Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Genomics 112 (2020) 3890–3892

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

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

Sk. Sarif Hassan[a,*], Pabitra Pal Choudhury[b], Bidyut Roy[c]

[a] Department of Mathematics, Pingla Thana Mahavidyalaya, Maligram, Paschim Medinipur, 721140, West Bengal, India
[b] Applied Statistics Unit, Indian Statistical Institute, Kolkata 700108, West Bengal, India
[c] Human Genetics Unit, Indian Statistical Institute, Kolkata 700108, West Bengal, India

ARTICLE INFO

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

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 (E) protein of SARS-CoV2 possesses several non-synonymous mutations over the transmembrane and C-terminus domains in 15 (0.414%) genomes among 3617 SARS-CoV2 genomes, analyzed. More precisely, 10(0.386%) out of 2588 genomes from the USA, 3(0.806%) from Asia, 1 (0.348%) from Europe and 1 (0.274%) from Oceania contained the missense mutations over the E-protein of SARS-CoV2 genomes. The C-terminus motif DLLV has been to DFLV and YLLV in the proteins from QJR88103 (Australia: Victoria) and QKI36831 (China: Guangzhou) respectively, which might affect the binding of this motif with the host protein PALS1.

1. Introduction

The present pandemic situation 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]. Several mutations have been detected in various proteins of the SARS-CoV2 over a short period of time, which are recently reported in various articles [2–4]. Genomic variations and evolution enabled the virus to escape host immunity [5,6]. So, such variability would help the scientists towards the drug development [1]. Among various proteins of SARS-CoV2, spike(S), envelope (E), membrane(M) and nucleocapsid (N) are the four structural proteins which help them 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-protein in pathogenesis is not yet clearly understood. Notably, this protein interacts with other structural proteins such as membrane(M) and other accessory proteins viz. ORF3a, ORF7a and host cell proteins [8]. Envelope protein of SARS-CoV2 is 76 amino acids long and 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 [9,10].

Four mutations including one deletion have been found in the envelope protein of SARS-CoV2 with reference to the SARS-CoV1, a species of coronavirus that also infects humans, bats and certain other mammals. The alignment of the envelope proteins of the SARS-CoV1 and SARS-CoV2 is given in Fig. 2.

It is reported that the C-terminus domain of the envelope protein contains the motif DLLV which binds to the host cell PALS1 protein to facilitate infection [9,11,12].

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 discussed.

2. Methods

From the NCBI virus database, all the protein sequences of 3617 SARS-CoV2 genomes were fetched. Then the amino acid sequences of envelope protein of SARS-CoV2 are exported in fasta format using file operations through Matlab. These sequences (fasta formatted) are blasted using Clustal-Omega and found the mismatched and from their mutations and their associated positions were detected [13].

* Corresponding author.
E-mail addresses: sarimif@gmail.com (S.S. Hassan), pabitra@isical.ac.in (P.P. Choudhury), broy@isical.ac.in (B. Roy).

https://doi.org/10.1016/j.ygeno.2020.07.001
Received 9 June 2020; Received in revised form 30 June 2020; Accepted 1 July 2020
Available online 05 July 2020

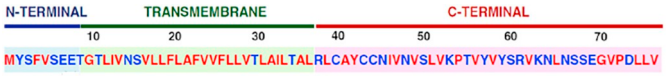

Fig. 1. Amino acid sequence and domains of the envelope protein of SARS-CoV2 [7].
Red and blue colors are representing hydrophobic and hydrophilic amino acid, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Results

Among these virus genomes from 3617 patients; 2588 were from the 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 E-protein protein over the available 3617 SARS-CoV2 genomes (Table 1). It is to be noted that 10 (0.386%) out of 2588 genomes from USA, 3 (0.806%) from Asia, 1 (0.348%) from Europe and 1 (0.274%) from Oceania) contained the missense mutations (Table 1) in the envelope proteins of SARS-CoV2 genomes. Changes of the R-group of each amino acid according to the mutations are also presented (Table-1). It is to be noted that 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.

- In less than 0.5% of the SARS-CoV2 genomes, the E-protein possesses the missense mutations as adumbrated in the Table 1. In TMD and C-terminus domain, there are nine different mutations where the R-group property changes. But only in QHZ00381, for the mutation L37H in the TMD of the envelope protein causes changes in amino acid from hydrophobic to hydrophilic.
- TMD was also observed to be conserved over the SARS-CoV1 and COV2 genomes, but 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 TMD of the envelope protein of the virus from South Korea may affect the ion channel activity of the envelope protein.
- The motif *'DLLV'* has been changed 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'* may mis-target the PALS1 at Golgi and delaying TJ formation and accordingly may influence replication and/or infectivity of the virus [10].
- In the C-terminus domain of the E-protein of SARS-CoV2 the amino acid S at 68th position changes to the amino acids F and C in the proteins {*QKG*87268, *QKG*88576} from the USA: Massachusetts and QKI36855 from China: Guangzhou respectively. Note that the mutation of the amino acid S to F keeps the R-group property unchanged (i.e. hydrophobic to hydrophilic) while that of the amino acid S to C changes the R-group from Hydrophilic to Hydrophobic. This would possibly make changes in protein functions and interactions.

```
CLUSTAL O(1.2.4) multiple sequence alignment


SARS-CoV1-Envelope      MYSFVSEETGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPTVYVYS    60
SARS-CoV2-Envelope      MYSFVSEETGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYS    60
                        ************************************************:.****

SARS-CoV1-Envelope      RVKNLNSSEGVPDLLV        76
SARS-CoV2-Envelope      RVKNLNSSR-VPDLLV        75
                        ********. ******
```

Fig. 2. Clustal alignment of the envelope protein of SARS-CoV1 and SARS-CoV2.
Mutations in (C)-terminus domain in the E protein 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).The deletion mutation of G at the 70th position with respect to the reference envelope protein of SARS-CoV1 is also noted.

Table 1
Non- synonymous mutation in the E-protein of SARS-CoV2.

Protein-ID	Geo-location	Mutation	Domain	Change of R-group
QJA42107	USA: VA	A36V	TMD[a]	Hydrophobic to Hydrophobic
QJQ84222	USA: KENNER, LA	L26F	TMD	Hydrophobic to Hydrophobic
QHZ00381	South Korea	L37H	TMD	Hydrophobic to Hydrophilic
QJS53352	Greece: Athens	L39M	TMD	Hydrophobic to Hydrophobic
QJR88103	Australia: Victoria	L73F	C-terminus	Hydrophobic to Hydrophobic
QKE45838	USA: CA	P71L	C-terminus	Hydrophobic to Hydrophobic
QKE45886	USA: CA	P71L	C-terminus	Hydrophobic to Hydrophobic
QKE45898	USA: CA	P71L	C-terminus	Hydrophobic to Hydrophobic
QKE45910	USA: CA	P71L	C-terminus	Hydrophobic to Hydrophobic
QJE38284	USA: CA	P71L	C-terminus	Hydrophobic to Hydrophobic
QIU81527	USA: WA	P71L	C-terminus	Hydrophobic to Hydrophobic
QKG87268	USA: Massachusetts	S68F	C-terminus	Hydrophobic to Hydrophobic
QKG88576	USA: Massachusetts	S68F	C-terminus	Hydrophobic to Hydrophobic
QKI36831	China: Guangzhou	D72Y	C-terminus	Hydrophilic to Hydrophobic
QKI36855	China: Guangzhou	S68C	C-terminus	Hydrophilic to Hydrophobic

[a] TMD: transmembrane domain.

4. Concluding remarks

Among all the proteins present in the novel RNA virus, some accessory proteins such as ORF6, ORF7b, ORF8, ORF10 contain the least number of missense mutation as reported in various studies [14–16]. And same is true for E-protein. We find 15 among 3617 (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 [17]. 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 and examined the mutations. SH, PPC, BR analyzed the data and result. SH wrote the initial draft which was checked and edited by all other authors to generate the final version.

Declaration of Competing Interest

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, J. Transl. Med. 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, Unpublished, May 2020.

[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, Genomics 112 (5) (2020) 3226–3237.

[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, Futur. Virol. (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 hace-2 transgenic mice, Virology 376 (2) (2008) 379–389.

[7] D. Schoeman, B.C. Fielding, Coronavirus envelope protein: current knowledge, Virol. J. 16 (1) (2019) 69.

[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, Mol. Biol. Cell 21 (22) (2010) 3838–3852.

[10] X. Liu, E.J. Fuentes, Emerging themes in pdz domain signaling: Structure, function, and inhibition, International Review of Cell and Molecular Biology, 343 Elsevier, 2019, pp. 129–218.

[11] T. Beuming, L. Skrabanek, M.Y. Niv, P. Mukherjee, H. Weinstein, Pdzbase: a protein–protein interaction database for pdz-domains, Bioinformatics 21 (6) (2005) 827–828.

[12] M. Peiris, R. Bruzzone, B. Margolis, B. Nal, The Sars Coronavirus e Protein Interacts With pals1 and alters 5 Tight Junction Formation and Epithelial Morphogenesis 6.

[13] 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 Res. 47 (W1) (2019) W636–W641.

[14] S. Chatterjee, Understanding the Nature of Variations in Structural Sequences Coding for Coronavirus Spike, Envelope, Membrane and Nucleocapsid Proteins of sars-cov-2, (March 28, 2020) (2020).

[15] 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, Virol. J. 2 (1) (2005) 51.

[16] 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, The Nidoviruses, Springer, 2006, pp. 115–120.

[17] M. López de Diego, J.L. Nieto-Torres, J.M. Jiménez-Guardeño, J.Á. Regla-Nava, C. Castaño-Rodrguez, R. Fernandez-Delgado, F. Usera, L. Enjuanes Sánchez, Coronavirus Virulence Genes With Main Focus on Sars-cov Envelope Gene, (2014).