Genetic variations in the Orf7a protein of SARS-CoV-2 and its possible role in vaccine development

Niti Yashvardhini1, Deepak Kumar Jha2,*,†, Amit Kumar3, Kumar Sayrav4, Manjush Gaurav3

ABSTRACT
Introduction: SARS-CoV-2 (Severe Acute Respiratory Syndrome) is the causative agent of the novel coronavirus disease (COVID-19) that has been creating an unprecedented situation globally. The recurrent mutations in SARS-CoV-2 genomes impact on the vaccine designing strategies. The Orf7a is a 121-amino acid long-type I transmembrane accessory protein encoded by the genome of SARS-CoV-2 and plays a crucial role in the virus–host interaction. The present study aimed to analyze the variations occurring in Orf7a due to multiple mutations and its immunological role in developing a promising therapeutic target to curb SARS-CoV-2 infections. Methods: 16,161 sequences of Orf7a reported from the onset of this disease until 13 June 2021 from five continents were compared to identify genetic variations in the protein. Results: A total of 470 point mutations were detected in the sequences submitted. Subsequently, the nature of mutations (deleterious or neutral) was determined. Furthermore, the physicochemical properties, antigenicity, allergenicity, toxicity, and stability of Orf7a protein were estimated to demonstrate the stability of the protein. Additionally, we identified three B-cell immune epitopes, and their MHC cluster analysis was also performed. Conclusion: The recurrent mutations in Orf7a of SARS-CoV-2 provide a deep understanding of its role in the virus–host interactions. Findings of our study revealed that the predicted epitopes could be promising candidates for a vaccine against COVID-19 infections.

Key words: COVID-19, Epitope, Mutation, Orf7a, SARS-CoV-2, Vaccine

INTRODUCTION
SARS-CoV-2 is responsible for the rapid emergence of novel coronavirus disease, first reported at the wet seafood market of Wuhan city of China in December 20191,2. COVID-19 is a contagious disease that induces mild to severe respiratory illness, including multi-organ dysfunction in the infected individuals3. SARS-CoV-2 transmission occurs via the inhalation of aerosols or direct contact with the droplets from an infected person. It has been observed that the incubation period of COVID-19 infection commonly varies between 2 to 14 days4. COVID-19 has been declared a pandemic on 11th March 2020 by the World Health Organization (WHO, 2020). As of July 13th, 2021, worldwide, 190,561,846 confirmed cases of COVID-19 had been reported to WHO, including 4,095,470 casualties5.

Coronaviruses (CoVs) are enveloped positive-sense single-stranded RNA viruses belonging to the coronaviridae family. The genetic material is of ~30 kb length encoding polyproteins of 9860 long chain of amino acids6. The genome of SARS-CoV-2 encodes four main structural proteins (spike S, envelope E, membrane M and nucleocapsid N), nine accessory open reading frames (Orf3a, Orf3b, Orf6, Orf7a, Orf7b, Orf8a, Orf8b and Orf99b) and several non-structural proteins ranging from NSP1 to NSP167,8. Orf7a is made up of 121 long amino acids chain of accessory protein in SARS-CoV-2 that plays an important role in virus-host interaction. ORF7a of SARS-CoV-2 consists of the information of a type I transmembrane protein, which is primarily located in the Golgi apparatus but can also be found on the cell surface9,10.

RNA viruses like SARS-CoV-2 exhibit higher rates of genetic mutation than DNA viruses which leads to genomic diversity. Thus, SARS-CoV-2 acquires genetic heterogeneity that modulates virulence properties in the host and thereby facilitating the immune evasion of host11–13. A total of 470 point mutations were detected from 16,161 sequences submitted since the onset of this disease up to 13th June 2021. Additionally, using the predictive tools of computational biology, we attempted to design the epitope-based vaccine candidates that can generate long-lasting B-cell immune responses against SARS-CoV-2 infections. This study also highlights the physicochemical properties, antigenicity, allergenicity, and toxicity of vaccine construct and their MHC cluster analysis, which revealed predicted epitopes can be a potent vaccine candidate.
vaccine candidate to minimize COVID-19 infections. The purpose of the present study was, to analyze the variations occurring in Orf7a protein due to multiple point mutations leading to the alterations in the structure of Orf7a and its immunological role in designing epitope-based vaccine candidates against COVID-19 infections. Moreover, this in silico research work further needs validation through in vitro and in vivo studies.

METHODS

Data mining

The full-length protein sequence of Orf7a protein of SARS-CoV-2 was downloaded from the NCBI virus database, submitted from five different continents: Asia, Africa, Europe, Oceania, and South America till 13th June 2021. There were nearly 16,161 sequences released from different continents since the onset of this pandemic. For the mutation studies, a reference sequence of the Orf7a protein of the Wuhan virus was also downloaded with accession number QWZ15014.

Multiple sequence alignment and identification of Orf7a mutants

The full-length Orf7a protein sequences were aligned using Clustal Omega online platform, and the aligned files were viewed using Jalview to detect the mutations regarding Wuhan type virus sequence. The frequency of mutations was calculated to check if different point mutations were from different continents. The non-synonymous amino acid variants were analyzed using Protein Variation Effect Analyzer known as PROVEAN v1.1.3 with a cutoff predicted score of -2.50 to detect the effect of mutation on the Orf7a protein.

Estimation of physicochemical properties and hydropathy index of Orf7a protein

The physicochemical properties, which include molecular weight, extinction coefficient, amino acid composition, instability index, estimated half-life, aliphatic index, and an average of hydrophobicity (GRAVY) was calculated using ProtParam tool of the Expasy online program. Protscale tool of expense was used for preparing hydropathy plot of Orf7a protein.

Identification of linear B-cell epitopes

IEDB was used to predict the linear B-cell epitopes in the Orf7a protein of SARS-CoV-2. IEDB web server constructs epitopes based on estimation of parameters such as flexibility, accessibility, hydrophilicity, turns, polarity, and the antigenic propensity of the protein using amino acid scales and HMMs.

MHC allele cluster analysis

MHCluster 2.0 online tool was used to analyze MHC class I and MHC class II alleles, which might interact with the epitopes leading to the immune responses. This online server predicts epitopes and the allele binding phylogenetically in the form of clusters and heatmap.

Antigenicity and allergenicity evaluation

The antigenicity of the Orf7a protein was estimated using the Vaxijen v2.0 server, which predicts antigens according to the auto cross-covariance (ACC) transformation of the protein sequences. To detect whether the Orf7a protein was allergenic, an AllerTOP server was used, which evaluates protein allergenicity on autocross variance (ACC method) that explains residues based on hydrophobicity, size, flexibility, and other parameters.

RESULTS

Identification of Orf7a mutants and detection of non-synonymous mutants

A total of 16,161 full-length protein sequences of Orf7a, 121 amino acids in length were submitted from all the five continents (Asia, Africa, Europe, Oceania, and South America) till June 2021 since the onset of this pandemic. These sequences were downloaded along with a reference sequence of Wuhan-type virus from the NCBI virus database. The multiple sequence alignment was performed to detect the variations in the isolates and visualized using Jalview. Among these point mutations, N43Y, T14I, V82A, S81L, and T39I were the most frequently occurring mutations and were used for further characterization in this study (Figure 1). All these five frequent mutations were deleterious for the Orf7a protein at 2.5 cutoff values of PROVEAN scores (Table 1).

Estimation of physicochemical properties and hydropathy index of Orf7a accessory protein

The estimation of physicochemical properties of Orf7a protein revealed that Orf7a protein is 121 amino acids long with a molecular weight 13744.17 Da, aliphatic index 48.66, instability index 48.66, and GRAVY score of 0.233 (Table 2). The structure of the
Figure 1: Frequency of mutations in Orf7a protein from five different continents.
https://doi.org/10.6084/m9.figshare.16529691.v1

Table 1: List of nonsynonymous amino acid substitutions in Orf7a protein (cutoff = -2.5)

| Variant | PROVEAN score | Prediction (cutoff=-2.5) |
|---------|---------------|--------------------------|
| N43Y    | -8.000        | Deleterious              |
| T14I    | -3.193        | Deleterious              |
| V82A    | -2.667        | Deleterious              |
| S81L    | -4.000        | Deleterious              |
| T39I    | -6.000        | Deleterious              |

Figure 2: Structure of Orf7a transmembrane protein as obtained by TMHMM server which predicts the occurrence of different amino acids in different locations of the membrane.
https://doi.org/10.6084/m9.figshare.16529694.v1
Table 2: Physicochemical properties of ORF7a protein

| Physicochemical properties          | ORF7a          | Amino acid composition | No. | Percent composition (%) |
|------------------------------------|----------------|------------------------|-----|-------------------------|
| Molecular weight                   | 13744.17       | Ala (A)                | 9   | 7.4                     |
| No. of amino acids                 | 121            | Arg (R)                | 1   | 4.1                     |
| Theoretical pI                      | 8.23           | Asn (N)                | 2   | 1.7                     |
| Instability index                   | 48.66          | Asp (D)                | 2   | 1.7                     |
| No. of negatively charged (Asp+Glu)| 10             | Cys (C)                | 6   | 5.0                     |
| No. of positively charged (Arg+Lys)| 12             | Gln (Q)                | 3   | 4.1                     |
| Aliphatic index                     | 48.66          | Glu (E)                | 8   | 6.6                     |
| Grand average of hydro-pathicity   | 0.233          | Gly (G)                | 4   | 3.3                     |
| Estimated half-life (mammalian reticulocytes, in vitro) | 30 hours       | His (H)               | 3   | 2.5                     |
| Atomic composition                 |                | Ile (I)               | 8   | 6.6                     |
| C                                  | 633            | Leu (L)               | 15  | 12.4                    |
| H                                  | 988            | Lys (K)               | 7   | 5.8                     |
| N                                  | 156            | Met (M)               | 1   | 0.8                     |
| O                                  | 171            | Phe (F)               | 10  | 8.3                     |
| S                                  | 7              | Pro (P)               | 6   | 5.0                     |
| Formula                            | C633H988N156O171S7 | Ser (S)            | 7   | 5.8                     |
| Total number of atoms              | 19             | Thr (T)               | 10  | 8.3                     |
|                                    |                | Trp (W)               | 0   | 0.0                     |
|                                    |                | Tyr (Y)               | 5   | 4.1                     |
|                                    |                | Val (V)               | 8   | 6.6                     |
|                                    |                | Phy (O)               | 0   | 0.0                     |
|                                    |                | Sec (U)               | 0   | 0.0                     |

Table 3: B-cell epitopes of Orf7a protein of SARS-CoV-2

| No. | Start | End | Peptide       | Length |
|-----|-------|-----|---------------|--------|
| 1   | 17    | 25  | LYHYQECVR     | 9      |
| 2   | 33    | 51  | EPCSSGTVYEGNSPFHPLAD | 19     |
| 3   | 71    | 96  | VKHYYQLRARSVSPKLIRQEEVQEL | 26     |
Figure 3: Hydropathy plot of wild type Orf7a protein showing hydrophobic amino acid residues. https://doi.org/10.6084/m9.figshare.16529697.v1

ORF7a protein is shown in Figure 2. The hydropathy plot showed the C-terminal amino acid to be relatively more hydrophobic than the N-terminal end of the Orf7a protein (Figure 3).

B-cell epitope prediction
A total of three linear B-cell epitopes were predicted for 121 amino acids long Orf7a protein, as shown in Figure 4 and Table 3. These epitopes can induce antibody production and hence play a crucial role in humoral immunity.

Cluster analysis of MHC alleles
The cluster analysis of the MHC class I allele is shown in Figure 5A&B while that of class II allele is shown in Figure 5C&D, where the red zone denotes strong interaction of the HLA allele with the epitopes of Orf7a protein, whereas yellow depicts weak interaction. We analyzed the binding ability of all the possible alleles with the Orf7a epitopes.
Assessment of antigenicity and allergenicity

To predict the antigenicity of Orf7a protein, the Vaxijen v2.0 server was used, which predicts antigenicity based on the ability of the vaccine candidate to bind with the B-cell and T-cell receptors and hence can enhance the immune response. This analysis revealed the antigenic nature of Orf7a protein with an antigenicity score of 0.6441 at a threshold of 0.4%. A good vaccine candidate needs to be non-allergenic; hence, the allergenicity and toxicity analysis of Orf7a protein revealed its non-allergenic nature, hence it is possibly a potent vaccine candidate.

DISCUSSION

The rapid spread of coronavirus disease started in China, in late December 2019 and has become a serious threat to human health across the globe. Therefore, efficacious and safe antiviral therapeutics are indispensable to curb COVID-19 infections. Primarily, the novel coronavirus causes a pulmonary obstruction with multi-organ dysfunction in humans, whose manifestation encompasses dyspnea (shortness of breath), sore throat, dry cough, and fever. The symptoms of the COVID-19 begin within two days, or it may take up to \( \geq 14 \) days. COVID-19 infections may have some symptoms, or the infected individuals may appear to be asymptomatic.

SARS-CoV-2 is an RNA virus and has an enormous capacity to exhibit high rates of mutation. It has been observed in previous studies that mutation plays a vital role in viral evolution and adaptations. Since these traits are found to be the key determinants for viruses to live in the dynamic host environment...
and enabling them to escape the pre-existing immunity of the host and quickly acquire drug resistance. Various factors are responsible for the rapid spread of SARS-CoV-2 infection, such as fidelity of its RNA polymerase, population density, different geographical regions, poor health or hygiene, and environmental conditions.

Mutational analysis of this contagious virus provides a better understanding of its epidemiology, pathogenesis, and design of suitable antiviral therapeutics to fight against COVID-19 infections. We detected 470 point mutations from 16,161 sequences of Orf7a proteins around the world. RNA viruses, including SARS-Cov-2, can accumulate genomic mutations through an error-prone viral enzyme reverse transcriptase and better adapt inside the host, which further creates hurdles in designing antiviral therapeutics against RNA viruses.

The main function of ORF7a is binding and preventing N-linked glycosylation of BST-2 (Bone marrow stromal antigen 2, also called CD317 or tetherin), therefore, blocking the tethering of SARS-CoV viruses to the cytoplasmic membrane after they are released from the cell. Taylor JK et al. (2015) have reported that SARS-CoV ORF7a antagonizes the function of BST-2 and suggested that therapeutics designed to inhibit the interaction between BST-2 and ORF7a might be inhibiting virus growth both in vitro and in vivo.

Epitope-based vaccine designing strategies using various tools of immunoinformatics gained much attention for various infectious diseases in recent times. The conventional methods of vaccine development are costly, time-consuming, and require lots of experimental work. However, the epitope-based approach of vaccine designing uses several predictive tools of bioinformatics and has proven to be highly advantageous over the traditional vaccine development strategies. As evident from the earlier studies, in silico vaccine development methods, seem to be specific, easily establish an immunological correlation between host and pathogens, and can elicit long-lasting immunity.

Previous studies have shown that epitope-based vaccine candidates might be a potential target to combat SARS-CoV-2 infections. Therefore, for designing the epitope-based vaccine candidate, antigenicity, allergenicity, physicochemical properties, toxicity, and stability of Orf7a protein were explored to demonstrate the stability of the protein. In addition, we identified 3 B-cell immune epitopes, and its MHC cluster analysis has also been performed, which revealed predicted epitopes might be a promising vaccine candidate to combat COVID-19 infections.

CONCLUSIONS
The occurrence of recurrent mutations in the Orf7a of SARS-CoV-2 provides a deep understanding of its role in the virus-host interaction. For designing vaccine construct, Orf7a of coronavirus has been chosen as a good target since Orf7a is a type I transmembrane protein. Moreover, our study highlights the high efficacy and durability of designed epitopes-based vaccine construct using predictive immunoinformatics tools; further, in vitro and in vivo studies are mandatory to validate designed vaccine candidates.

ABBREVIATIONS
COVID-19: Coronavirus disease 2019
MHC: Major Histocompatibility Complex
Orf7a: Open Reading Frame 7a
SARS: Severe acute respiratory syndrome

ACKNOWLEDGMENTS
None.

AUTHOR’S CONTRIBUTIONS
NY DKJ performed all the analysis, AK MG KS performed mutational study, NY DKJ wrote the manuscript. All authors read and approved the final manuscript.

FUNDING
None.

AVAILABILITY OF DATA AND MATERIALS
Not applicable.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE
Not applicable.

CONSENT FOR PUBLICATION
Not applicable.

COMPETING INTERESTS
The authors declare that they have no competing interests.

REFERENCES
1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395(10223):497–506. PMID:31986264. Available from: 10.1016/S0140-6736(20)30183-5.
2. Lu H, Stratton CW, Tang YW. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. J Med Virol. 2020;92(4):401–402. PMID:31950516. Available from: 10.1002/jmv.25678.
4503
16. Gasteiger E, Hoogland C, Gattiker A. Protein Identification and Analysis Tools on the ExPaSy Server. Prot Protoc Handb. 2005. Available from: 10.1385/1-59259-890-6:571.

17. Kim Y, Ponomarenko J, Zhu Z, Tamang D, Wang P, Greenbaum J. Immune epitope database analysis resource. Nucleic Acids Research. 2012;40(Web Server issue):525–30. PMID: 22610854.

18. Thomsen M, Lundegaard C, Buus S, Lund O, Nielsen M. MHC-cluster, a method for functional clustering of MHC molecules. Immunogenetics. 2013;65(9):653–65. PMID: 23775223. Available from: 10.1007/s00251-013-0714-9.

19. Doytchinova IA, Flower DR. VaxiGen: a server for predicting protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007;8(1):4. PMID: 17207271. Available from: 10.1186/1471-2105-8-4.

20. Dimitrov I, Flower DR, Doytchinova I. AllerTOP - a server for in silico prediction of allergens. BMC Bioinformatics. 2013;14:4. PMID: 23735058. Available from: 10.1186/1471-2105-14-6.

21. Benvenuto D, Giovanetti M, Ciccozzi A, Spoto S, Angeletti S, Ciccozzi M. The 2019-coronavirus epidemic: evidence for virus evolution. Journal of Medical Virology. 2020;92(4):455–9. PMID: 31994738. Available from: 10.1002/jmv.25688.

22. Yashvardhini N, Jha DK. In silico analysis of mutational variants of SARS-CoV-2 RdRp protein. Res J Life Sci Bioinfo Pharma Chem sc. 2020;6(5):68–76.

23. Yashvardhini N, Jha DK. Occurrence of Recurrent Mutations in SARS-CoV-2 Genome and its Implications in the Drug Designing Strategies. Journal of Pharmacy - Pharmacognosy Research. 2020:96–102.

24. Wang MA, et al. Temperature Significantly Change COVID-19 Transmission in 429 cities. medRxiv. 2020;2020.02.22.20025791. Available from: 10.1101/2020.02.22.20025791.

25. Mishra CB, Pandey P, Sharma RD, Malik MZ, Mongre RK, Lynn AM, et al. Identifying the natural polyphenal catechin as a multi-targeted agent against SARS-CoV-2 for the plausible therapy of COVID-19: an integrated computational approach. Brief Bioinform. 2021;22(2):1346–1360. PMID: 33386025. Available from: 10.1093/bib/bbab78.

26. Taylor JK, Coleman CM, Postel S, Sisk JM, Bernbaum JG, Venkataraman T. Severe Acute Respiratory Syndrome Coronavirus ORF7a Inhibits Bone Marrow Stromal Antigen 2 Virion Tethering through a Novel Mechanism of Glycosylation Interference. Journal of Virology. 2015;89(23):11820–33. PMID: 26378163. Available from: 10.1128/JVI.02274-15.

27. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579(7798):265–9. PMID: 32015508. Available from: 10.1038/s41586-020-2008-3.

28. Tai W, He L, Zhang X, Pu J, Voronin D, Jiang S, et al. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. Cellular & Molecular Immunology. 2020;17(6):163–20. PMID: 32023189. Available from: 10.1038/s41423-020-0400-4.

29. Jakhar R, Gakhar SK. An Immunoinformatics Study to Predict Epitopes in the Envelope Protein of SARS-CoV-2. The Canadian Journal of Infectious Diseases & Medical Microbiology. 2020;2020:7097356. PMID: 33299503. Available from: 10.1155/2020/7097356.

30. Chiov SS, Fan YC, Crill WD, Chang Ry, Chang GJ. Mutation analysis of the cross-reactive epitopes of Japanese encephalitis virus envelope glycoprotein. The Journal of General Virology. 2012;93(Pt 6):1185–92. PMID: 22337639. Available from: 10.1099/vir.0.040238-0.
Ready to submit your manuscript? Choose Biomedpress and benefit from:

- Fast, convenient online submission
- Through peer-review by experienced researchers
- Rapid publication on acceptance
- Free of charge (without publication fees)

Learn more [http://www.biomedpress.org/journals/](http://www.biomedpress.org/journals/)

| Journal Name | Indexed | Acceptance Rate | Article Publishing Charge | Submission to first editorial decision |
|--------------|---------|-----------------|---------------------------|----------------------------------------|
| Biomedical Research and Therapy | Web of Science (ESCI), Embase, Google Scholar | 54.32% | Free | 27 days |
| Progress in Stem Cell | Embase, Google Scholar | 78.19% | Free | 19 days |
| Asian Journal of Health Sciences | Google Scholar | 72.89% | Free | 16.5 days |