Notable sequence homology of the ORF10 protein introspects the architecture of SARS-COV-2

Sk. Sarif Hassan\textsuperscript{1,*}, Diksha Attrish\textsuperscript{2,+}, Shinjini Ghosh\textsuperscript{3,+}, Pabitra Pal Choudhury\textsuperscript{4}, Vladimir N. Uversky\textsuperscript{5,†}, Bruce D. Uhal\textsuperscript{6}, Kenneth Lundstrom\textsuperscript{7}, Nima Rezaei\textsuperscript{8,9}, Alaa A. A. Aljabali\textsuperscript{10}, Murat Seyran\textsuperscript{11}, Damiano Pizzol\textsuperscript{12}, Parise Adadi\textsuperscript{13}, Tarek Mohamed Abd El-Aziz\textsuperscript{14,15}, Antonio Soares\textsuperscript{14}, Ramesh Kandimalla\textsuperscript{16}, Murtaza Tambuwala\textsuperscript{17}, Amos Lai\textsuperscript{18}, Gajendra Kumar Azad\textsuperscript{19}, Samendra P. Sherchan\textsuperscript{20}, Wagner Baetas-da-Cruz\textsuperscript{21}, Giorgio Palu\textsuperscript{22}, and Adam M. Brufsky\textsuperscript{23}

\textsuperscript{1}Department of Mathematics, Pingla Thana Mahavidyalaya, Maligram, Paschim Medinipur, 721140, West Bengal, India
\textsuperscript{2}Dr. B. R. Ambedkar Centre For Biomedical Research (ACBR), University of Delhi (North Campus), Delhi 110007, India
\textsuperscript{3}Department of Biophysics, Molecular Biology and Bioinformatics, University of Calcutta, Kolkata 700009, West Bengal, India
\textsuperscript{4}Applied Statistics Unit, Indian Statistical Institute, Kolkata 700108, West Bengal, India
\textsuperscript{5}Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA
\textsuperscript{6}Department of Physiology, Michigan State University, East Lansing, MI 48824, USA
\textsuperscript{7}PanTherapeutics, Rte de Lavaux 49, CH1095 Lutry, Switzerland
\textsuperscript{8}Research Center for Immunodeficiencies, Pediatrics Center of Excellence, Children's Medical Center, Tehran University of Medical Sciences, Tehran, Iran
\textsuperscript{9}Network of Immunity in Infection, Malignancy and Autoimmunity (NIIMA), Universal Scientific Education and Research Network (USERN), Stockholm, Sweden
\textsuperscript{10}Department of Pharmaceutics and Pharmaceutical Technology, Yarmouk University-Faculty of Pharmacy, Irbid 566, Jordan
\textsuperscript{11}Doctoral studies in natural and technical sciences (SPL 44), University of Vienna, Austria
\textsuperscript{12}Italian Agency for Development Cooperation - Khartoum, Sudan Street 33, Al Amarat, Sudan
\textsuperscript{13}Department of Food Science, University of Otago, Dunedin 9054, New Zealand
\textsuperscript{14}Department of Cellular and Integrative Physiology, University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Dr, San Antonio, TX 78229-3900, USA
\textsuperscript{15}Zoology Department, Faculty of Science, Minia University, El-Minia 61519, Egypt
\textsuperscript{16}CSIR-Indian Institute of Chemical Technology Uppal Road, Tarnaka, Hyderabad-500007, Telangana State, India
\textsuperscript{17}School of Pharmacy and Pharmaceutical Science, Ulster University, Coleraine BT52 1SA, Northern Ireland, UK
\textsuperscript{18}Division of Pulmonary and Critical Care Medicine, Mayo Clinic, Rochester, Minnesota, USA
\textsuperscript{19}Department of Zoology, Patna University, Patna-800005, Bihar, India
\textsuperscript{20}Department of Environmental Health Sciences, Tulane University, New Orleans, LA, 70112, USA
\textsuperscript{21}Translational Laboratory in Molecular Physiology, Centre for Experimental Surgery, College of Medicine, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil
\textsuperscript{22}Department of Molecular Medicine, University of Padova, Via Gabelli 63, 35121, Padova, Italy
\textsuperscript{23}University of Pittsburgh School of Medicine, Department of Medicine, Division of Hematology/Oncology, UPMC Hillman Cancer Center, Pittsburgh, PA, USA

\textsuperscript{*}vuversky@usf.edu
\textsuperscript{†}sarimif@gmail.com
\textsuperscript{+}These authors contributed equally to this work

ABSTRACT
Introduction

Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2), responsible for the global pandemic, has brought the whole world to a stand-still\textsuperscript{12}. The contagious nature of this virus is concerning as it has infected more than 25 million people worldwide claiming 850,000 deaths, so far\textsuperscript{3-5}. In addition to low-pathogenicity and endemic coronaviruses, high pathogenic Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) and Middle East Respiratory Syndrome Coronavirus (MERS-CoV) in 2002 and 2013, respectively, caused severe human illnesses, e.g. pneumonia, and renal failure but without any pandemic grade transmission capacities\textsuperscript{6}. SARS-CoV had a 9.7\% infection fatality rate and MERS a 37\% infection fatality rate, but SARS-CoV-2 has a lower than 1\% infection fatality\textsuperscript{7}. Therefore, it is vital to monitor critical mutations in the proteins such as ORF10 (open reading frame 10) that could possibly change viral pathogenicity. SARS-CoV-2 is a Baltimore class IV positive-sense, single-stranded RNA virus with four structural proteins, sixteen non-structural proteins, and six accessory proteins\textsuperscript{8}.

The smallest accessory protein in SARS-CoV-2, the 38-residue peptide ORF10, and distinguishes the infection more rapidly than PCR based strategies\textsuperscript{9}. The protein SARS-CoV-2 ORF10 has the highest number of immunogenic epitopes of all putative ORF proteins, therefore making it a potential target for vaccine development\textsuperscript{10}. Due to its short length, ORF10 has been suggested to be an insertion mutation. However, this is unlikely as the ORF10 gene is present at the terminal its sgRNA sequence. It has been hypothesised that ORF10 is a transposon, but this is also unlike as transposons are of larger size\textsuperscript{9}.

ORF10 consists of a Molecular Recognition Feature (MoRF) region from amino acid residue 3 to 7, which is a molecular recognition site for interaction with other proteins\textsuperscript{11}. It is one of the critical properties of intrinsically disordered proteins that allow proteins to adapt an ensemble of conformations when bound to different proteins, and this permits interaction with multiple proteins\textsuperscript{12}. Through high-throughput analysis it was revealed that ORF10 can interact with a large number of host proteins despite its small structure; therefore, this aspect can be likely attributed to the MoRF region\textsuperscript{11}. Through bioinformatics, it was previously reported that the SARS-CoV-2 ORF10 exhibits interaction with multiple members of the Cullin-ubiquitin-ligase complex and controls the host-ubiquitin machinery for viral pathogenesis\textsuperscript{13-16}.

Humans may not have been able to utilize any memory B and T cells elicited against other microorganisms to target ORF10 and fight SARS-CoV-2, contributing to its contagious nature\textsuperscript{17}. It was further reported that no sequence homology was found with any protein in the NCBI protein depository. Recently, SARS-CoV-2 ORF10 is found to have 99.15\% nucleotide similarity to that of Pangolin-CoV-2020\textsuperscript{18,19}.

The present study examines mutations discovered in SARS-CoV-2 ORF10 variants, which along with their physicochemical and immunological properties suggests the significance of these mutations to alter pathogenesis and to possibly identify some potential vaccine candidates. A inclusive parity and disparity analysis between the two ORF10 proteins of SARS-CoV-2 and Pangolin-CoV was also conducted.

Results

Mutations in SARS-CoV-2 ORF10

Each unique ORF10 sequence was aligned using the National Center for Biotechnology Information (NCBI) protein p-blast and omega blast suites to determine the mismatches and thereby, the missense mutations (amino acid changes) were identified\textsuperscript{20,21} (Figure 1(A)). A mutation from one amino acid A\textsubscript{1} to another A\textsubscript{2} at the position p is denoted by A\textsubscript{1}pA\textsubscript{2} or A\textsubscript{(p)}A\textsubscript{2}. Based on the mutations, conserved and non-conserved residues in ORF10 proteins are identified and marked in different colors in (Figure...
Also, the molecular recognition features (MoRF) (YINVF) are predicted using the server MoRFchibi for the ORF10 Wuhan sequence.

Figure 1. (A): Mutations and their amino acid positions in ORF10 proteins of SARS-CoV-2; (B): Conserved, mutated residues and molecular recognition features of ORF10 (YP_009725255) of SARS-CoV-2.

There are altogether 22 distinct missense mutations which were examined across 22 unique ORF10 variants of SARS-CoV-2. These missense mutations are found in the entire ORF10 sequence starting from the amino acid position 2 to 38. The amino acids arginine (R), valine (V), and leucine (L) are substituted to more than one amino acid at fixed positions (marked magenta in Figure 1(B). The largest conserved region across all the 24 ORF10 variants is "SLLLC" at positions 15–19.

Note that each unique variant (Table 2) of SARS-CoV-2 ORF10 possesses a single missense mutation (Table 1).

<table>
<thead>
<tr>
<th>Accession ID</th>
<th>Mutations</th>
<th>Type of Mutations</th>
<th>PROVEAN Score</th>
<th>Effect of Mutations on Structure</th>
<th>RI</th>
<th>Polarity Changes</th>
<th>Charge</th>
</tr>
</thead>
<tbody>
<tr>
<td>QNI23218.1</td>
<td>G2D</td>
<td>Deleterious</td>
<td>-7</td>
<td>Decrease</td>
<td>7</td>
<td>NP to P</td>
<td>Neutral to Acidic</td>
</tr>
<tr>
<td>QIS29991.1</td>
<td>V6I</td>
<td>Neutral</td>
<td>-1</td>
<td>Decrease</td>
<td>7</td>
<td>NP to NP</td>
<td>Neutral to Neutral</td>
</tr>
<tr>
<td>QLE33451.1</td>
<td>Y14C</td>
<td>Deleterious</td>
<td>-9</td>
<td>Decrease</td>
<td>2</td>
<td>P to P</td>
<td>Neutral to Neutral</td>
</tr>
<tr>
<td>QNC04532.1</td>
<td>R20I</td>
<td>Deleterious</td>
<td>-8</td>
<td>Decrease</td>
<td>3</td>
<td>NP to NP</td>
<td>Basic (strongly) to Neutral</td>
</tr>
<tr>
<td>QLA48060.1</td>
<td>R20K</td>
<td>Deleterious</td>
<td>-3</td>
<td>Decrease</td>
<td>8</td>
<td>NP to P</td>
<td>Basic (strongly) to Basic</td>
</tr>
<tr>
<td>QMT97141.1</td>
<td>S23F</td>
<td>Deleterious</td>
<td>-6</td>
<td>Increase</td>
<td>2</td>
<td>P to NP</td>
<td>Neutral to Neutral</td>
</tr>
<tr>
<td>QMU93213.1</td>
<td>R24C</td>
<td>Deleterious</td>
<td>-8</td>
<td>Decrease</td>
<td>7</td>
<td>P to P</td>
<td>Basic (strongly) to Neutral</td>
</tr>
<tr>
<td>QMT54534.1</td>
<td>R24L</td>
<td>Deleterious</td>
<td>-7</td>
<td>Decrease</td>
<td>9</td>
<td>P to NP</td>
<td>Basic (strongly) to Neutral</td>
</tr>
<tr>
<td>QKU34102.1</td>
<td>Y26H</td>
<td>Deleterious</td>
<td>-5</td>
<td>Decrease</td>
<td>8</td>
<td>P to P</td>
<td>Neutral to Basic (weakly)</td>
</tr>
<tr>
<td>QNI25281.1</td>
<td>V30A</td>
<td>Deleterious</td>
<td>-4</td>
<td>Decrease</td>
<td>9</td>
<td>NP to NP</td>
<td>Neutral to Neutral</td>
</tr>
<tr>
<td>QNC49349.1</td>
<td>V30L</td>
<td>Deleterious</td>
<td>-3</td>
<td>Decrease</td>
<td>4</td>
<td>NP to NP</td>
<td>Neutral to Neutral</td>
</tr>
<tr>
<td>QNA70543.1</td>
<td>L37F</td>
<td>Deleterious</td>
<td>-4</td>
<td>Decrease</td>
<td>7</td>
<td>NP to NP</td>
<td>Neutral to Neutral</td>
</tr>
<tr>
<td>QKV37245.1</td>
<td>T38I</td>
<td>Deleterious</td>
<td>-6</td>
<td>Decrease</td>
<td>5</td>
<td>P to P</td>
<td>Neutral to Neutral</td>
</tr>
<tr>
<td>QKI80176.1</td>
<td>L37P</td>
<td>Deleterious</td>
<td>-7</td>
<td>Decrease</td>
<td>8</td>
<td>NP to P</td>
<td>Neutral to Neutral</td>
</tr>
<tr>
<td>QNB17780.1</td>
<td>F35S</td>
<td>Deleterious</td>
<td>-8</td>
<td>Decrease</td>
<td>9</td>
<td>NP to P</td>
<td>Neutral to Neutral</td>
</tr>
<tr>
<td>QMT94417.1</td>
<td>D31Y</td>
<td>Deleterious</td>
<td>-9</td>
<td>Decrease</td>
<td>6</td>
<td>P to P</td>
<td>Acidic to Neutral</td>
</tr>
<tr>
<td>QLG75514.1</td>
<td>N22T</td>
<td>Deleterious</td>
<td>-6</td>
<td>Decrease</td>
<td>5</td>
<td>NP to NP</td>
<td>Neutral to Neutral</td>
</tr>
<tr>
<td>QLQ99973.1</td>
<td>I13M</td>
<td>Deleterious</td>
<td>-3</td>
<td>Decrease</td>
<td>1</td>
<td>P to P</td>
<td>Neutral to Neutral</td>
</tr>
<tr>
<td>QNG42985.1</td>
<td>P10S</td>
<td>Deleterious</td>
<td>-8</td>
<td>Decrease</td>
<td>8</td>
<td>NP to P</td>
<td>Neutral to Neutral</td>
</tr>
<tr>
<td>QLM54161.1</td>
<td>A8V</td>
<td>Deleterious</td>
<td>-4</td>
<td>Increase</td>
<td>3</td>
<td>NP to NP</td>
<td>Neutral to Neutral</td>
</tr>
<tr>
<td>QNG41574.1</td>
<td>I4L</td>
<td>Neutral</td>
<td>-2</td>
<td>Increase</td>
<td>1</td>
<td>NP to NP</td>
<td>Neutral to Neutral</td>
</tr>
</tbody>
</table>

Table 1. Twenty-two ORF10 proteins (SARS-CoV-2) and their corresponding mutations and predicted effects with changes in chemical properties. "PROVEAN score: If the PROVEAN score is equal to or below a predefined threshold (e.g., -2.5), the protein variant is predicted to have a "deleterious" effect. If the PROVEAN score is above the threshold, the variant is predicted to have a "neutral" effect. "RI: Reliability Index ranges from 0 to 9.

From Table 1, it was established that the majority of the diversified mutations are deleterious and cause the stability of the protein to decrease, thus indicating the amplification of intricate virulence of SARS-CoV-2.

Sequence Homology and Mutations of SARS-CoV-2 ORF10
It was reported that SARS-CoV-2 ORF10 is not homologous with other proteins in the NCBI depository. The SARS-CoV-2 ORF10 was blasted in the NCBI depository and no significant homology was detected for ORF10 SARS-CoV as well as Bat-CoV ORF10. Surprisingly, SARS-CoV-2 ORF10 showed 97.37% homology to Pangolin-CoV ORF10 (QIG55954.1 (Release date: 2020-05-18; Collection date: 2019-03-29; Geo-location: China; Host: Sunda pangolin (Manis javanica))) (Figure 2).
Only the serine (S) has been mutated to asparagine (N) at amino acid position 25 in SARS-CoV-2 ORF10 from the Pangolin-CoV ORF10 and the mutation is deleterious (PROVEAN score -3). Due to this mutation, the stability of the protein structure is predicted to be decreased and consequently the intricate virulence of SARS-CoV-2 will escalate.

Analysis of the per-residue intrinsic disorder predispositions of the ORF10 of SARS-CoV-2 and ORF10 proteins from SARS-CoV and Pangolin-CoV provide further evidence of their differences. Figure 3A represents the results of this analysis and shows that while ORF10 proteins from SARS-CoV-2 and Pangolin-CoV show very similar disorder profiles, the per-residue disorder propensity of the ORF10 protein from SARS-CoV is remarkably different, especially within the C-terminal half of this protein. This is in agreement with the results of other analyses conducted in this study.

![Figure 2](https://example.com/figure2.png)

**Figure 2.** Alignment of two ORF10 sequences (37 out of 38 identical residues) of Pangolin-CoV.

![Figure 3](https://example.com/figure3.png)

**Figure 3.** (A) Comparison of the intrinsic disorder profile of the reference ORF10 protein from SARS-CoV-2 (YP_009725255) from the NC_045512 SARS-CoV2 genome (China, Wuhan) (bold black curve) with those of ORF10 proteins from the Pangolin-CoV (QIG55954.1) and SARS-CoV TW-HP1 (UniProt ID: Q6SRY8).

(B) Analysis of the intrinsic disorder predisposition of the unique variants of SARS-CoV2 ORF10 in comparison with the reference ORF10 protein from SARS-CoV-2 (YP_009725255) from the NC_045512 SARS-CoV2 genome (China, Wuhan) (bold black curve). Analysis is conducted using PONDR-VSL2 algorithm\(^{23}\), which is one of the more accurate standalone disorder predictors\(^{24-26}\). A disorder threshold is indicated as a thin line (at score = 0.5). Residues/regions with the disorder scores > 0.5 are considered as disordered, whereas residues with disorder scores between 0.25 and 0.5 are considered highly flexible, and residues with disorder scores between 0.1 and 0.25 are taken as moderately flexible.

Figure 3B compares intrinsic disorder predispositions of the 24 unique variants of ORF10 protein from different isolates of SARS-CoV-2. It is seen that intrinsic disorder predispositions can vary significantly, especially within the C-terminal half of the protein. In fact, majority of substitutions found within the N-terminal region (residues 1-15; i.e., mutations G2D, I4L, V6I, A8V, P10S, I13M, and Y14C) have very little effect on the local intrinsic disorder predisposition of ORF10. On the other hand, ORF10 variants with the mutations within the C-terminal region (residues 20-38; i.e., mutations R20I/K, N22T, S23F, R24C/L, Y26H, A28V, V30A/L, D31Y, F35S, L37P/F, and T38I, as well as shortened QJR96431.1 variant, which is truncated due to a nonsense mutation at the position 29) typically show rather substantial variability in their local disorder predispositions. The most significant changes are observed within the “disorder hump” region (residues 20-30), intensity of which is increased in QKU54102.1 (Y26H), QNI25281.1 (V30A), and QNB17780.1 (F35S) ORF10 variants, whereas in the variants QMT54534.1 (R24L), QNC04532.1 (R20I), QMU93213.1 (R24C), and QMT97141.1 (S23F), this hump is either eliminated or noticeably
Interestingly, comparison of the Figure 3A and 3B shows that the variability in the disorder predisposition between many variants of the ORF10 protein from various SARS-CoV-2 isolates is noticeably greater than that between the reference ORF10 from SARS-CoV-2 and ORF10 from Pangolin-CoV. On the other hand, none of the SARS-CoV-2 ORF10 variants (with the exception for the truncated QJR96431.1 variant) has as disordered C-terminal half as the ORF10 protein from SARS-CoV does.

**Comparison of SARS-CoV2 ORF10 and Pangolin-CoV ORF10**

Considering the highest amount of sequence homology of ORF10 proteins of SARS-CoV-2 and Pangolin-CoV, we intended to discover the parity and disparity between the ORF10 proteins of SARS-CoV-2 and Pangolin-CoV. We, therefore, performed a multi-dimensional analysis of both ORF10 proteins from structural, physicochemical, biophysical and immunological aspects to understand the origin of SARS-CoV-2 from the ORF10 perspective.

**Figure 4.** (A): Basic properties of ORF10 proteins of SARS-CoV-2 and Pangolin-CoV; (B): Peptide and solvent accessibility properties of ORF10 proteins of SARS-CoV-2 and Pangolin-CoV

Exploration of similarities between SARS-CoV-2 and Pangolin ORF10 sequences (Figure 4A) revealed that neither of them had disulfide linkages. However, many differences were detected. The SARS-CoV-2 ORF10 protein was classified as an alpha-helical transmembrane protein (with probability 0.489) owing to the server ABTMpro as well as the presence of a majority of hydrophobic amino acids, whereas the Pangolin-CoV ORF10 sequence was predicted to be a non-transmembrane protein (with probability 0.513). Also, it was discerned that the predicted probability of antigenicity of SARS-CoV-2 ORF10 was slightly higher than that of Pangolin-CoV ORF10. It was predicted that both proteins are located in the capsid region of the virus as both of them have a positive distance score, with a higher score for Pangolin-CoV (0.1502) than for SARS-CoV-2 (0.1141).
To achieve deeper insights into the ORF10 proteins of SARS-CoV-2 and Pangolin-CoV, we characterized their secondary structure (Figure 4B) and found them to be very much similar except for a significant difference at the position 26, Tyr (Y), which for SARS-CoV-2 ORF10 is in the coil region whereas for Pangolin-CoV ORF10, it is located in the strand region. Most of the residues, 23 in SARS-CoV-2 ORF10 and 24 in Pangolin-CoV ORF10, are buried and consequently, the solubility of SARS-CoV-2 ORF10 is slightly higher than that of Pangolin-CoV.

![Figure 5](image-url)

**Figure 5.** (A): Physicochemical properties and hydropathy of ORF10 of SARS-CoV-2 and Pangolin-CoV; (B): Enzymes and numbers of associated cleavages and their positions.

After structural and fundamental property studies, a subsequent thorough analysis of the physicochemical properties of two ORF10 proteins of SARS-CoV-2 and Pangolin-CoV was performed, which unveiled the high similarity based on extinction coefficient, isoelectric point and net charge (Figure 5A). However, the molecular weight (4449.18 g/mol) of the SARS-CoV-2 ORF10 was higher compared to Pangolin-CoV ORF10 (4422.16 g/mol), due to the substitution of S (low molecular weight) of Pangolin-CoV to N (high molecular weight) of SARS-CoV-2. The enzyme cleavage sites for the SARS-CoV-2 and Pangolin-CoV ORF10 were also indistinguishable for all proteases (Figure 5B).
Protein intrinsic disorder analysis disclosed the presence of hotloops in both sequences within the same span of amino-acids (26-38). However, the presence of loops/coils (22-29) was a distinct characteristic of SARS-CoV-2 ORF10 and no such structures were observed for Pangolin ORF10 (Figure 6).

![Diagram showing disordered loops and hotloops of ORF10 of SARS-CoV-2 and Pangolin-CoV]

**Figure 6.** Disordered loops and hotloops of ORF10 of SARS-CoV-2 and Pangolin-CoV

Eleven distinct epitopes in the SARS-CoV-2 ORF10 were identified and analysed for binding affinity using PICKPOCKET across 12 HLA subtypes. The IDEB score was predicted using the IDEB immunogenicity tool. Eleven epitopes (marked in orange) from the Wuhan SARS-CoV-2 ORF10 sequence. Scores in red/blue show an increase/decrease concerning the score associated with nine epitopes. Green marked scores convey the immunogenicity value remaining unchanged.

![Table showing epitopes and IDEB scores for SARS-CoV-2 ORF10 across HLA subtypes]

**Figure 7.** Eleven distinct epitopes in the SARS-CoV-2 ORF10 were identified and analysed for binding affinity using PICKPOCKET across 12 HLA subtypes. The IDEB score was predicted using the IDEB immunogenicity tool. Eleven epitopes (marked in orange) from the Wuhan SARS-CoV-2 ORF10 sequence. Scores in red/blue show an increase/decrease concerning the score associated with nine epitopes. Green marked scores convey the immunogenicity value remaining unchanged.

Discussion

A detailed study of the ORF10 protein was carried out to evaluate its potential to yield to variants that could possibly alter viral pathogenicity. It was observed that each SARS-CoV-2 ORF10 sequence possesses one distinct mutation. Each of the twenty-two SARS-CoV-2 ORF10 variants is at a uniquely different position. None of these mutations in the SARS-CoV-2 ORF10, however, contributes to the determination of clades of SARS-CoV-2. Of all variants, a total of 13 variants were...
identified to possess mutations at amino acid positions 22-38 and in a region predicted to contain overlapping loops/coils and hot-loop regions of the ORF10 protein. All mutations were predicted to be deleterious with decreased effect on protein structure stability except S23F, which increased stability, denoting that these mutations play an active role in enhancing intrinsic propensity disorder (IPD) and allowing the protein to undergo more favorable interactions with other proteins. Two other mutations, I4L and V6I, were found to be in the MoRF region of ORF10, and which may also possibly contribute to the IPD as well.

The mutations at positions 20 and 24 were also significant due to their sensitivity for trypsin activity. Four ORF10 variants (QNC04532.1, QMT54534.1, QMU93213.1 and QLA48060.1) possess four mutations at these two positions. Among them, three variants harboring the mutations R20I, R24L and R24C provide trypsin resistance, while the fourth variant (QLA48060.1) with the R20K mutation is susceptible to protease degradation.

An amino acid homology of 97.37% was observed between SARS-CoV-2 ORF10 and Pangolin-CoV ORF10. Although most physicochemical and peptide properties are similar, the probability of antigenicity is greater for SARS-CoV-2 ORF10 than that of Pangolin-CoV ORF10 and consequently a stronger immune response is predicted for SARS-CoV-2 ORF10. A change from strand (Pangolin-CoV ORF10) to coil (SARS-CoV-2 ORF10) at position 26 (tyrosine (Y)), is predicted indicating the higher disordered state of the protein. A sequence with the Y26H mutation was also detected in SARS-CoV-2 ORF10, which showed that a hydrophobic amino acid was replaced by a hydrophilic amino acid, thus increasing the probability for more ionic interactions.

Analysis identified ORF10 mutations predicted to alter binding affinity to respective HLA alleles and to possibly correspondingly change the immunogenicity of SARS-CoV-2 ORF10. Eight ORF10 variants (containing one of the following mutations each G2D, I4L, I13M, Y14C, Y26H, F35S, L37S and L37P (Table 1)) accounted for 40% of total mutations and demonstrated decreased affinity for MHC class I, 25% of the variants (carrying mutations R20K, R20I, R24C, R24L and D31Y) predict for increased affinity, and 35% of the variants (carrying mutations V6I, A8V, P10S, S23F, A28V and V30A) contain both high and low binding affinity epitopes. This may indicates that mutations in ORF10 are predominantly decrease the affinity of epitopes to escape the host-immune system, while in the mixed cases the effect of increased affinity by mutations is nullified by the presence of mutations contributing to decreased affinity. For mutations showing only increased binding affinity epitopes, it is hypothesized that acquiring more than one mutation in a single sequence in the future will nullify them as well. In addition, the immunogenicity score prediction revealed that a large number of mutations had decreased or no effect and very few of them exhibited an increased immunogenicity score, which may be a possible strategy adopted by SARS-CoV-2 to evade the host-immune response. Six mutation-bearing sequences (QL57416.1, QMT97141.1, QLY88596.1, QNC49349.1, QMT54534.1, and QLG76514.1) were found to contain epitopes showing both high affinity binding for MHC class 1 and high immunogenicity, indicating that these epitopes can mount significant immune response and might serve as potential targets for vaccine candidates. More critical study in ORF10 SARS-CoV2 is necessary to monitor high frequency mutations that could change viral pathogenesis.

ORF10 protein of SARS-CoV-2 and Pangolin-CoV are similar. However, there are predicted notable differences detected between these two ORF10 proteins in terms of loop/coil structure, antigenicity, solubility, and in mutational diversification of SARS-CoV-2. These significant disagreements of various physicochemical, structural, immunological properties despite an amino acid homology (97.37%) between the ORF10 proteins of SARS-CoV-2 and Pangolin-CoV are quite surprising, and deserving of further study.

Data and Methods

Data acquisition

There were 11,288 complete genomes of SARS-CoV-2 available on the NCBI (National Center for Biotechnology Information) database, as of 28th August 2020. Each genome contains the ORF10 accessory protein and among them only 34 sequences were found to be unique. Among these unique ORF10 protein sequences, only 22 sequences possess only one missense mutation each and the remaining sequences possess ambiguous mutations. It is noted that, there was only one ORF10 sequence (QJR96431.1) which was truncated due to a nonsense mutation at amino acid position 29. The present study focused on these 23 ORF10 proteins (Table 2).
A reference ORF10 protein (YP_009725255.1) of the SARS-CoV-2 genome (NC_045512) from Wuhan, China was used to identify the mutations. The miscellany of ORF10 variants of SARS-CoV-2 is clearly observed in the sequence-based homology (Figure 8(A)) and phylogeny (Figure 8(B)).
Figure 8. (A): Multiple sequence alignment (MSA) of 24 SARS-CoV-2 ORF10 proteins; (B): Phylogeny of 24 SARS-CoV-2 ORF10 sequences.

Each ORF10 of SARS-CoV-2 is different from the Wuhan SARS-CoV-2 ORF10 sequence utilizing a single amino acid change at a distinct position. Noticeably, these positions (18) are widely varying from the position 2 to 38 for the 22 SARS-CoV-2 ORF10 variants.

Methods
Webserver based predictions
The prediction of various properties of ORF10 proteins was determined by several web servers which are briefly described as follows.
• For the prediction of the effect of identified mutations, the PROVEAN webserver was used and also for the structural effects of mutations, and another webserver, I-MUTANT, was used\textsuperscript{28–30}. The QUARK webserver was used for the prediction of secondary structure of ORF10 proteins\textsuperscript{31–33}.

• Given an amino acid sequence, the ABPTMpro webserver predicts whether the given sequence is a transmembrane protein. If the given sequence is a transmembrane protein, it further predicts the probabilities of the protein being an alpha-helix transmembrane protein or a Beta Barrel transmembrane protein. In addition, for various peptide property findings, the INNOVAGEN webserver was used\textsuperscript{34}.

• The DIpro can predict whether the given protein sequence contains a cysteine disulfide bond, based on 2D recurrent neural network, support vector machine, graph matching and regression algorithms\textsuperscript{35}.

• The protein antigenicity is predicted using the webserver ANTIGENpro, which is a sequence-based, alignment-free and pathogen-independent predictor. A two-stage architecture makes the probability of prediction based on multiple representations of the primary sequence and five machine learning algorithms\textsuperscript{36}. The intrinsic disorder prediction of a given protein sequence was made using the server DisEMBL\textsuperscript{37}.

• Epitopes of a given amino acid sequence were spotted and analyzed for binding affinity using across 12 HLA (human leukocyte antigen) subtypes (HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*24:02, HLA-A*26:01, HLA-B*07:02, HLA-B*08:01, HLA-B*27:05 B*39:01, B*40:01, B*58:01 and B*15:01). The IDEB (The Immune Epitope Database) score was predicted using the IDEB immunogenicity tool\textsuperscript{38,39}.

Evaluating the per-residue predisposition of various ORF10 proteins for intrinsic disorder
Per-residue disorder distribution within ORF10 protein sequences was evaluated by PONDR – VSL2\textsuperscript{23}, which is one of the more accurate standalone disorder predictors\textsuperscript{24–26}. The per-residue disorder predisposition scores are on a scale from 0 to 1, where values of 0 indicate fully ordered residues, and values of 1 indicate fully disordered residues. Values above the threshold of 0.5 are considered disordered residues, whereas residues with disorder scores between 0.25 and 0.5 are considered highly flexible, and residues with disorder scores between 0.1 and 0.25 are taken as moderately flexible.

References


Acknowledgements
Authors acknowledge the NCBI sequence (SARS-CoV-2 and Pangolin-CoV-2020) depositors.

Author contributions statement
SSH conceived the problem and experiment(s). DA, SG, SSH, VNU examined the mutations. SSH, PPC, DA, SG and VNU analysed the results. SSH wrote the primary draft of the article. All authors reviewed, edited and approved the final manuscript.

Conflict of interest
The authors have no conflicts of interest to declare.