 1, also the changes of the R-group of each amino acid according
 to the mutations are presented. The Table-1 depicts the following:

Table 1: Non-synonemous mutation in the envelope protein of SARS-CoV2

Protein-ID	Geo-location	Mutation	Domain	Transition of R-group	Collection date
QJA42107	USA: VA	A36V	TMD*	Hydrophobic to Hydrophobic	2020-04
QJQ84222	USA: KENNER, LA	L26F	TMD	Hydrophobic to Hydrophobic	04/04/20
QHZ00381	South Korea	L37H	TMD	Hydrophobic to Hydrophilic	2020-01
QJS53352	Greece: Athens	L39M	TMD	Hydrophobic to Hydrophobic	20/03/20
QJR88103	Australia: Victoria	L73F	C-terminus	Hydrophobic to Hydrophobic	22/03/20
QKE45838	USA: CA	P71L	C-terminus	Hydrophobic to Hydrophobic	30/04/20
QKE45886	USA: CA	P71L	C-terminus	Hydrophobic to Hydrophobic	30/04/20
QKE45898	USA: CA	P71L	C-terminus	Hydrophobic to Hydrophobic	30/04/20
QKE45910	USA: CA	P71L	C-terminus	Hydrophobic to Hydrophobic	30/04/20
QJE38284	USA: CA	P71L	C-terminus	Hydrophobic to Hydrophobic	28/03/20
QIU81527	USA: WA	P71L	C-terminus	Hydrophobic to Hydrophobic	19/03/20
QKG87268	USA: Massachusetts	S68F	C-terminus	Hydrophobic to Hydrophobic	01/04/20
QKG88576	USA: Massachusetts	S68F	C-terminus	Hydrophobic to Hydrophobic	06/04/20
QKI36831	China: Guangzhou	D72Y	C-terminus	Hydrophilic to Hydrophobic	27/02/20
QKI36855	China: Guangzhou	S68C	C-terminus	Hydrophilic to Hydrophobic	25/02/20

*TMD: transmembrane domain

- In less than 0.5% of the SARS-CoV2 genomes, the envelope protein possesses the missense mutations as
 adumbrated in the Table 1. There are ten different mutations where the R-group property changes only
 50 in the case of the protein QHZ00381 for the mutation L37H in the TMD of the envelope protein.
- TMD was also observed to be conserved over the SARS-CoV1 and COV2 genomes, the protein sequences
 of QJA42107 (USA: VA), QJQ84222(USA: KENNER, LA), QHZ00381(South Korea) and QJS53352(Greece:
 Athens) possess four mutations A36V, L26F, L37H and L39M, respectively, in the TMD of the envelope
 protein. Change in the R-group property from Hydrophobic to Hydrophilic in the virus TMD from
 55 South Korea may affect the function of the envelope protein.
- The motif '*DLLV*' has been changes to '*DFLV*' and '*YLLV*' in the proteins QJR88103 (Australia:
 Victoria) and QKI36831 (China: Guangzhou) due to the mutations L73F and D72Y respectively. These
 mutations having changes in the motif '*DFLV*' might affect in binding this motif to the PALS1 of the
 host and accordingly may influence replication and/or infectivity of the virus.
- In the C-terminus of the envelope protein of SARS-CoV2 the amino acid S at 68th position changes
 60 to the amino acids F and C in the proteins {*QKG87268*, *QKG88576*} from USA: Massachusetts and

QKI36855 from China: Guangzhou respectively. Note that the mutation of the amino acid S to F keeps the R-group unchanged while that of the amino acid S to C changes the R-group Hydrophilic to Hydrophobic. This would possibly make changes in protein functions and interactions.

65 4. Concluding remarks

Among all the proteins associated to the novel RNA virus, some accessory proteins such as ORF6, ORF7b, ORF8, ORF10 contain least number of missense mutation as reported in various articles [12, 13, 14]. And same is true for envelope protein, E. We find 0.414% of the SARS-CoV2 genome contains eight different types of mutations in TMD and C-terminus of the envelope protein. Mutated E-protein might affect replication and propagation of the SARS-CoV2 as has been observed in cases of SARS-CoV and MERS-CoV in mouse model. Potential studies have also shown that vaccine against the E-protein mutated viruses can reduce the infectivity in mouse model.

Author Contributions

SH conceived the problem. SH determined the mutations. SH, PPC, BR analysed the data and result. SH wrote the initial draft which was checked and edited by all other authors to generate the final version.

Conflict of Interests

The authors do not have any conflicts of interest to declare.

References

- [1] M. Pachetti, B. Marini, F. Benedetti, F. Giudici, E. Mauro, P. Storici, C. Masciovecchio, S. Angeletti, M. Ciccozzi, R. C. Gallo, et al., Emerging sars-cov-2 mutation hot spots include a novel rna-dependent-rna polymerase variant, *Journal of Translational Medicine* 18 (2020) 1–9.
- [2] K. Xu, B.-J. Zheng, R. Zeng, W. Lu, Y.-P. Lin, L. Xue, L. Li, L.-L. Yang, C. Xu, J. Dai, et al., Severe acute respiratory syndrome coronavirus accessory protein 9b is a virion-associated protein, *Virology* 388 (2) (2009) 279–285.
- [3] S. S. Hassan, P. P. Choudhury, B. Roy, S. S. Jana, Missense mutations in sars-cov2 genomes from indian patients (May 2020). doi:10.31219/osf.io/2wm8h.
- [4] S. S. Hassan, P. P. Choudhury, P. Basu, S. S. Jana, Molecular conservation and differential mutation on orf3a gene in indian sars-cov2 genomes, *BioRxiv* (2020).
- [5] Y. Li, X. Yang, N. Wang, H. Wang, B. Yin, X. Yang, W. Jiang, The divergence between sars-cov-2 and ratg13 might be overestimated due to the extensive rna modification, *Future Virology* (0) (2020).

- [6] M. L. DeDiego, L. Pewe, E. Alvarez, M. T. Rejas, S. Perlman, L. Enjuanes, Pathogenicity of severe acute respiratory coronavirus deletion mutants in hACE2 transgenic mice, *Virology* 376 (2) (2008) 379–389.
- [7] D. Schoeman, B. C. Fielding, Coronavirus envelope protein: current knowledge, *Virology journal* 16 (1) (2019) 69.
- 95 [8] R. McBride, M. Van Zyl, B. C. Fielding, The coronavirus nucleocapsid is a multifunctional protein, *Viruses* 6 (8) (2014) 2991–3018.
- [9] K.-T. Teoh, Y.-L. Siu, W.-L. Chan, M. A. Schlüter, C.-J. Liu, J. M. Peiris, R. Bruzzone, B. Margolis, B. Nal, The SARS coronavirus E protein interacts with pals1 and alters tight junction formation and epithelial morphogenesis, *Molecular biology of the cell* 21 (22) (2010) 3838–3852.
- 100 [10] T. Beuming, L. Skrabanek, M. Y. Niv, P. Mukherjee, H. Weinstein, Pdbbase: a protein–protein interaction database for PDZ-domains, *Bioinformatics* 21 (6) (2005) 827–828.
- [11] F. Madeira, Y. M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. R. Tivey, S. C. Potter, R. D. Finn, et al., The EMBL-EBI Search and Sequence Analysis Tools APIs in 2019, *Nucleic acids research* 47 (W1) (2019) W636–W641.
- 105 [12] S. Chatterjee, Understanding the nature of variations in structural sequences coding for coronavirus spike, envelope, membrane and nucleocapsid proteins of SARS-CoV-2, *Envelope, Membrane and Nucleocapsid Proteins of SARS-CoV-2* (March 28, 2020) (2020).
- [13] T. H. Tan, T. Barkham, B. C. Fielding, C.-F. Chou, S. Shen, S. G. Lim, W. Hong, Y.-J. Tan, Genetic lesions within the 3a gene of SARS-CoV, *Virology journal* 2 (1) (2005) 51.
- 110 [14] A. Pekosz, S. R. Schaecher, M. S. Diamond, D. H. Fremont, A. C. Sims, R. S. Baric, Structure, expression, and intracellular localization of the SARS-CoV accessory proteins 7a and 7b, in: *The Nidoviruses*, Springer, 2006, pp. 115–120.