

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

Sk. Sarif Hassan ^{a,*}, Diksha Attrish ^b, Shinjini Ghosh ^c, Pabitra Pal Choudhury ^d, Vladimir N. Uversky ^{e,*}, Alaa A.A. Aljabali ^f, Kenneth Lundstrom ^g, Bruce D. Uhal ^h, Nima Rezaei ^{i,z}, Murat Seyran ^j, Damiano Pizzol ^k, Parise Adadi ^l, Antonio Soares ^m, Tarek Mohamed Abd El-Aziz ^{m,n}, Ramesh Kandimalla ^{o,aa}, Murtaza M. Tambuwala ^p, Gajendra Kumar Azad ^q, Samendra P. Sherchan ^r, Wagner Baetas-da-Cruz ^s, Amos Lal ^t, Giorgio Palù ^u, Kazuo Takayama ^v, Ángel Serrano-Aroca ^w, Debmalya Barh ^x, Adam M. Brufsky ^y

^a Department of Mathematics, Pingla Thana Mahavidyalaya, Maligram 721140, India

^b Dr. B. R. Ambedkar Centre For Biomedical Research (ACBR), University of Delhi (North Campus), Delhi 110007, India

^c Department of Biophysics, Molecular Biology and Bioinformatics, University of Calcutta, Kolkata 700009, West Bengal, India

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

^e Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA

^f Department of Pharmaceutics and Pharmaceutical Technology, Yarmouk University-Faculty of Pharmacy, Irbid 566, Jordan

^g PanTherapeutics, Rte de Lavaux 49, CH1095 Lutry, Switzerland

^h Department of Physiology, Michigan State University, East Lansing, MI 48824, USA

ⁱ Research Center for Immunodeficiencies, Pediatrics Center of Excellence, Children's Medical Center, Tehran University of Medical Sciences, Tehran, Iran

^j Doctoral Studies in Natural and Technical Sciences (SPL 44), University of Vienna, Austria

^k Italian Agency for Development Cooperation - Khartoum, Sudan Street 33, Al Amarat, Sudan

^l Department of Food Science, University of Otago, Dunedin 9054, New Zealand

^m 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

ⁿ Zoology Department, Faculty of Science, Minia University, El-Minia 61519, Egypt

^o CSIR-Indian Institute of Chemical Technology, Uppal Road, Tarnaka, Hyderabad 500007, Telangana State, India

^p School of Pharmacy and Pharmaceutical Science, Ulster University, Coleraine BT52 1SA, Northern Ireland, UK

^q Department of Zoology, Patna University, Patna 800005, Bihar, India

^r Department of Environmental Health Sciences, Tulane University, New Orleans, LA 70112, USA

^s Translational Laboratory in Molecular Physiology, Centre for Experimental Surgery, College of Medicine, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

^t Division of Pulmonary and Critical Care Medicine, Mayo Clinic, Rochester, MN, USA

^u Department of Molecular Medicine, University of Padova, Via Gabelli 63, 35121 Padova, Italy

^v Center for iPS Cell Research and Application, Kyoto University, Kyoto 606-8397, Japan

^w Biomaterials and Bioengineering Lab, Centro de Investigación Traslacional San Alberto Magno, Universidad Católica de Valencia, San Vicente Mártir 46001, Valencia, Spain

^x Centre for Genomics and Applied Gene Technology, Institute of Integrative Omics and Applied Biotechnology (IIQAB), Purba Medinipur, India

^y University of Pittsburgh School of Medicine, Department of Medicine, Division of Hematology/Oncology, UPMC Hillman Cancer Center, Pittsburgh, PA, USA

^z Network of Immunity in Infection, Malignancy and Autoimmunity (NIIMA), Universal Scientific Education and Research Network (USERN), Stockholm, Sweden

^{aa} Department of Biochemistry, Kakatiya Medical College, Warangal, Telangana State, India

ARTICLE INFO

Article history:

Received 7 January 2021

Received in revised form 30 March 2021

Accepted 30 March 2021

Available online xxxx

Keywords:

ORF10

SARS-CoV-2

Pangolin-CoV-2020

Mutations

ABSTRACT

The current Coronavirus Disease 19 (COVID-19) pandemic, caused by Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) shows similar pathology to MERS and SARS-CoV, with a current estimated fatality rate of 1.4%. Open reading frame 10 (ORF10) is a unique SARS-CoV-2 accessory protein, which contains eleven cytotoxic T lymphocyte (CTL) epitopes each of nine amino acids in length. Twenty-two unique SARS-CoV-2 ORF10 variants have been identified based on missense mutations found in sequence databases. Some of these mutations are predicted to decrease the stability of ORF10 in silico physicochemical and structural comparative analyses were carried out on SARS-CoV-2 and Pangolin-CoV ORF10 proteins, which share 97.37% amino acid (aa) homology. Though there is a high degree of ORF10 protein similarity of SARS-CoV-2 and Pangolin-CoV, there are differences of these two ORF10 proteins related to their sub-structure (loop/coil region), solubility, antigenicity and shift from strand to coil at aa position

* Corresponding authors.

E-mail addresses: sarimif@gmail.com (S.S. Hassan), vuversky@usf.edu (V.N. Uversky), alaaj@yu.edu.jo (A.A.A. Aljabali), rezaei_nima@tums.ac.ir (N. Rezaei), parise.adadi@postgrad.otago.ac.nz (P. Adadi), soaresa@uthscsa.edu (A. Soares), mohamedt1@uthscsa.edu (T.M. Abd El-Aziz), ramesh.kandimalla@iict.res.in (R. Kandimalla), m.tambuwala@ulster.ac.uk (M.M. Tambuwala), gkazad@patnauniversity.ac.in (G.K. Azad), sshercha@tulane.edu (S.P. Sherchan), giorgio.palu@unipd.it (G. Palù), kazuo.takayama@cira.kyoto-u.ac.jp (K. Takayama), angel.serrano@ucv.es (Á. Serrano-Aroca), brufskyam@upmc.edu (A.M. Brufsky).

1. Introduction

The Coronavirus Disease 19 (COVID-19) pandemic has affected the whole world with more than 131 million people infected and 2.85 million fatalities worldwide as of April, 5, 2021 [1–5]. The high fatality rate of 9.7% for SARS-CoV and 37% for Middle East Respiratory Syndrome Coronavirus (MERS-CoV) in comparison to 1.4% for SARS-CoV-2 has made it vital to monitor mutations within proteins such as ORF10 that could influence viral pathogenicity [6–8]. SARS-CoV-2 is a positive-sense, single-stranded RNA virus with four structural, sixteen non-structural, and six accessory proteins [9]. ORF10 is the smallest accessory protein (38 aa) in SARS-CoV-2, which can distinguish the infection faster than PCR techniques [10]. ORF10, present at the C-terminal of the genome, is hypothesized to be a transposon, although being distinct from larger transposons [10,11]. ORF10 contains a MoRF region from amino acid residue 3 to 7, a protein interaction site, which enables the intrinsic protein to adopt a set of conformations connected with different proteins [12,13]. High-throughput analysis revealed that ORF10 could interact with many host proteins such as multiple members of the Cullin-ubiquitin-ligase complex, which is essential for viral pathogenesis despite its small structure [12,14–17]. Humans may not utilize any memory B and T cells elicited against other microorganisms to target ORF10 and fight SARS-CoV-2 [18]. No sequence homology was found with any protein in the NCBI protein depository. SARS-CoV-2 ORF10 was reported to have a 99.15% nucleotide similarity to Pangolin-CoV-2020 [19,20].

Here we explore mutations described for SARS-CoV-2 ORF10 variants, which in addition to their physiochemical and immunological properties; may possibly have an impact on pathogenesis. The SARS-CoV-2 and Pangolin-CoV ORF10 proteins are also compared.

2. Data and methods

2.1. Data acquisition

11,288 complete genomes of SARS-CoV-2 were retrieved from the National Center for Biotechnology Information (NCBI) database. There were 34 unique ORF10 accessory protein sequences. Only 22 sequences of these ORF10 proteins have one missense mutation, each with an ambiguous mutation in the remaining sequences. A nonsense mutation at position 29 resulted in a truncated ORF10 sequence (QJR96431.1) (Table 1).

The SARS-CoV-2 genome (NC_045512) reference ORF10 (YP_009725255.1) was used to identify mutations [21]. The ORF10 variants of SARS-CoV-2 were compared by sequence-based homology (Fig. 1A) and phylogeny (Fig. 1B).

SARS-CoV-2 ORF10 sequences utilize a single amino acid change at a distinct position. These positions (18) vary widely from position 2 to 38 for the 22 SARS-CoV-2 ORF10 variants.

2.2. Methods

2.2.1. Webserver based predictions

The prediction of various properties of ORF10 proteins was determined by several webserver.

- The web server PROVEAN was used to estimate the effect of known mutations and the structural effect of these mutations, and another

I-MUTANT webserver was used [22–24]. The QUARK webserver was used to predict the secondary structure of ORF10 proteins [25–27].

- The ABTMpro webserver predicts whether the protein is a transmembrane protein from its sequence, and further predicts alpha-helices and beta sheets. Besides, the INNOVAGEN webserver was used for peptide property predictions [28].
- The Dlprowebserver predicts whether the protein sequence contains disulfide bonds, based on 2D recurrent neural network, support vector machine, graph matching and regression algorithms [29].
- Protein antigenicity is predicted using the webserver ANTIGENpro. A two-stage architecture makes a prediction likelihood centered on several primary sequence representations and five machine learning algorithms [30]. The DisEMBL server uses the intrinsic distortion estimation of a single protein sequence [31].
- Epitopes of a given aa specific 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 Immune Epitope Database (IDEB) score was predicted using the IDEB immunogenicity tool [32,33].

2.2.2. 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 [34], which is an accurate standalone disorder predictor [35–37]. Predisposition scores for the per-residual condition are 0 to 1, where 0 indicates residues entirely arranged, and 1 indicates residues completely disordered. Residues of disorder scores between 0.25 and 0.5 were considered extremely versatile, and disorder scores between 0.1 and 0.25 were considered mildly versatile. Residues with values higher than 0.5 were considered disordered.

Table 1

Twenty-four distinct ORF10 protein IDs correlated with geolocation.

Accession	Geo_location	Collection_date
YP_009725255	China	2019-12
QIJ57416	USA: WA	2020
QIS29991	China: Hubei, Wuhan	2020-01-10
QJR96431	USA: CA	2020-03-13
QKU54102	USA: Washington, King County	2020-03-15
QLA48060	USA: NY	2020-03-24
QNG41574	USA: Minnesota	2020-03-25
QKV08176	USA: Washington, King County	2020-03-26
QKV37245	Australia: Northern Territory	2020-03-27
QNI23218	USA: Virginia	2020-04
QLG99793	USA: CA	2020-04-16
QLY88596	USA: GA	2020-04-27
QNC04532	USA	2020-04-29
QNI25281	USA: Virginia	2020-05
QLI33453	USA	2020-05-12
QNC49349	Pakistan	2020-05-15
QMT94417	USA: Washington, Yakima County	2020-05-27
QMT54534	USA: Washington, Yakima County	2020-06-17
QLG76514	Australia: Victoria	2020-06-20
QNG42985	USA: FL	2020-06-23
QMT97141	USA: FL	2020-06-30
QNB17780	Bangladesh	2020-07-07
QMU93213	USA: Wisconsin, Dane county	2020-07-13
QNA70543	Bangladesh	2020-07-19

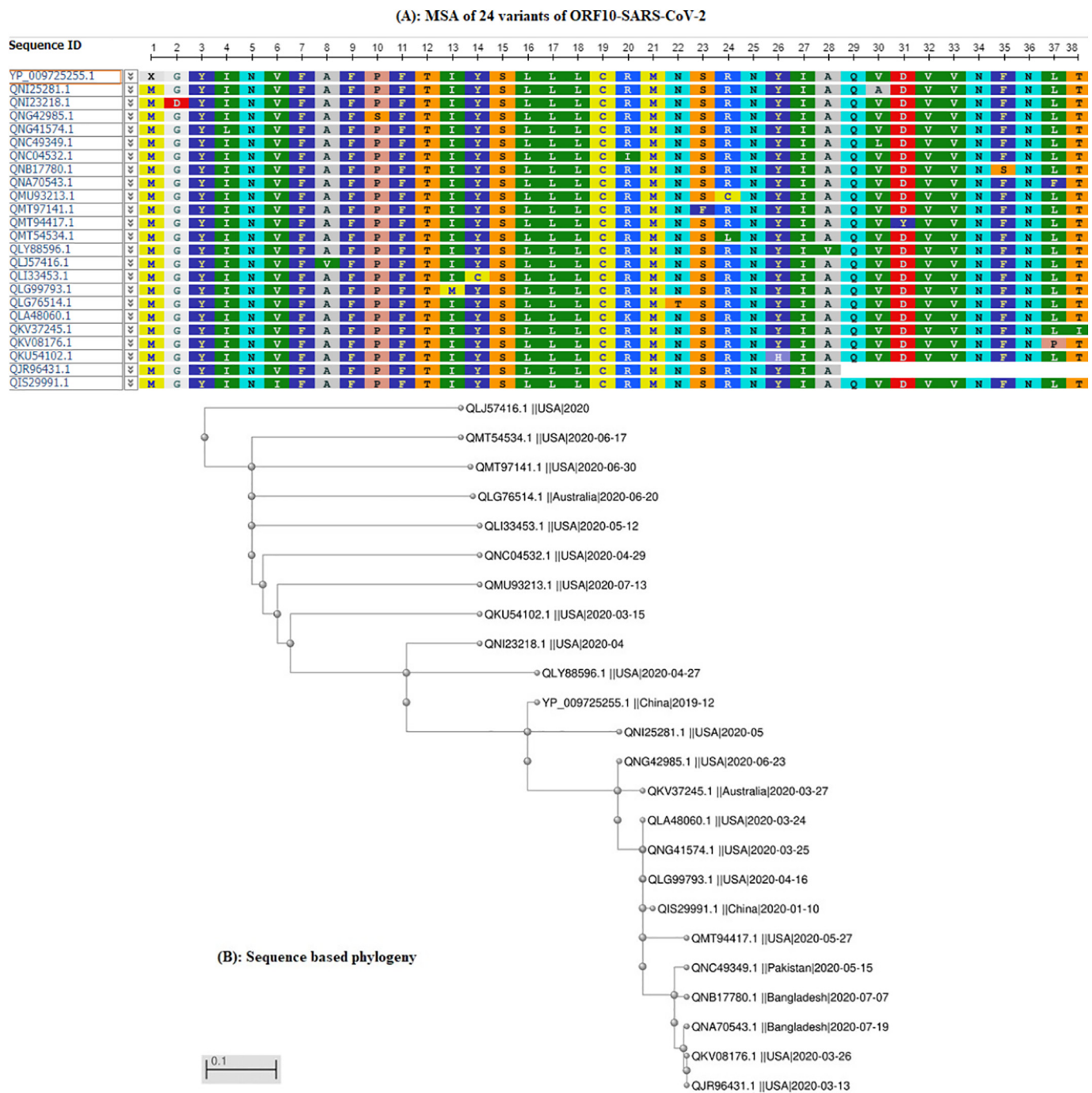


Fig. 1. (A): Multiple sequence alignment (MSA) of 24 SARS-CoV-2 ORF10 proteins; (B): phylogeny of 24 SARS-CoV-2 ORF10 sequences.

3. Results

3.1. SARS-CoV-2 ORF10 mutations

Each ORF10 sequence has been aligned using the p-blast protein and omega blast suites of the NCBI, and missense mutations were identified in (Fig. 2A) [38,39]. Conserved and non-conserved residues in ORF10 proteins are identified and marked in different colors (Fig. 2B). MoRF (YINVF) is also predicted using MoRFchibi server for the ORF10 Wuhan sequence [40].

There are 22 unique mutations in 22 SARS-CoV-2 ORF10 variants. These missense mutations are found in the entire ORF10 sequence from the aa position 2 to 38. Arginine (R), valine (V), and leucine (L) are substituted to more than one aa at fixed positions (marked

magenta in Fig. 2B). The largest conserved region across all the 24 ORF10 variants is 'SLLLC' at positions 15–19.

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

It was established that the most diversified mutations are deleterious and resulted in decreased protein stability, thus indicating the amplification of intricate virulence of SARS-CoV-2 (Table 2).

3.2. Sequence homology and mutations of SARS-CoV-2 ORF10

SARS-CoV-2 ORF10 does not show homology with other proteins in the NCBI depository including the Bat CoV ORF10 [10]. Surprisingly, SARS-CoV-2 ORF10 showed 97.37% homology to Pangolin-CoV ORF10

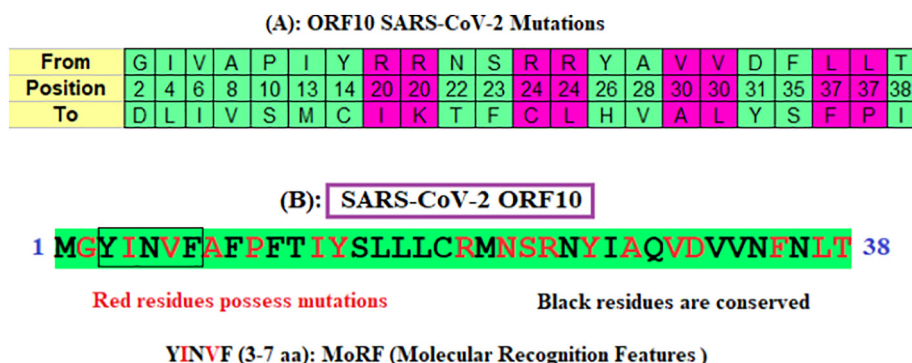


Fig. 2. (A): Mutations and their aa positions in ORF10 proteins; (B): conserved, mutated residues and molecular recognition features of ORF10 (YP_009725255) of SARS-CoV-2.

(QIG55954.1 (release date: 2020-05-18; collection date: 2019-03-29; geo-location: China; host: *Sunda pangolin* (*Manis javanica*))) (Fig. 3) [20].

The only difference in the ORF10 sequences is between the serine (S) in the Pangolin-CoV and asparagine (N) in the SARS-CoV-2 at position 25, which according to the PROVEAN score (−3) is deleterious. Subsequently, the protein structural stability is predicted to decrease.

Analysis of the per-residue intrinsic disorder predispositions of the ORF10 of SARS-CoV-2 and Pangolin-CoV provide evidence of their differences. The findings indicate that while the ORF10 SARS-CoV-2 and Pangolin-CoV proteins have very close disorder profiles, the residual disorder tendency of SARS-CoV ORF10 differs significantly, especially within its C-terminal half (Fig. 4A).

Fig. 4B compares intrinsic disorder predispositions of the 24 unique variants of ORF10 protein from different SARS-CoV-2 isolates. 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 flattened. Interestingly, comparison of Fig. 4A and B 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.

3.3. Comparison of SARS-CoV-2 ORF10 and Pangolin-CoV ORF10

Provided that SARS-CoV-2 and Pangolin-CoV ORF10 have the highest sequence homology, we aimed to detect parity and difference between SARS-CoV-2 and Pangolin-CoV. Therefore, we 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.

The correlations between sequences of SARS-CoV-2 and Pangolin-CoV ORF10 showed no disulfide connections (Fig. 5A). However, there

Table 2

Twenty-two ORF10 proteins, their corresponding mutations, and predicted effects with chemical property changes.

Accession ID	Mutations	Type of mutations	PROVEAN score ^a	Effect of mutations on structure	RI ^b	Polarity changes	Charge
QNI23218.1	G2D	Deleterious	−7	Decrease	7	NP to P	Neutral to acidic
QIS29991.1	V6I	Neutral	−1	Decrease	7	NP to NP	Neutral to neutral
QLI33453.1	Y14C	Deleterious	−9	Decrease	2	P to P	Neutral to neutral
QNC04532.1	R20I	Deleterious	−8	Decrease	3	NP to NP	Basic (strongly) to neutral
QLA48060.1	R20K	Deleterious	−3	Decrease	8	NP to P	Basic (strongly) to basic
QMT97141.1	S23F	Deleterious	−6	Increase	2	P to NP	Neutral to neutral
QMU93213.1	R24C	Deleterious	−8	Decrease	7	P to P	Basic (strongly) to neutral
QMT54534.1	R24L	Deleterious	−7	Decrease	9	P to NP	Basic (strongly) to neutral
QKU54102.1	Y26H	Deleterious	−5	Decrease	8	P to P	Neutral to basic (weakly)
QNI25281.1	V30A	Deleterious	−4	Decrease	9	NP to NP	Neutral to neutral
QNC49349.1	V30L	Deleterious	−3	Decrease	4	NP to NP	Neutral to neutral
QNA70543.1	L37F	Deleterious	−4	Decrease	7	NP to NP	Neutral to neutral
QKV37245.1	T38I	Deleterious	−6	Decrease	5	P to NP	Neutral to neutral
QKV08176.1	L37P	Deleterious	−7	Decrease	8	NP to NP	Neutral to neutral
QNB17780.1	F35S	Deleterious	−8	Decrease	9	NP to P	Neutral to neutral
QMT94417.1	D31Y	Deleterious	−9	Decrease	6	P to P	Acidic to neutral
QLY88596.1	A28V	Deleterious	−4	Decrease	5	NP to NP	Neutral to neutral
QLG76514.1	N22T	Deleterious	−6	Decrease	1	P to P	Neutral to neutral
QLG99793.1	I13M	Deleterious	−3	Decrease	8	NP to NP	Neutral to neutral
QNG42985.1	P10S	Deleterious	−8	Decrease	8	NP to P	Neutral to neutral
QLJ57416.1	A8V	Deleterious	−4	Increase	3	NP to NP	Neutral to neutral
QNG41574.1	I4L	Neutral	−2	Increase	1	NP to NP	Neutral to neutral

^a 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.

^b RI: Reliability Index ranges from 0 to 9.

ORF10-Pangolin-CoV
ORF10-SARS-CoV-2

MGYINVFAPFPTIYSLLLCRMNSRSYIAQVDVNFNLT 38
MGYINVFAPFPTIYSLLLCRMNSRNYIAQVDVNFNLT 38

Fig. 3. Alignment of two ORF10 sequences (37 out of 38 identical residues) of Pangolin-CoV.

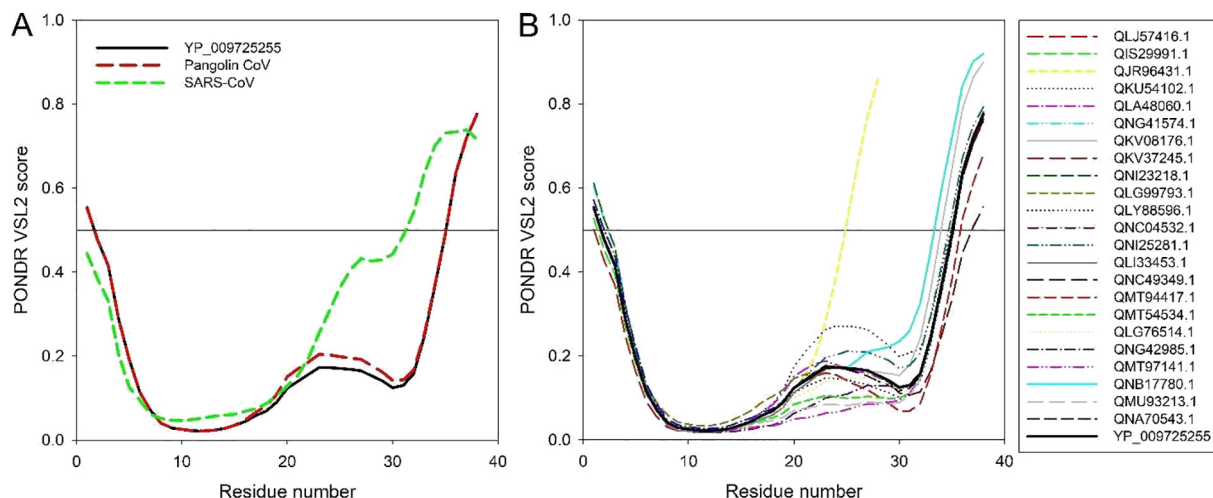


Fig. 4. (A) Comparison of the intrinsic disorder profile of the reference ORF10 protein from SARS-CoV-2 (YP_009725255) from the NC_045512 SARS-CoV-2 genome (China, Wuhan) (black curve) with ORF10 proteins from the Pangolin-CoV (QIG55954.1) and SARS-CoV TW-HPI (UniProt ID: Q6SRV8). (B) Predisposition of intrinsic disease of SARS-CoV-2 ORF10 single variants relative to SARS-CoV-2 (YP 009725255) ORF10 protein of the NC_045512 SARS-CoV-2 genome (China, Wuhan) (black curve). The analysis is conducted using the PONDR-VSL2 algorithm [34], one of the more accurate standalone disorder predictors [35–37]. A thin line (score = 0.5) is the threshold separating order from disorder. Residues with the predicted disorder scores ≥ 0.5 are considered as disordered, residues with the disorder scores ranging between 0.25 and 0.5 are flexible, whereas disorder scores below 0.25 correspond to ordered residues.

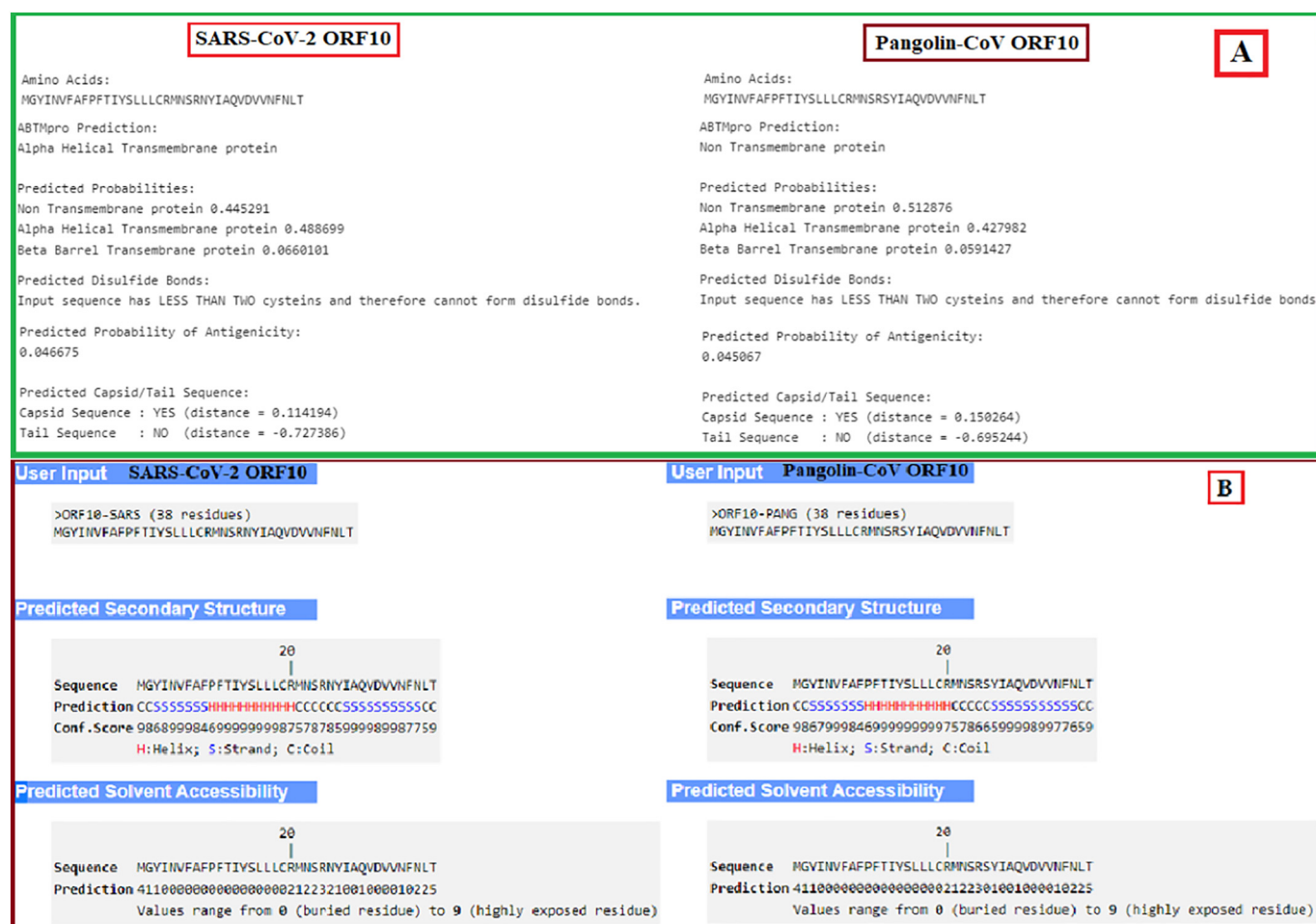


Fig. 5. (A): Basic properties of ORF10 proteins of SARS-CoV-2 and Pangolin-CoV; (B): peptide and solvent accessibility properties of ORF10 proteins of SARS-CoV-2 and Pangolin-CoV.

are several variations. Due to the ABTMpro server and the inclusion of the bulk of hydrophobic amino acids, the SARS-CoV-2 ORF10 was predicted to be an alpha-helical transmembrane protein (probability 0.489) while the Pangolin-CoV ORF10 series was predicted to be a non-transmembrane protein (probability 0.513). The predicted probability of antigenicity of SARS-CoV-2 ORF10 was slightly higher than that of Pangolin-CoV ORF10. Both proteins were expected to be located in the capsid area of the virus as they both show a positive distance with Pangolin-CoV (0.1502) at a higher rate than SARS-CoV-2 (0.1141).

We characterized their secondary structure (Fig. 5B) for detailed insights into ORF10 proteins from SARS-CoV-2 and Pangolin-CoV and found that these are almost the same except for a significant variation of tyrosine (Y) at position 26 for SARS-CoV-2 ORF10. In SARS-CoV-2 ORF10, 23 of the residues are buried, and the solubility is significantly greater in 24 residues in SARS-CoV-2 ORF10 compared to Pangolin-CoV.

Subsequent in-depth physicochemical properties study of two of ORF10 SARS-CoV-2 and Pangolin-CoV proteins revealed the high similarity of extinction, isoelectric point, and net charging dependent on structural and fundamental propriety studies (Fig. 6A). However, in contrast to Pangolin-CoV ORF10 (4422 g/mol), the molecular weight of SARS-CoV-2 ORF10 was higher because of the replacement of Pangolin-CoV S (low molecular weight) for N (high molecular weight) for SARS-CoV-2. The enzyme cleavage sites for the SARS-CoV-2 and Pangolin-CoV ORF10 were also indistinguishable for all proteases (Fig. 6B).

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-CoV ORF10 (Fig. 7).

We studied and identified nine amino acid epitopes 11 cytotoxic T-lymphocytes (CTLs), from the SARS-CoV-2 ORF10 series, in all 12 HLA subtypes to demonstrate the immunogenic properties of ORF10 and their associated epitopic mutations (Fig. 8). The scores were contrasted with the initial epitopes, thereby predicting that the binding affinity for Class-I MHC molecules will increase/decline due to mutations. All eleven epitopes and mutational epitopes have been analyzed using the IDEB tool to take their immunogenicity into account.

4. Discussions

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

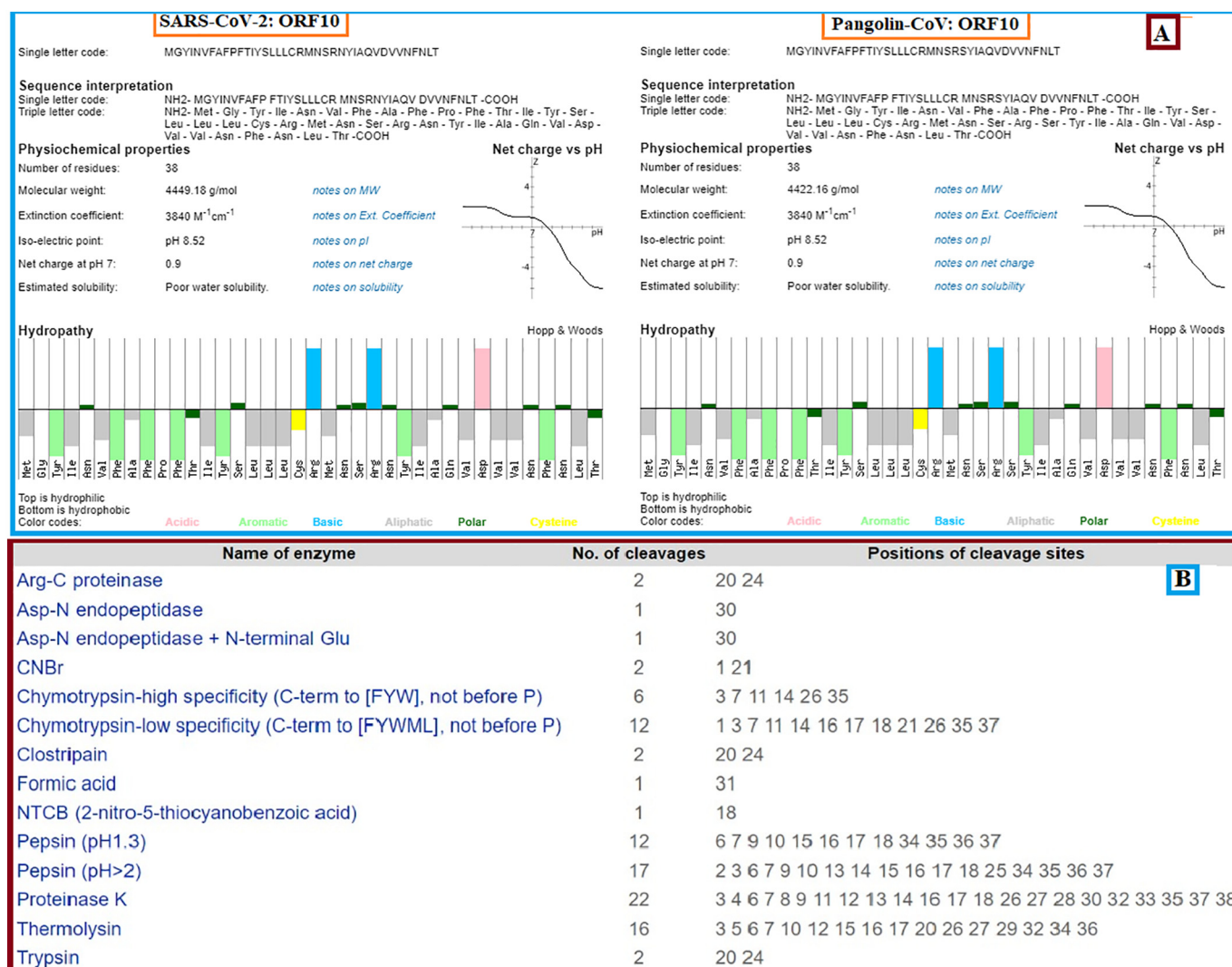


Fig. 6. (A): Physicochemical properties and hydropathy of ORF10 of SARS-CoV-2 and Pangolin-CoV; (B): enzymes and numbers of associated cleavages and their positions.

Pangolin-CoV-ORF10	SARS-CoV-2 ORF10
Disordered by Loops/coils definition	Disordered by Loops/coils definition
>none_LOOPS none mgyinvfafp ftiyslllcr mnsrsyiaqv dvvnfnlt	>none_LOOPS 22-29 mgyinvfafp ftiyslllcr mNSRNYIAQV dvvnfnlt
Disordered by Hot-loops definition	Disordered by Hot-loops definition
>none_HOTLOOPS 26-38 mgyinvfafp ftiyslllcr mnsrsYIAQV DVVNFNLT	>none_HOTLOOPS 26-38 mgyinvfafp ftiyslllcr mnsrnYIAQV DVVNFNLT

Fig. 7. Disordered loops and hotloops of ORF10 of SARS-CoV-2 and Pangolin-CoV.

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 2)) 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 indicate that mutations in ORF10 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 (QLJ57416.1, QMT97141.1, QLY88596.1, QNC49349.1, QMT54534.1, and QLG76514.1) were found to contain epitopes showing both high affinity binding for

Position	Epitope	HLA subtypes	1-log50k (affinity score)	IDEB Score	Position	Epitope	HLA subtypes	1-log50k (affinity score)	IDEB Score	Position	Epitope	HLA subtypes	1-log50k (affinity score)	IDEB Score
1–9	MGYINVFAF	HLA-B*58:01	0.514	0.28694	6–14	VFAFFPTIY	HLA-B*15:01	0.311	0.34042	11–19	FTIYSLLC	HLA-A*03:01	0.545	-0.1479
	MDYINVFAF		0.388	0.28694		IFAFPTIY		0.337	0.34042		FTMYSLLC		0.044	-0.2416
	MGYLNFAF		0.504	0.14186		VVFVPTIY		0.295	0.34112		FTICSLLC		0.228	-0.1984
	MGYINFAF		0.5	0.37336		VFAFSFTIY		0.331	0.19012		RMNSRNYIA		0.582	-0.0496
	MGYINVVF		0.517	0.2882		VFAFFPTMY		0.285	0.16006		IMNSRNYIA		0.596	-0.0496
3–11	YINVEAFP	HLA-B*58:01	0.474	0.28259	8–16	VFAFFPTIC	HLA-A*24:02	0.108	0.34042	20–28	RMNSRNYIA	HLA-A*02:01	0.602	-0.0496
	YLNVEAFP		0.447	0.28259		AFPTIYSL		0.483	0.1775		RMNSRNYIA		0.568	0.23465
	YINIFAFP		0.468	0.37492		VFPFTIYSL		0.512	0.1775		RMNSRNYIA		0.61	0.06301
	YINVEVFP		0.477	0.28462		AFSFTIYSL		0.406	0.1274		RMNSLNYIA		0.607	-0.1108
	YINVEAFSE		0.491	0.19241		AFPTIYSL		0.47	-0.1131		RMNSRNYIA		0.548	-0.0192
5–13	NVEAFPFTI	HLA-A*02:01	0.533	0.30241	9–17	AFPTICSL	HLA-B*07:02	0.412	0.13512		RMNSRNYIV		0.719	-0.0496
	NIFAFPFTI		0.535	0.30241		FPFTIYSL		0.595	0.05708		RMTSRNYIA		0.53	-0.03492
	NVVFVFTI		0.505	0.30458		FSFTIYSL		0.282	0.05708					
	NVEAFSFTI		0.581	0.15712		FPFTIYSL		0.581	-0.2435					
	NVEAFPFTM		0.445	0.30241		FPFTIYSL		0.591	0.00981					
Position	Epitope	HLA subtypes	1-log50k (affinity score)	IDEB Score	Position	Epitope	HLA subtypes	1-log50k (affinity score)	IDEB Score	Position	Epitope	HLA	Affinity Index	IDEB Score
22–30	NSRNYIAQV	HLA-A*02:01	0.321	0.09546	27–35	IAQVDVNF	HLA-B*58:01	0.458	0.09546	29–37	QVDVNFNL	HLA-A*02:01	0.437	0.17797
	NFRNYIAQV		0.36	0.09731		IVQVDVNF		0.369	0.09546		QADVNFNL		0.29	0.17787
	NSCNYIAQV		0.442	0.06301		IAQADVNF		0.494	0.09329		QLDVNFNL		0.613	0.17787
	NSLNYIAQV		0.535	0.07691		IAQLDVNF		0.442	0.04276		QVYVNFNL		0.508	0.16947
	NSRNHIAQV		0.292	0.13241		IAQVYVNF		0.499	0.07026		QVDVNSNL		0.324	-0.0606
	NSRNYIVQV		0.329	0.09913		IAQVDVNS		0.193	0.09546		QVDVNFNF		0.221	0.17787
	NSRNYIAQA		0.184	0.09731							QVDVNFNF		0.244	0.17787
	NSRNYIAQL		0.592	0.09731										
	TSRNYIAQV		0.316	0.09731										

Fig. 8. In 12 HLA subtypes, 11 distinct epitopes were described and analyzed in the SARS-CoV-2 ORF10 for binding affinity using PICKPOCKET. Using the IDEB tool, the IDEB value was estimated. Eleven epitopes of the Wuhan SARS-CoV-2 ORF10 series (marked in orange). Red/blue scores indicate an increase/decline of the score for nine epitopes. The immunogenicity attribute remains unchanged with significant green values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

MHC class I and high immunogenicity, indicating that these epitopes can mount significant immune response and might serve as potential targets for vaccine candidates. More critical studies in ORF10 SARS-CoV-2 are 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 of 97.37% between the ORF10 proteins of SARS-CoV-2 and Pangolin-CoV are quite surprising, and deserving of further study.

A question exists as to the expression of ORF10 both in vivo and in virally infected cell lines. In a small case series of two subjects, a SARS-CoV-2 strain with a truncation mutation of ORF10 was associated with mild disease and in vitro this strain replicated with the same efficiency as strains with non-truncated ORF10 [41]. It should be noted that both individuals infected with this strain had mild disease, and the VeroE6 cells used for viral culture lack native interferon (IFN) production. Evasion of IFN production as well as IFN signaling appears to be important in the pathogenesis of SARS-CoV-2, and interference with IFN induction or interferon sensitivity by ORF10 cannot be ruled out by these experiments [42]. In this vein, ORF10 expression is found in immune cells of subjects infected with SARS-CoV-2, and expression levels of ORF 10 are associated with disease severity [43]. Finally, T cells of acute and convalescent subjects with SARS-CoV-2 infection react to ORF10 in vitro [44]. Taken together, these data suggest that ORF10 is indeed expressed during infection and may be involved in disease severity. The analysis of mutations described in our study through various in vivo and in vitro models of COVID-19 infection and disease severity should be explored, and appear crucial in our understanding of disease pathogenesis.

CRedit authorship contribution statement

SSH conceived the problem and experiment(s). DA, SG, SSH, VNU examined the mutations. SSH, PPC, DA, SG and VNU analyzed the results. SSH wrote the primary draft of the article. AAAA and KL have made major editing to reach a final form. All authors critically reviewed, edited, and approved the final manuscript.

Declaration of competing interest

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

References

- [1] W.-H. Kong, Y. Li, M.-W. Peng, D.-G. Kong, X.-B. Yang, L. Wang, M.-Q. Liu, Sars-cov-2 detection in patients with influenza-like illness, *Nat. Microbiol.* 5 (5) (2020) 675–678.
- [2] B. Ju, Q. Zhang, J. Ge, R. Wang, J. Sun, X. Ge, J. Yu, S. Shan, B. Zhou, S. Song, et al., Human neutralizing antibodies elicited by sars-cov-2 infection, *Nature* 584 (7819) (2020) 115–119.
- [3] S. Yousefzadegan, N. Rezaei, Case report: death due to covid-19 in three brothers, *Am. J. Trop. Med. Hyg.* 102 (6) (2020) 1203–1204.
- [4] Z. Xu, L. Shi, Y. Wang, J. Zhang, L. Huang, C. Zhang, S. Liu, P. Zhao, H. Liu, L. Zhu, et al., Pathological findings of covid-19 associated with acute respiratory distress syndrome, *Lancet Respir. Med.* 8 (4) (2020) 420–422.
- [5] G. Li, Y. Fan, Y. Lai, T. Han, Z. Li, P. Zhou, P. Pan, W. Wang, D. Hu, X. Liu, et al., Coronavirus infections and immune responses, *J. Med. Virol.* 92 (4) (2020) 424–432.
- [6] V.M. Corman, N.L. Ithete, L.R. Richards, M.C. Schoeman, W. Preiser, C. Drosten, J.F. Drexler, Rooting the phylogenetic tree of middle east respiratory syndrome coronavirus by characterization of a conspecific virus from an african bat, *J. Virol.* 88 (19) (2014) 11297–11303.
- [7] D.S. Hui, E.I. Azhar, T.A. Madani, F. Ntoumi, R. Kock, O. Dar, G. Ippolito, T.D. Mchugh, Z.A. Memish, C. Drosten, et al., The continuing 2019-ncov epidemic threat of novel coronaviruses to global health—the latest 2019 novel coronavirus outbreak in Wuhan, China, *Int. J. Infect. Dis.* 91 (2020) 264–266.
- [8] W. Yang, S. Kandula, M. Huynh, S.K. Greene, G. Van Wye, W. Li, H.T. Chan, E. McGibbon, A. Yeung, D. Olson, et al., Estimating the infection-fatality risk of sars-cov-2 in New York city during the spring 2020 pandemic wave: a model-based analysis, *Lancet Infect. Dis.* 21 (2020) 203–212.
- [9] C. S. G. of the International, et al., The species severe acute respiratory syndrome-related coronavirus: classifying 2019-ncov and naming it sars-cov-2, *Nat. Microbiol.* 5 (4) (2020) 536.
- [10] T. Koyama, D. Platt, L. Parida, Variant analysis of sars-cov-2 genomes, *Bull. World Health Organ.* 98 (7) (2020) 495.
- [11] K. Kiyotani, Y. Toyoshima, K. Nemoto, Y. Nakamura, Bioinformatic prediction of potential t cell epitopes for sars-cov-2, *J. Hum. Genet.* 65 (7) (2020) 569–575.
- [12] R. Giri, T. Bhardwaj, M. Shegane, B.R. Gehi, P. Kumar, K. Gadhave, C.J. Oldfield, V.N. Uversky, Understanding covid-19 via comparative analysis of dark proteomes of sars-cov-2, human sars and bat sars-like coronaviruses, *Cell. Mol. Life Sci.* (2020) 1–34.
- [13] V.N. Uversky, Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes, *Chem. Soc. Rev.* 40 (3) (2011) 1623–1634.
- [14] D.E. Gordon, G.M. Jang, M. Bouhaddou, J. Xu, K. Obernier, K.M. White, M.J. O'Meara, V.V. Rezelj, J.Z. Guo, D.L. Swaney, et al., A sars-cov-2 protein interaction map reveals targets for drug repurposing, *Nature* (2020) 1–13.
- [15] J. Diaz, Sars-cov-2 molecular network structure, *Front. Physiol.* 11 (2020) 870.
- [16] Q. Liang, J. Li, M. Guo, X. Tian, C. Liu, X. Wang, X. Yang, P. Wu, Z. Xiao, Y. Qu, et al., Virus-host interactome and proteomic survey of pmbs from covid-19 patients reveal potential virulence factors influencing sars-cov-2 pathogenesis, *Med.* 2 (1) (2020) 99–112.
- [17] R. Cagliani, D. Forni, M. Clerici, M. Sironi, Coding potential and sequence conservation of sars-cov-2 and related animal viruses, *Infect. Genet. Evol.* (2020) 104353.
- [18] N. Le Bert, A.T. Tan, K. Kunasegaran, C.Y. Tham, M. Hafezi, A. Chia, M.H.Y. Chng, M. Lin, N. Tan, M. Linster, et al., Sars-cov-2-specific t cell immunity in cases of covid-19 and sars, and uninfected controls, *Nature* 584 (7821) (2020) 457–462.
- [19] D. Kim, J.-Y. Lee, J.-S. Yang, J.W. Kim, V.N. Kim, H. Chang, The architecture of sars-cov-2 transcriptome, *Cell* 181 (4) (2020) 914–921.
- [20] P. Liu, J.-Z. Jiang, X.-F. Wan, Y. Hua, L. Li, J. Zhou, X. Wang, F. Hou, J. Chen, J. Zou, et al., Are pangolins the intermediate host of the 2019 novel coronavirus (sars-cov-2)? *PLoS Pathog.* 16 (5) (2020), e1008421.
- [21] X. Wu, Y. Cai, X. Huang, X. Yu, L. Zhao, F. Wang, Q. Li, S. Gu, T. Xu, Y. Li, et al., Co-infection with sars-cov-2 and influenza a virus in patient with pneumonia, *China, Emerg. Infect. Dis.* 26 (6) (2020) 1324.
- [22] Y. Choi, G.E. Sims, S. Murphy, J.R. Miller, A.P. Chan, Predicting the functional effect of amino acid substitutions and indels, *PLoS One* 7 (10) (2012) e46688.
- [23] Y. Choi, A fast computation of pairwise sequence alignment scores between a protein and a set of single-locus variants of another protein, *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine 2012*, pp. 414–417.
- [24] Y. Choi, A.P. Chan, Proven web server: a tool to predict the functional effect of amino acid substitutions and indels, *Bioinformatics* 31 (16) (2015) 2745–2747.
- [25] D. Xu, Y. Zhang, Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field, *Proteins Struct. Funct. Bioinformatics* 80 (7) (2012) 1715–1735.
- [26] D. Xu, Y. Zhang, Toward optimal fragment generations for ab initio protein structure assembly, *Proteins Struct. Funct. Bioinformatics* 81 (2) (2013) 229–239.
- [27] 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.
- [28] J. Cheng, A.Z. Randall, M.J. Sweredoski, P. Baldi, Scratch: a protein structure and structural feature prediction server, *Nucleic Acids Res.* 33 (suppl_2) (2005) W72–W76.
- [29] J. Cheng, H. Saigo, P. Baldi, Large-scale prediction of disulphide bridges using kernel methods, two-dimensional recursive neural networks, and weighted graph matching, *Proteins Struct. Funct. Bioinformatics* 62 (3) (2006) 617–629.
- [30] C.N. Magnan, M. Zeller, M.A. Kayala, A. Vigil, A. Randall, P.L. Felgner, P. Baldi, High-throughput prediction of protein antigenicity using protein microarray data, *Bioinformatics* 26 (23) (2010) 2936–2943.
- [31] R. Linding, L.J. Jensen, F. Diella, P. Bork, T.J. Gibson, R.B. Russell, Protein disorder prediction: implications for structural proteomics, *Structure* 11 (11) (2003) 1453–1459.
- [32] H. Zhang, O. Lund, M. Nielsen, The pickpocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to mhc-peptide binding, *Bioinformatics* 25 (10) (2009) 1293–1299.
- [33] R. Vita, S. Mahajan, J.A. Overton, S.K. Dhanda, S. Martini, J.R. Cantrell, D.K. Wheeler, A. Sette, B. Peters, The Immune Epitope Database (IEDB): 2018 update, *Nucleic Acids Res.* 47 (D1) (2019) D339–D343.
- [34] Z. Obradovic, K. Peng, S. Vucetic, P. Radivojac, A.K. Dunker, Exploiting heterogeneous sequence properties improves prediction of protein disorder, *Proteins Struct. Funct. Bioinformatics* 61 (S7) (2005) 176–182.
- [35] F. Meng, V.N. Uversky, L. Kurgan, Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions, *Cell. Mol. Life Sci.* 74 (17) (2017) 3069–3090.
- [36] Z.-L. Peng, L. Kurgan, Comprehensive comparative assessment of in-silico predictors of disordered regions, *Curr. Protein Pept. Sci.* 13 (1) (2012) 6–18.
- [37] X. Fan, L. Kurgan, Accurate prediction of disorder in protein chains with a comprehensive and empirically designed consensus, *J. Biomol. Struct. Dyn.* 32 (3) (2014) 448–464.
- [38] M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis, T.L. Madden, Ncbi blast: a better web interface, *Nucleic Acids Res.* 36 (suppl_2) (2008) W5–W9.

- [39] 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.
- [40] N. Malhis, M. Jacobson, J. Gsponer, Morfchibi system: software tools for the identification of morfs in protein sequences, *Nucleic Acids Res.* 44 (W1) (2016) W488–W493.
- [41] K. Pancer, A. Milewska, K. Owczarek, A. Dabrowska, M. Kowalski, P.P. Łabaj, W. Branicki, M. Sanak, K. Pyrc, The sars-cov-2 orf10 is not essential in vitro or in vivo in humans, *PLoS Pathog.* 16 (12) (2020), e1008959.
- [42] H. Xia, Z. Cao, X. Xie, X. Zhang, J.Y.-C. Chen, H. Wang, V.D. Menachery, R. Rajsbaum, P.-Y. Shi, Evasion of type i interferon by sars-cov-2, *Cell Rep.* 33 (1) (2020), 108234.
- [43] T. Liu, P. Jia, B. Fang, Z. Zhao, Differential expression of viral transcripts from single-cell rna sequencing of moderate and severe covid-19 patients and its implications for case severity, *Front. Microbiol.* 11 (2020).
- [44] P. Bacher, E. Rosati, D. Esser, G.R. Martini, C. Saggau, E. Schiminsky, J. Dargvainiene, I. Schröder, I. Wieters, Y. Khodamoradi, et al., Low-avidity cd4+ t cell responses to sars-cov-2 in unexposed individuals and humans with severe covid-19, *Immunity* 53 (6) (2020) 1258–1271.