



Revelation of Potent Epitopes Present in Unannotated ORF Antigens of SARS-CoV-2 for Epitope-Based Polyvalent Vaccine Design Using Immunoinformatics Approach

Patil Pranita Uttamrao, Chakkarai Sathyaseelan[†], L. Ponoop Prasad Patro[†] and Thenmalarchelvi Rathinavelan^{*}

OPEN ACCESS

Edited by:

Anthony L. Cunningham,
Westmead Institute for Medical
Research, Australia

Reviewed by:

Kerrie Sandgren,
Westmead Institute for Medical
Research, Australia
Katie Lineburg,
The University of Queensland,
Australia

*Correspondence:

Thenmalarchelvi Rathinavelan
tr@bt.iith.ac.in

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Viral Immunology,
a section of the journal
Frontiers in Immunology

Received: 09 April 2021

Accepted: 31 July 2021

Published: 23 August 2021

Citation:

Uttamrao PP, Sathyaseelan C,
Patro LPP and Rathinavelan T (2021)
Revelation of Potent Epitopes Present
in Unannotated ORF Antigens of
SARS-CoV-2 for Epitope-Based
Polyvalent Vaccine Design Using
Immunoinformatics Approach.
Front. Immunol. 12:692937.
doi: 10.3389/fimmu.2021.692937

Department of Biotechnology, Indian Institute of Technology Hyderabad, Kandi, India

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) kills thousands of people worldwide every day, thus necessitating rapid development of countermeasures. Immunoinformatics analyses carried out here in search of immunodominant regions in recently identified SARS-CoV-2 unannotated open reading frames (uORFs) have identified eight linear B-cell, one conformational B-cell, 10 CD4+ T-cell, and 12 CD8+ T-cell promising epitopes. Among them, ORF9b B-cell and T-cell epitopes are the most promising followed by M.ext and ORF3c epitopes. ORF9b₄₀₋₄₈ (CD8+ T-cell epitope) is found to be highly immunogenic and antigenic with the highest allele coverage. Furthermore, it has overlap with four potent CD4+ T-cell epitopes. Structure-based B-cell epitope prediction has identified ORF9b₆₁₋₆₈ to be immunodominant, which partially overlaps with one of the linear B-cell epitopes (ORF9b₆₅₋₆₉). ORF3c CD4+ T-cell epitopes (ORF3c₂₋₁₆, ORF3c₃₋₁₇, and ORF3c₄₋₁₈) and linear B-cell epitope (ORF3c₁₄₋₂₂) have also been identified as the candidate epitopes. Similarly, M.ext and 7a.iORF1 (overlap with M and ORF7a) proteins have promising immunogenic regions. By considering the level of antigen expression, four ORF9b and five M.ext epitopes are finally shortlisted as potent epitopes. Mutation analysis has further revealed that the shortlisted potent uORF epitopes are resistant to recurrent mutations. Additionally, four N-protein (expressed by canonical ORF) epitopes are found to be potent. Thus, SARS-CoV-2 uORF B-cell and T-cell epitopes identified here along with canonical ORF epitopes may aid in the design of a promising epitope-based polyvalent vaccine (when connected through appropriate linkers) against SARS-CoV-2. Such a vaccine can act as a bulwark against SARS-CoV-2, especially in the scenario of emergence of variants with recurring mutations in the spike protein.

Keywords: uORFs, SARS-CoV-2, ORF9b, T-cell epitopes, B-cell epitopes, epitope-based polyvalent vaccine, Canonical protein epitopes, HLA I supergroups

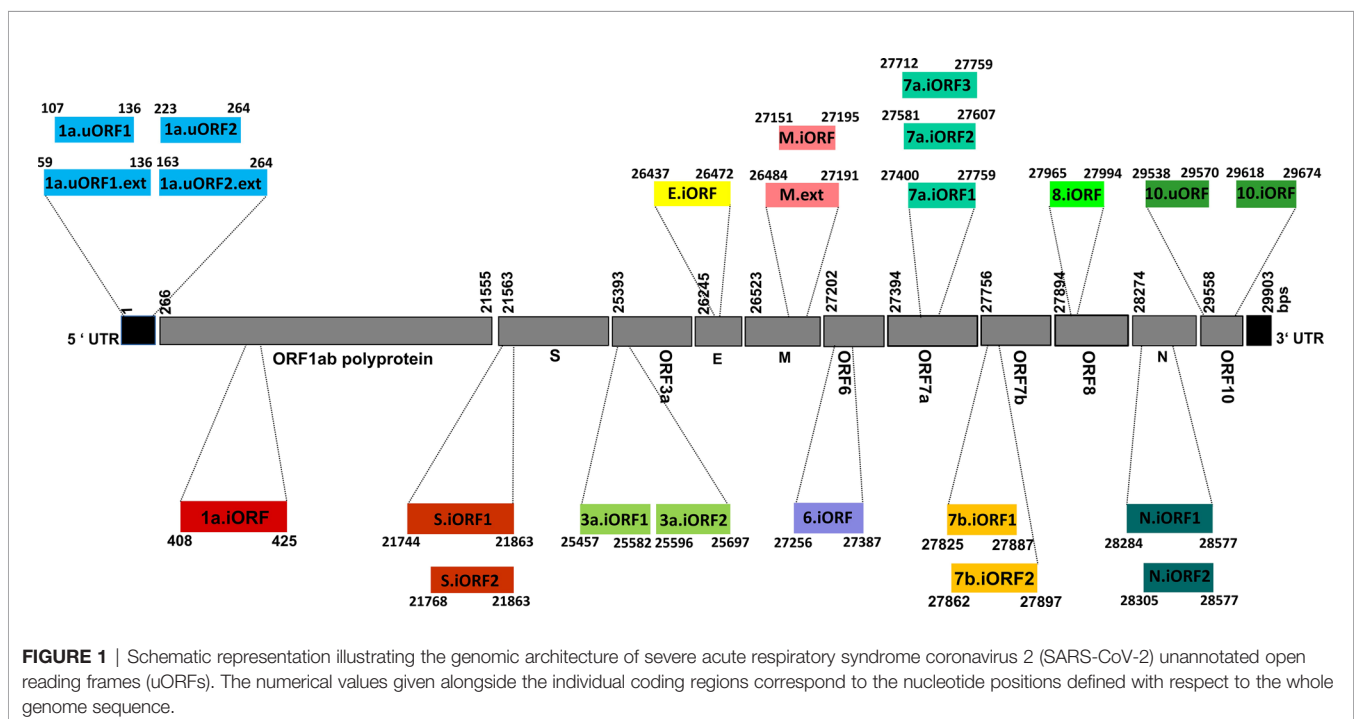
INTRODUCTION

Even 18 months after the official declaration of the SARS-CoV-2 pandemic by the World Health Organization (<https://www.who.int/>), the world is losing thousands of lives, and nearly half a million people around the globe are being infected by the virus every day (<https://www.worldometers.info/coronavirus/>). Although spike glycoprotein-based vaccines have been developed in a fast-track mode to combat SARS-CoV-2 (1–3), the viral evolution with mutations (4, 5) in spike protein and the associated enhanced pathogenicity, transmissibility, and immune escape are of major concerns (6). Indeed, there are reports (7, 8) about the reduced efficacy of the vaccines against the new variants (4). Reports indicate that the number of mutations in the spike protein has increased to 1.4-fold in a time span of 6 months (9, 10). This is indicative of challenges in using the existing spike protein antigen-based vaccines (11) when new variants emerge.

The efficient approach for vaccine development is the multiepitope-based vaccine, which uses short synthetic amino acid stretches that are present in the antigenic protein(s) and are capable of inducing a broad immune response (12). Experimental investigations in recent times have revealed immunodominant epitopes present in the canonical proteins of SARS-CoV-2 (13–25). Several immunoinformatics approaches, which are cost-effective and time-saving compared to the traditional methods, have also been used in this direction to identify potential epitopes in the canonical proteins of SARS-CoV-2 (26–35). Using comparative genomics and ribosome-profiling techniques, a recent experimental study has confirmed the translation of 23 additional unannotated open reading frame (uORF) proteins along with the proteins expressed

by canonical ORFs (36). Despite being expressed in equivalence to the canonical ORF proteins and having functional and regulatory roles, uORFs are being neglected while analyzing the SARS-CoV-2 proteome dynamics. Nonetheless, to the best of our knowledge, there is no systematic investigation carried out to identify immunodominant regions in the uORF proteins of SARS-CoV-2.

Thus, an immunoinformatics approach has been employed here to identify the potential T-cell and B-cell epitopes present in the antigens expressed by SARS-CoV-2 uORFs (**Figure 1** and **Table S1**). Although the SARS-CoV-2 uORFs express 23 proteins (36), nine of them are short polypeptide chains (*viz.*, less than 15 amino acids length) (36). Thus, 14 uORF proteins are considered in the current investigation to identify the potential B-cell linear epitopes and CD4+ T-cell [major histocompatibility complex (MHC) II/human leukocyte antigen (HLA) II] epitopes. Subsequently, 21 uORF proteins (with length of ≥ 9 amino acids) have been considered for CD8+ T-cell (MHC I/HLA I) epitope prediction. Furthermore, variations in these uORFs have also been analyzed by considering the 775,392 SARS-CoV-2 whole genome sequences deposited to GISAID until April 30, 2021. The results reveal several high, moderate, and low recurrent mutations located in the predicted promiscuous epitopes. However, promiscuous ORF9b epitopes are found to be resistant to mutations as well as immunogenic. It is noteworthy that ORF9b plays a role in inhibiting the host innate immune response (37), and it also has a good level of expression. Thus, the potent B-cell and T-cell epitopes identified in ORF9b make it a promising vaccine candidate. Similarly, N and M.ext/M proteins (**Figure 1** and **Table S1**) also possess potent epitopes. Finally, a vaccine construct has been proposed here by considering the potent epitopes of ORF9b, N and M.ext/M proteins.



METHODS

The published SARS-CoV-2 uORF sequences (36) were used as the reference to predict the T-cell and B-cell epitopes present in the uORF proteins. For the mutation analyses, the coding regions corresponding to the uORFs were translated to amino acid sequences using the in-house scripts. **Figure 2** describes the epitope prediction methodology.

T-Cell Epitope Prediction

CD8+ T-cell (MHC I/HLA I) epitope prediction was done against the 78 HLA class I alleles [HLA-A, HLA-B, HLA-C, HLA-E, and HLA-G (given in **Table S2**)] using the TepiTool resource from IEDB web tool (38). The epitopes (9-mer peptides) having the percentile rank ≤ 1 (estimated using the combination

of ANN, SMM, CombLib, and NetMHCpan EL methods) were shortlisted for further analysis.

Parallely, CD4+ T-cell (MHC II/HLA II) epitope prediction was done using the combination of NN-align, SMM-align, CombLib, Sturniolo, and NetMHCII-pan methods against the 27 HLA II alleles [HLA-DR, HLA-DP, and HLA-DQ (**Table S2**)]. The epitopes (15-mer) having the percentile rank ≤ 1 were shortlisted.

From the pool of epitopes, promiscuous epitopes were chosen based on their ability to bind multiple alleles, *viz.*, ≥ 3 and ≥ 2 for CD8+ and CD4+ T-cell epitopes, respectively. Furthermore, immunogenicity of the CD8+ T-cell epitopes was predicted using the MHC I immunogenicity analysis resource of IEDB (39), and the epitopes with an immunogenicity score ≥ 0.25 were shortlisted. Similarly, immunogenicity for CD4+ T-cell epitopes was checked using the CD4+ T-cell immunogenicity prediction tool of

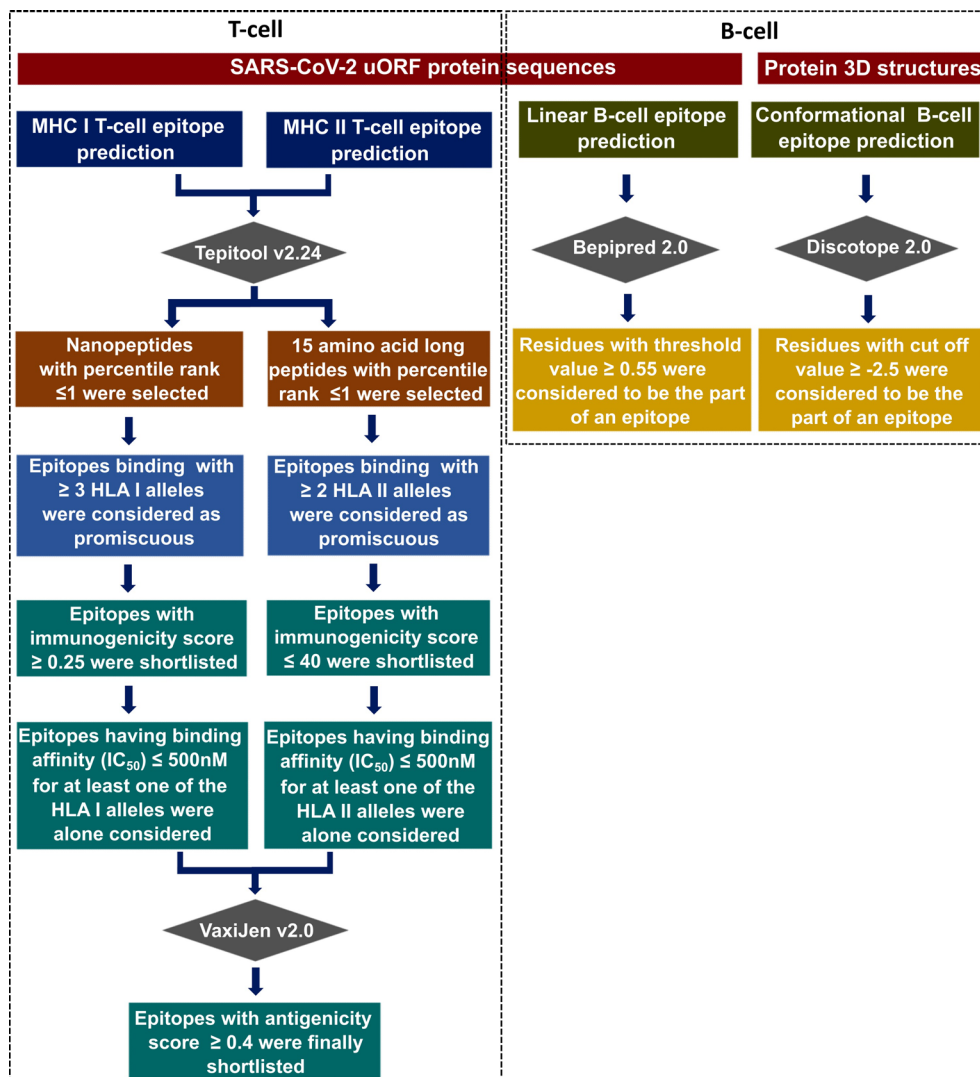


FIGURE 2 | Flowchart illustrating the immunoinformatics protocol used in this study to identify the potential CD8+ T-cell [major histocompatibility complex (MHC) I/ human leukocyte antigen (HLA) I] epitopes, CD4+ T-cell (MHC II/HLA II) epitopes, B-cell linear, and B-cell conformational epitopes corresponding to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) canonical and unannotated open reading frames (uORFs) protein antigens.

IEDB (40, 41), and epitopes having a combined immunogenicity score of ≤ 40 were shortlisted for further analysis.

Characterization and Profiling of Predicted T-Cell Epitopes

Among the shortlisted epitopes, the ones that are having $IC_{50} \leq 500$ nM (42) for at least one of its corresponding HLA alleles were alone considered to have good binding affinity. Antigenicity scores for both HLA I and HLA II epitopes (which fulfill the abovementioned IC_{50} criterion) were predicted using VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) (43). Subsequently, the epitopes with a threshold of 0.4 antigenicity score were alone considered. Furthermore, the worldwide coverage of individual shortlisted epitopes was predicted using the population coverage analysis tool (44). For this, only 78 HLA class I alleles were considered, as they have more than 1% population frequency. The epitopes binding with HLA class I supergroups [seven supergroups (45)] and supertypes [10 supertypes (46)] were also analyzed to confirm the population coverage of the promiscuous epitopes.

Linear and Conformational B-Cell Epitope Prediction

For linear (continuous) B-cell epitope prediction, protein sequences were examined for putative B-cell epitopes using the Bepipred 2.0 server (47) by applying the threshold value of ≥ 0.55 (corresponding to 80% specificity). For the conformational (discontinuous) B-cell epitope prediction, the Discotope 2.0 server (48) (cutoff ≥ -2.5 , corresponding to 80% specificity) was employed. Predicted epitopes were then projected onto the 3D structure(s) of protein antigen(s) using the PyMOL suite (49).

Determination of SARS-CoV-2 uORF Sequence Conservation

To further evaluate the epitopes based on their sequence conservation, 775,392 SARS-CoV-2 whole genome sequences (deposited in the GISAID on or before April 30, 2021) were subjected to nucleotide and amino acid mutation analyses. For this, the gene sequences corresponding to the uORFs were

translated using the reference sequences (36) with the help of in-house scripts. The amino acid mutation analyses were done as discussed elsewhere (9, 10). The mutations were categorized as highly recurring (HR, occurring with $\geq 10\%$ percentage frequency), moderately recurring (MR, occurring with 1%–10% percentage frequency), and low recurring (LR, occurring below 1% percentage frequency but should have occurred at least three times) based on their recurrence in the 775,392 viral proteomes.

RESULTS

IEDB, a widely used resource to identify epitopes (www.iedb.org) (50), is employed in this study to predict the conformational B-cell (≥ 5 -mer peptide), linear B-cell (5–30-mer peptide), CD8+ T-cell (HLA I) (9-mer peptide), and CD4+ T-cell (HLA II) (15-mer peptide) epitopes present in the SARS-CoV-2 uORF proteins. Based on the stringent criteria described in the *Methods* section (**Figure 2**), the potent CD8+ T-cell and CD4+ T-cell epitopes are shortlisted from the pool of predicted epitopes. A CD8+ or CD4+ T-cell epitope is considered a potent epitope only when it fulfills the criteria of promiscuity, immunogenicity, antigenicity, and binding affinity. For instance, although a CD8+ T-cell epitope binds with more than three HLA I alleles and has immunogenicity and antigenicity scores above 0.25 and 0.4, respectively, it is not considered a potent epitope if the IC_{50} value is not less than 500 nM for at least one of its HLA I-binding partners.

By applying the above criteria (**Figure 2**), one conformational B-cell, 10 linear B-cell, 13 CD8+ T-cell (HLA I), and 17 CD4+ T-cell (HLA II) epitopes are shortlisted (**Tables 1–3**). Very interestingly, five of the ORF9b T-cell epitopes have partial overlap with the epitopes found in the SARS-CoV Tor2 strain (Source: IEDB) (www.iedb.org).

Potent B-Cell Epitopes Present in the Proteins Expressed by uORFs

Among the 23 proteins expressed by the SARS-CoV-2 uORFs, only six of them, namely, 1auORF2.ext, M.ext, ORF3c, 7a.iORF1,

TABLE 1 | Promiscuous B-cell epitopes from severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) unannotated open reading frame (uORF) proteins are listed along with the highly recurring (HR) or moderately recurring (MR) mutations, if any.

| uORF proteins | Positions | Epitope | Mutation (HR/MR) | Percentage frequency (%) |
|---------------------------------------|-----------|------------------|------------------|--------------------------|
| Linear B-cell epitopes | | | | |
| 1auORF2.ext | 22–29 | STSRFRPG | R27C | 98.99 |
| M.ext | 196–202 | ASQRVAG | | |
| | 213–230 | RIGNYKLNTHSSSDNI | | |
| ORF3c | 14–22 | YRYKPHLS | L21F; K17E | 1.58; 1.25 |
| 7a.iORF1 | 34–47 | SSGTYEGNSPFHPL | | |
| | 74–83 | QLRARSVSPK | | |
| N.iORF2 | 22–32 | AVGRDQNNVGP | | |
| ORF9b | 5–10 | ISEMHP | | |
| | 29–39 | AVGRDQNNVGP | | |
| | 65–69 | EDKAF | | |
| Conformational B-cell epitopes | | | | |
| ORF9b (PDB ID:6Z4U) | 61–68 | LNSLEDKA | | |

Note that the residue numbers of an epitope (column 2) located in the particular protein (column 1) and its amino acid sequence (column 3) are given along with the (if any) HR/MR (column 4) mutation(s) present in the epitope. Note that column 5 represents the percentage frequency of the HR/MR. For the uORF HR/MR mutations, refer to **Figure 4**.

TABLE 2 | Promiscuous CD8+ T-cell epitopes from severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) unannotated open reading frame (uORF) proteins are listed along with the number of human leukocyte antigen (HLA) I alleles.

| uORF proteins | Positions | Epitopes | No. of alleles | CD8+ immunogenicity score | Vaxijen antigenicity score | Population coverage of the alleles | HR/MR mutation | Percentage frequency (%) |
|---------------|-----------|------------|----------------|---------------------------|----------------------------|------------------------------------|----------------|--------------------------|
| ORF3c | 28–36 | ALHFLFFFR | 5 | 0.27426 | 0.7293 | 29.31% | R36I | 20.18 |
| 6.iORF | 5–13 | KVSIWNLDY | 9 | 0.29343 | 0.8195 | 59.07% | | |
| 7a.iORF1 | 96–104 | SPIFLIVAA | 6 | 0.33318 | 0.6265 | 33.20% | | |
| | 99–107 | FLIVAAIVF | 7 | 0.29611 | 0.6203 | 24.47% | | |
| | 102–110 | VAAIVFITL | 28 | 0.43202 | 0.8679 | 100.00% | | |
| 7b.iORF1 | 3–11 | IIFWFSLLEL | 17 | 0.2683 | 0.8291 | 99.98% | | |
| M.ext | 89–97 | ITGGIAIAM | 11 | 0.34671 | 0.7715 | 54.64% | | |
| | 63–71 | KLIFLWLLW | 5 | 0.34287 | 0.4968 | 18.61% | | |
| | 19–27 | GTTTVEELK | 4 | 0.29473 | 1.0976 | 37.50% | | |
| | 151–159 | LVIGAVILR | 7 | 0.2601 | 1.1027 | 46.67% | | |
| | 149–157 | SELVIGAVI | 11 | 0.25658 | 0.6409 | 43.14% | | |
| ORF9b | 41–48 | KVYPIILRL | 59 | 0.2634 | 0.6489 | 100.00% | | |
| | 87–95 | LPDEFVWT | 6 | 0.31977 | 0.5057 | 24.2% | | |

Note that the residue numbers of an epitope (column 2) located in the particular protein (column 1) and its amino acid sequence (column 3) are given along with its immunogenicity (column 5) and antigenicity (column 6) scores. Note that the number of HLA I alleles to which the epitope binds with (column 4) is also given along with the population coverage of the alleles (column 7). Column 8 represents the highly recurring (HR)/moderately recurring (MR) mutation(s) present in the epitope, if any. The percentage frequencies of the HR/MR mutation(s) are given in column 9. For the uORF HR/MR mutations, refer to **Figure 4**.

TABLE 3 | Promiscuous CD4+ T-cell epitopes from severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) unannotated open reading frame (uORF) proteins are listed along with the number of human leukocyte antigen (HLA) II alleles.

| uORF proteins | Positions | Epitope | No. of alleles | CD4+ immunogenicity score | Vaxijen antigenicity score | Population coverage of the alleles | HR/MR mutation | Percentage frequency (%) |
|---------------|-----------|------------------|----------------|---------------------------|----------------------------|------------------------------------|----------------|--------------------------|
| 3a.iORF1 | 2–16 | LLLQILFALLQRYRY | 2 | 35.9554 | 0.6920 | 14.77% | | |
| (ORF3c) | 26–40 | LLALHFLLFFRALPK | 3 | 38.66696 | 0.4136 | – | R36I | 20.18 |
| | 4–18 | LQILFALLQRYRYKYP | 2 | 39.97024 | 1.1376 | 14.77% | K17E | 1.25 |
| | 3–17 | LLQILFALLQRYRYK | 2 | 40.33808 | 0.7138 | 14.77% | K17E | 1.25 |
| M.ext | 101–115 | VGLMWLSYFIASFRL | 2 | 34.087 | 0.6658 | – | | |
| | 55–69 | RNRFLYIILFLWL | 2 | 39.4372 | 0.5114 | 10.54% | | |
| | 187–201 | RTLKYKLGASQRVA | 2 | 39.12528 | 0.5644 | 17.55% | | |
| 7a.iORF 1 | 96–110 | SPIFLVAAIVFITL | 2 | 40.0232 | 0.7118 | 28.63% | | |
| ORF9b & | 42–56 | YPIILRLGSPSLNMA | 2 | 38.29864 | 0.6786 | 28.79% | | |
| N.iORF2 | 43–57 | PIILRLGSPSLNMA | 2 | 38.73068 | 0.7535 | 28.79% | | |
| | 40–54 | KVYPIILRLGSPSL | 3 | 39.42084 | 0.623 | 37.64% | | |
| | 41–55 | VYPIILRLGSPSLN | 2 | 39.58644 | 0.5837 | 28.79% | | |
| S.iORF 2 | 24–38 | YHLMVMVILLPLRSL | 3 | 40.22584 | 0.8466 | | | |
| S.iORF1 | 4–18 | GSMLYMSLGPMLVRG | 2 | 38.62404 | 1.014 | 17.55% | Y8-, M9- | 42.85, 42.97 |
| | 3–17 | LGSMLYMSLGPMLVR | 2 | 38.86964 | 1.0898 | 17.55% | Y8-, M9- | 42.85, 42.97 |
| | 2–16 | LLGSMLYMSLGPMLV | 2 | 39.86928 | 0.8619 | 17.55% | Y8-, M9- | 42.85, 42.97 |
| | 5–19 | SMLYMSLGPMLVRGL | 2 | 40.08156 | 1.0383 | 17.55% | Y8-, M9- | 42.85, 42.97 |

Note that the residue numbers of an epitope (column 2) located in a particular protein (column 1) and its amino acid sequence (column 3) are given along with its immunogenicity (column 5) and antigenicity (column 6) scores. Note that the number of HLA II alleles to which the epitope bind with (column 4) is also given along with the population coverage of the alleles (column 7). Column 8 represents the HR/MR mutation(s) present in the epitope, if any. The percentage frequencies of the HR/MR mutation(s) are given in column 9. For the uORF HR/MR mutations, refer to **Figure 4**.

N.iORF2, and N.iORF1 (ORF9b), have the B-cell linear epitopes (**Tables 1, SD1**). In total, 10 B-cell linear epitopes of ≥ 5 amino acids length have been identified in the above uORF proteins. Notably, “ASQRVAG” and “RIGNYKLNTHSSSSDNI” epitopes identified from M.ext have overlap with the epitopes identified from the canonical ORF M protein (current study). The predicted M protein epitopes have also been reported in previous immunoinformatics study (34).

Since the structure of ORF9b protein (PDB ID: 6Z4U) alone is known among the uORF proteins, the conformational epitopes

present in the ORF9b protein are alone investigated. A recent SARS-CoV-2 immunoglobulin G (IgG) epitope profiling study has shown that along with the spike protein and N protein, ORF9b protein also elicits IgG-specific SARS-CoV-2 response (17). The prediction reveals the presence of one conformational epitope in the ORF9b protein (for the criteria, refer to **Figure 2**), which is of 8 amino acids length (residue numbers 61–68, LNSLEDKA; **Tables 1, SD2**). The projection of the predicted epitope onto the protein structure indicates that the epitope is surface exposed (**Figure 3**). It is noteworthy that LNSLEDKA

(residue numbers 61–68) conformational epitope has partial overlap with one of the linear B-cell epitopes (EDKAF, residue numbers 65–69; overlapping regions are underlined).

Potent CD8+ T-Cell Epitopes Present in the uORF Proteins

A total of 340 CD8+ T-cell (HLA I) epitopes (which have percentile rank ≤ 1) are predicted from 21 uORF proteins [1a.uORF1.ext, 1a.uORF2.ext, S.iORF1, S.iORF2, 3a.iORF1 (ORF3c), 3a.iORF2, M.ext, 6.iORF, 7a.iORF1, 7b.iORF1, N.iORF1 (ORF9b), N.iORF2, 10.iORF, 1a.uORF1, M.iORF, 7a.iORF3, 1a.uORF2, E.iORF, 7b.iORF2, 8.iORF, and 10.uORF] (Table SD3A). Among these, 248 epitopes show binding with ≥ 3 HLA I alleles. Among the 248 epitopes, only 20 have CD8+ immunogenicity score above 0.25 (Table SD3B). Refer Table SD3 for the complete list of CD8+ T-cell epitopes. Finally, 13 CD8+ T-cell epitopes are shortlisted as potent epitopes (Table 2) based on the antigenicity and IC₅₀ values. Notably, one of the promising ORF9b epitopes, “KVYPILRL,” has an overlap with the SARS-CoV tor2 strain epitopes (Source: IEDB) (www.iedb.org). Most interestingly, another ORF9b promising epitope (ORF9b₈₇₋₉₅; “LPDEFVVVT”) predicted here has also been identified in a recent study through IgG epitope profiling (17).

Potent CD4+ T-Cell Epitopes Present in uORF Proteins

Fourteen uORF proteins (1a.uORF1.ext, 1a.uORF2.ext, S.iORF1, S.iORF2, 3a.iORF1 (ORF3c), 3a.iORF2, M.ext, 6.iORF, 7a.iORF1, 7b.iORF1, N.iORF1 (ORF9b), N.iORF2, 10.iORF, and 7a.iORF3) are predicted to have 140 CD4+ T-cell epitopes with a percentile rank of ≤ 1 (Table SD4A). Among them, 25 epitopes bind with at least two of the HLA II alleles and have a CD4+ immunogenicity score of ≤ 40 (Table SD4B). By applying the criteria as described in Figure 2, 17 epitopes are finally shortlisted as the potent CD4+

T-cell epitopes (Table 3). Refer Table SD4C for the complete list of CD4+ T-cell epitopes

Interestingly, one of the potent ORF9b CD8+ T-cell epitopes, “KVYPILRL,” has complete overlap with a potent CD4+ T-cell epitope (“KVYPILRLGSPSLSL”; overlapping region is underlined). It also has an overlap with three other ORF9b potent CD4+ T-cell epitopes, namely, “PIILRLGSPSLNMA,” “YPIILRLGSPSLNM,” and “VYPILRLGSPSLSN” (Figure 3).

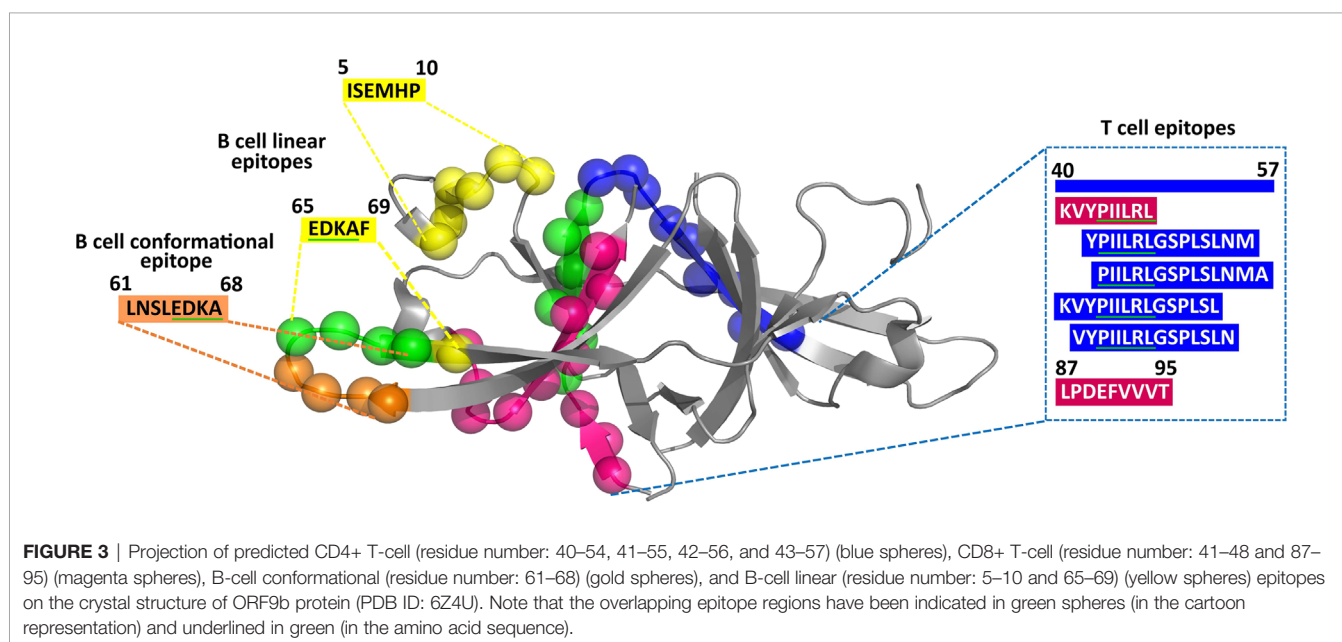
Population Coverage Exhibited by the Potent CD8+ T-Cell and CD4+ T-Cell Epitopes

The population coverage of the potent CD8+ T-cell and CD4+ T-cell epitopes are subsequently investigated. For this, the 13 CD8+ T-cell and 17 CD4+ T-cell epitopes respectively are tested against the IEDB HLA I and HLA II allele repository. In the case of potent CD8+ T-cell epitopes, the population coverage of the individual epitopes ranges between 18% and 100% (Tables 2, 3). Three of the CD8+ T-cell epitopes have population coverage of about 100%, and one of them is ORF9b epitope “KVYPILRL.” It further has the highest number of (59 out of 78) HLA I allele-binding partners. Note that population coverage is not simply depicted by the number of alleles that an epitope binds with. Rather, it represents the genotypic frequency of the allele it binds with.

Furthermore, 17 potent CD4+ T-cell epitopes are predicted to have individual population coverage between 10.54% and 37.64%. The ORF9b epitope “KVYPILRLGSPSLSL” has the highest population coverage among the CD4+ T-cell epitopes (*viz.*, 37.64%).

HLA Class I Supergroup Coverage of Potent CD8+ T-Cell Epitopes

To further investigate the binding specificity or flexibility of the CD8+ T-cell epitopes to HLA class I alleles, the allele-binding



coverage of the predicted promiscuous epitopes is analyzed by considering seven HLA class I supergroups (45) and 10 HLA class I supertypes (46). Analysis indicates that only 30% of the promiscuous uORF CD8+ T-cell epitopes fall into the same HLA class I supergroup. For instance, four out of 13 uORF epitopes are specific only to a particular HLA supergroup (Table SD5). The rest of the promiscuous epitopes bind to the HLA class I alleles that belong to at least two supergroups.

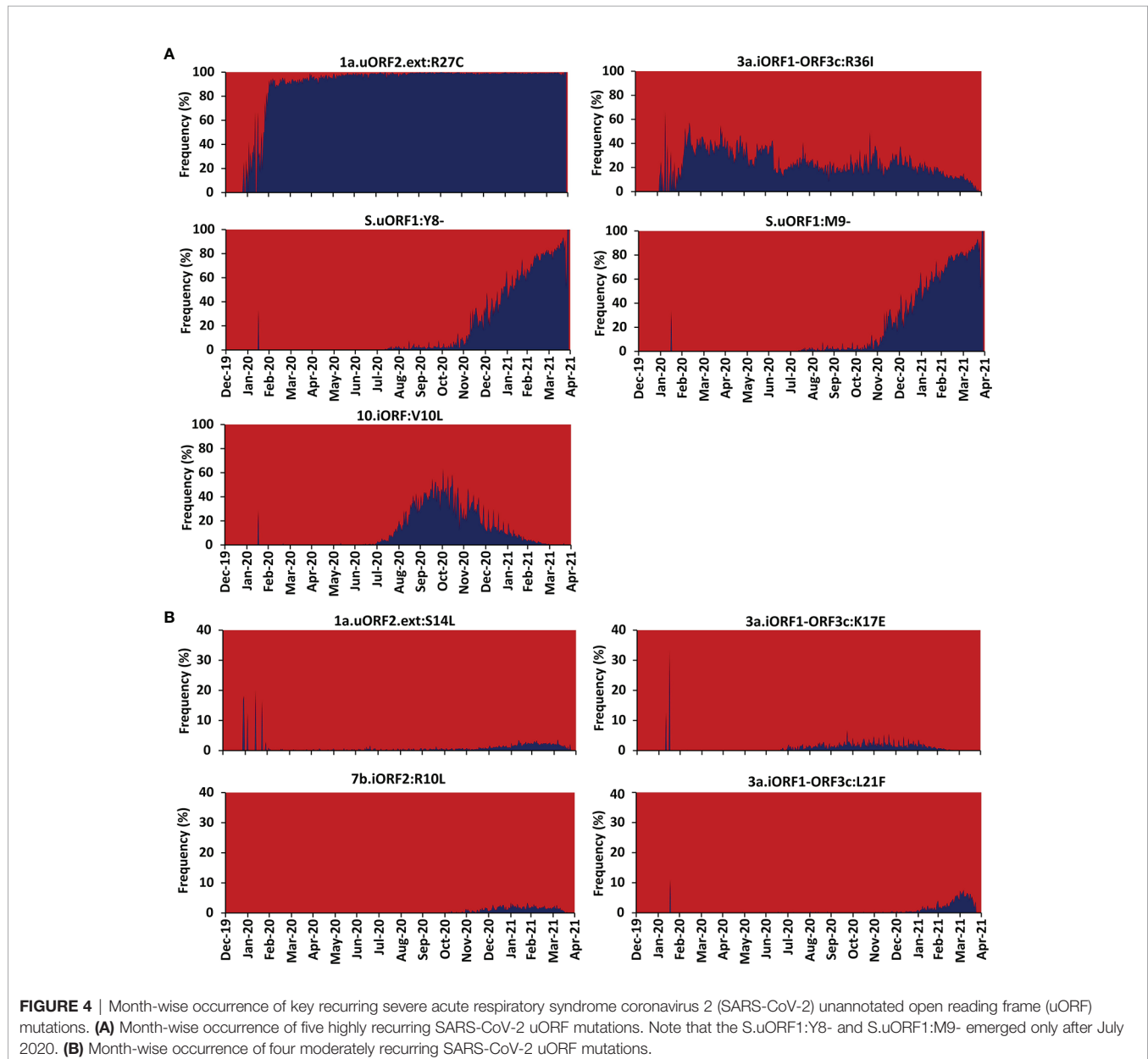
Conservation of Epitope Regions

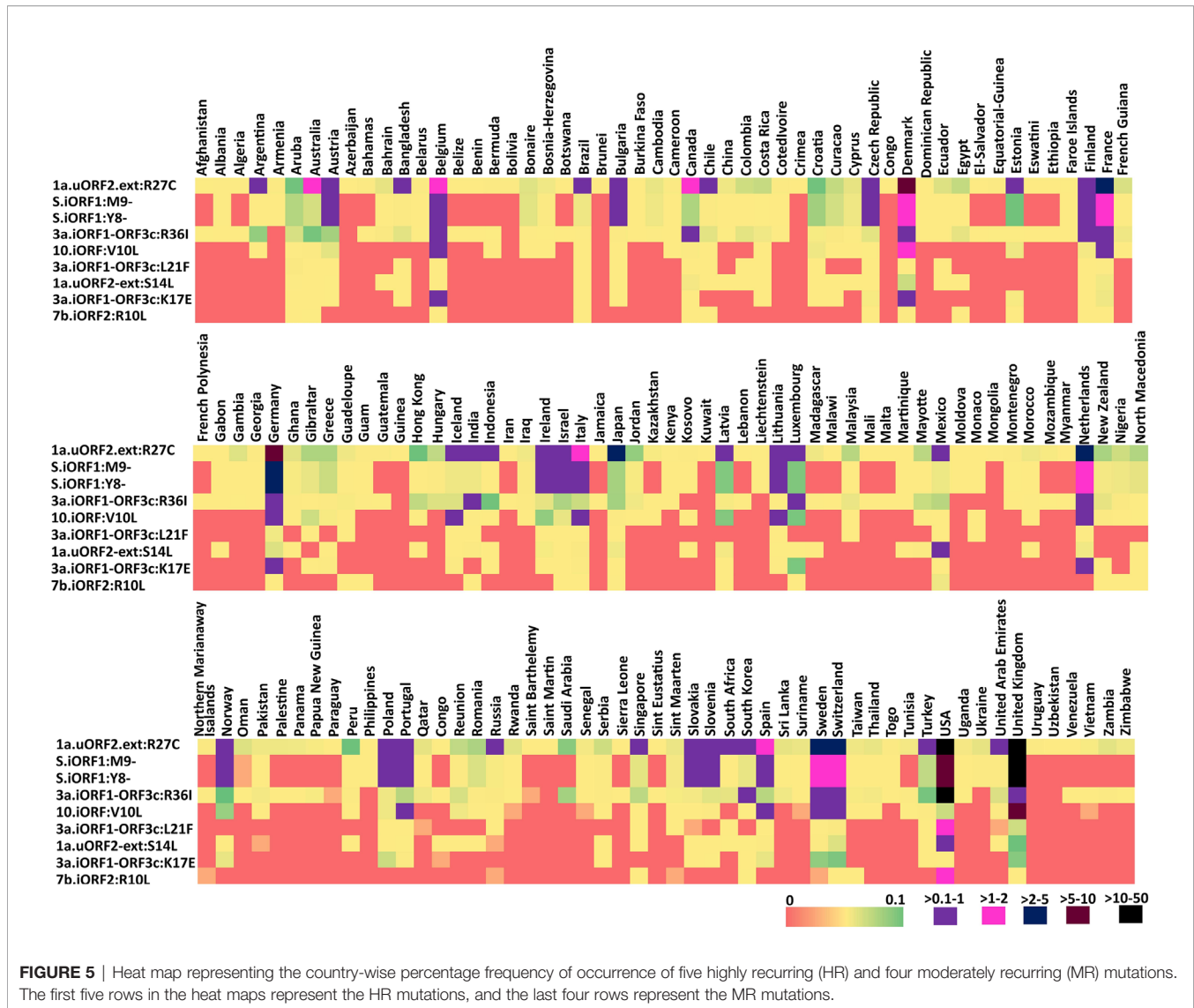
To investigate the conservation of the predicted SARS-CoV-2 uORF conformational B-cell, linear B-cell, CD8+ T-cell (HLA I), and CD4+ T-cell (HLA II) epitopes, mutational analysis is carried out for the proteins expressed by uORFs. The results reveal that uORF proteins are having 5, 4, and 2,642 high, moderate, and low

recurrent mutations, respectively (Figures 4, 5; Table SD6). While some of the low-recurring mutations are found in the predicted potential epitope regions, four highly recurring (1a-uORF2-ext: R27C, S.iORF1:Y8-, S.iORF1:M9-, and ORF3c:R36I) and two moderately recurring mutations (ORF3c:L21F and ORF3c:K17E) are found to occur only in 10 out of 41 shortlisted potent epitopes (Tables 1–3). Thus, these epitopes are not considered as potential epitopes.

DISCUSSION

Even 18 months after the coronavirus disease 2019 (COVID-19) outbreak, treating and preventing SARS-CoV-2 infection are still a big challenge. Especially, the emergence of variants with





recurring mutations in the spike protein poses major challenges in treating the SARS-CoV-2 infection. An immunoinformatics approach is employed here to facilitate the multi-epitope vaccine design, which may aid in overcoming the challenges in the traditional vaccine design. Although several studies have been carried out in this regard (26–35), they have mainly focused on predicting the epitopes from the proteins expressed by the canonical ORFs. In addition to the canonical ORF proteins, the uORF proteins of SARS-CoV-2 also exhibit antigenicity (Table S1) (17, 51). Nonetheless, there is no systematic investigation carried out to identify the epitopes present in uORF proteins. To this end, the current investigation aims to predict the linear B-cell, conformational B-cell, CD4+ T-cell, and CD8+ T-cell epitopes present in SARS-CoV-2 uORF proteins.

Using the recently published SARS-CoV-2 uORF sequences as the reference sequences (36), the epitopes present in the uORF proteins have been scanned in the IEDB database (www.iedb.org) (50). Based on the cutoff criteria described in Figure 2, 10

linear and one conformational B-cell epitopes have been shortlisted as potent epitopes (Table 1). Similarly, 17 CD4+ T-cell and 13 CD8+ T-cell epitopes have been shortlisted by considering their antigenicity, immunogenicity, and IC₅₀ value (Tables 2, 3). Additionally, the selected CD4+ T-cell and CD8+ T-cell epitopes exhibit an allele coverage of least two (out of 27 HLA II alleles) and three (out of 78 HLA I alleles), respectively. Interestingly, three of the CD8+ T-cell epitopes [“KVYPILRL” (ORF9b), “IIFWFSLEL” (ORF7b.iORF1), “VAAIVFITL” (ORF7a.iORF1)] exhibit 100% world population coverage, indicating that they are more promising epitopes. Furthermore, CD8+ T-cell epitope “KVYPILRL” (Figure 3) from ORF9b protein exhibits binding with the highest number of HLA I alleles (59 out of 78). This epitope also has an overlapping region with four of the potent CD4+ T-cell epitopes. Thus, these epitopes are the promising T-cell epitopes. Further mutational analyses have confirmed that HR and MR mutations are found only in 10 out of 41 shortlisted potent epitopes (Tables S1–S3).

Thus, a total of 31 epitopes, *viz.*, nine B-cell and 22 T-cell epitopes are finally shortlisted as potent epitopes from the uORF proteins.

Potent SARS-CoV-2 Unannotated ORF Epitopes for Multi-Epitope Vaccine Design

From the pool of 31 potent epitopes, 13 epitopes are proposed to be highly suitable for the multi-epitope SARS-CoV-2 vaccine design as discussed below. Since ORF9b followed by M.ext have better expression level among the uORF proteins, epitopes from these proteins can be considered for the vaccine design. The only conformational uORF B-cell epitope [“LNSLEDKA” (ORF9b₆₁₋₆₈)] and four linear B-cell epitopes [“AVGRDQNNVGP” (ORF9b₂₉₋₃₉, N_iORF2₂₂₋₃₂), “EDKAF” (ORF9b₆₅₋₆₉), “ASQRVAG” (M.ext₁₉₆₋₂₀₂), and “RIGNYKLNTDHSSSSDNI” (M.ext₂₁₃₋₂₃₀)] can be considered for the vaccine design. However, “EDKAF” (ORF9b₆₅₋₆₉) is excluded for the vaccine design as it overlaps (underlined) with the conformational B-cell epitope “LNSLEDKA” (ORF9b₆₁₋₆₈). Thus, a total of four B-cell epitopes are considered for the vaccine design.

Among the shortlisted potent CD8+ T-cell epitopes, “KVYPIILRL” (ORF9b₄₀₋₄₈) is predicted to bind with six (S1–S4, S6, and S7) out of seven HLA class I supergroups (**Figure 6A**, **Table SD5**). Interestingly, “SELVIGAVI” (M.ext₁₄₉₋₁₅₇) is the only shortlisted epitope that covers HLA class I supergroup S5 (in addition to S3) that is not covered by “KVYPIILRL” (ORF9b₄₀₋₄₈); thus, it becomes a valuable candidate for the vaccine design. Although “LPDEFVVVT” (ORF9b₈₇₋₉₅) (HLA class I supergroup S3) and GTITVEELK (M.ext₁₉₋₂₇) (HLA class I supergroup S1) cover only one of the HLA class I supergroups, they are also considered for the vaccine design, as they have been reported in earlier investigations (**Figures 6A, B**) (17, 32). In fact, “LPDEFVVVT” (ORF9b₈₇₋₉₅) is identified to be an epitope of high confidence in a recent IgG profiling experiment (17).

Similarly, the following CD4+ T-cell epitopes can be promising vaccine candidates: “KVYPIILRLGSPLSL” (ORF9b₄₀₋₅₄), VGLMWLSYFIASFRL” (M.ext₁₀₁₋₁₁₅), and “RTLSYYKLGASQRVA” (M.ext₁₈₇₋₂₀₁ and it overlaps with M.ext₁₈₈₋₂₀₂ CD4+ T-cell epitope). Notably, “KVYPIILRLGSPLSL” (ORF9b₄₀₋₅₄) overlaps with three other ORF9b CD4+ T-cell epitopes [“YPIILRLGSPLSLNM” (ORF9b₄₂₋₅₆), “PIILRLGSPLSLNMA” (ORF9b₄₃₋₅₇), and “VYPIILRLGSPLSLN” (ORF9b₄₁₋₅₅)]. Furthermore, “KVYPIILRLGSPLSL” (ORF9b₄₀₋₅₄) overlaps (underlined) with the ORF9b CD8+ T-cell epitope “KVYPIILRL” (ORF9b₄₀₋₄₈) that covers six out of seven HLA class I supergroups (**Figure 6A**; **Tables SD5A, B**). Thus, “KVYPIILRLGSPLSL” (ORF9b₄₀₋₅₄) is the utmost promising CD4+ T-cell epitope and is considered for the vaccine design. Similarly, “RTLSYYKLGASQRVA” (M.ext₁₈₇₋₂₀₁) is considered to be a potential epitope. Although some of the uORF proteins CD8+ T-cell epitopes have good world population coverage, they are not considered for the vaccine design due to the low expression level of the corresponding protein (36). Such examples include “VAAIVFITL” (7a_iORF1₁₀₂₋₁₁₀, population coverage = 100%) and “IIFWFSLEL” (7b_iORF1₃₋₁₁, population coverage ~100%) epitopes.

Thus, nine epitopes from ORF9b and M.ext proteins are proposed for the multi-epitope SARS-CoV-2 vaccine design (**Figure 6C**). Among the nine uORF epitopes considered for the vaccine design, “KVYPIILRLGSPLSL” (ORF9b₄₀₋₅₄) is reported in SARS-CoV Tor 2 strain [Source: IEDB] (www.iedb.org). The epitopes “SELVIGAVI” (M.ext₁₄₉₋₁₅₇) (19), GTITVEELK (M.ext₁₉₋₂₈) (32), “RIGNYKLNTDHSSSSDNI” (M.ext₂₁₃₋₂₃₀) (34), “RTLSYYKLGASQRVA” (M.ext₁₈₇₋₂₀₁) (19), and “LPDEFVVVT” (ORF9b₈₇₋₉₅) (17) are reported in previous experimental investigations. Since the M.ext protein expressed by the uORF has overlap with the M protein expressed by the corresponding canonical ORF, the M.ext protein epitopes proposed here are also found in M protein.

Due to the high immunodominance and high expression level of canonical ORF proteins, the present study aims to propose a vaccine construct that has the epitopes from both the canonical ORF and uORF proteins. Thus, the potent linear B-cell, conformational B-cell, CD4+ T-cell, and CD8+ T-cell epitopes from canonical ORF proteins have also been investigated independently in this study to propose an efficient multi-epitope-based vaccine construct that encompasses the epitopes from both the canonical ORF and uORF proteins. **Tables SD7–SD10** have the information about the epitopes predicted in the 26 SARS-CoV-2 canonical proteins. By following the same criteria used in the screening and shortlisting of uORF protein epitopes, 41 linear B-cell, five conformational B-cell, 115 CD8+ T-cell, and 71 CD4+ T-cell epitopes are shortlisted as potent epitopes from the canonical ORF proteins (**Tables S3–S6**). Since several epitope prediction studies have been carried out for canonical ORF proteins, the results are not discussed here in detail. The diversity in epitope-HLA class I allele binding is further confirmed by analyzing the epitope binding diversity with respect to different HLA class I supergroups (see *Methods*). In the case of promiscuous epitopes shortlisted from canonical proteins, only 17% of them bind to HLA class I alleles that fall into the same HLA supergroup (**Table SD11**). For instance, 10 out of 71 Nsp1–Nsp16 epitopes and nine out of 44 ORF2 (Spike)–ORF10 epitopes fall into the same HLA I supergroup. Among the shortlisted promising epitopes from the canonical ORF proteins, 36 and 91 have complete and partial overlap (>60%), respectively, with the earlier reported/predicted SARS-CoV-2 epitopes (13–20, 26, 31–35, 52) (**Tables S7A, B**; **Tables SD12, 13**). Thus, the linear B-cell, conformational B-cell, CD8+ T-cell, and CD4+ T-cell epitopes predicted here from the canonical ORF proteins act as a benchmark to validate the prediction of uORF epitopes. Indeed, there is a possibility of excluding the epitopes with good immunogenicity while applying additional criteria like antigenicity and/or binding affinity (IC₅₀) (**Figure 2**). However, a detailed comparison between the previously predicted/reported canonical ORF epitopes and the epitopes that are excluded (in the current study) despite having a good immunogenicity score (above 0.25 for CD8+ T-cell and below 40 for CD4+ T-cell epitopes) indicates that only a fraction (<15%) of such epitopes have been excluded (**Tables SD14, 15**).

For the multi-epitope vaccine design, the epitopes predicted from the N protein (encoded by one of the canonical ORFs) are considered, as it tops among the canonical and uORF proteins in

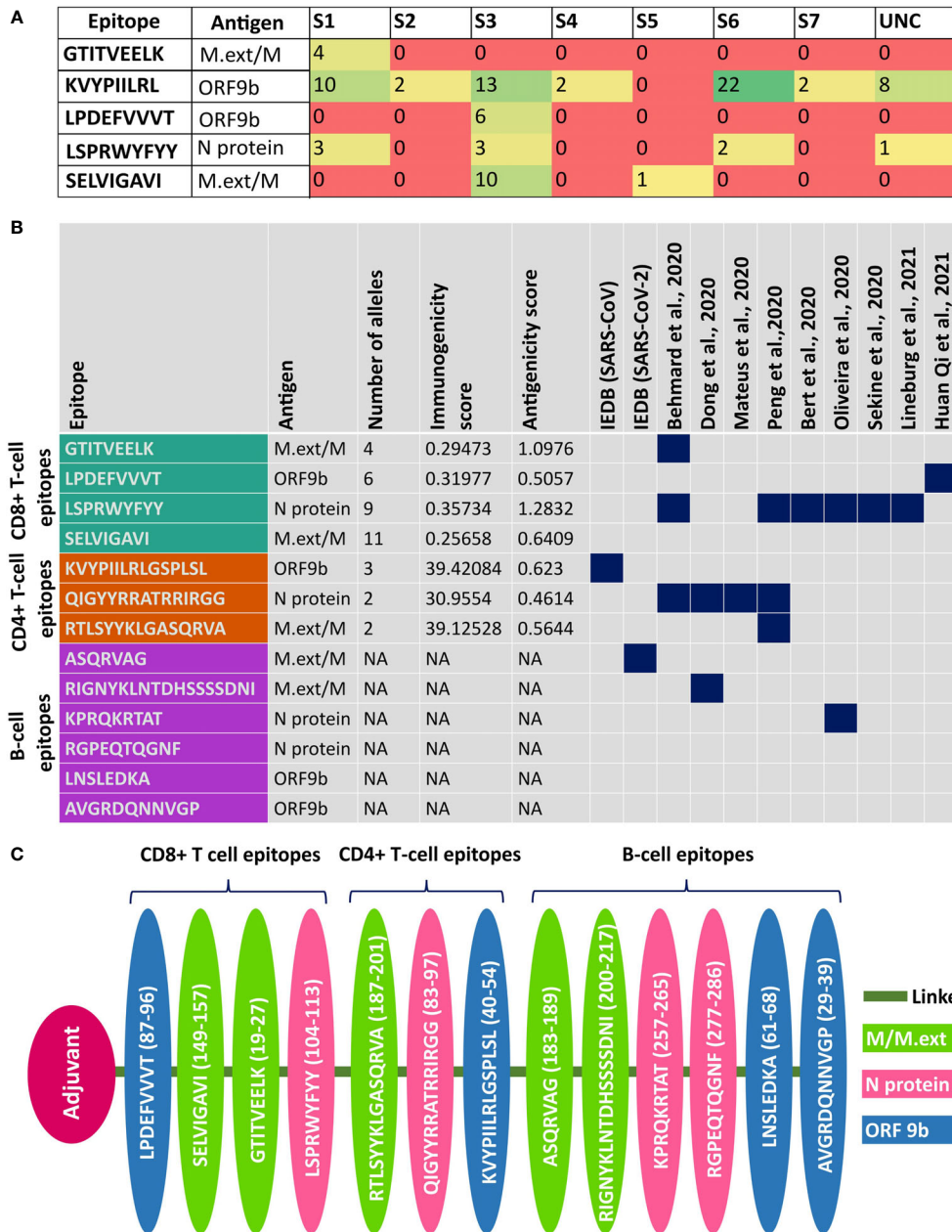


FIGURE 6 | Details about the potent epitopes and the design of multi-epitope vaccine construct. **(A)** Table showing the human leukocyte antigen (HLA) class I supergroup coverage of potent CD8+ T-cell epitopes. **(B)** Summary of the potent CD8+ T-cell, CD4+ T-cell, B-cell linear, and B-cell conformational epitopes used for the vaccine design. Note that if a potent epitope has already been reported elsewhere, it is highlighted in dark blue. **(C)** Linear multi-epitope vaccine construct designed using potent CD8+ T-cell, CD4+ T-cell, and B-cell epitopes predicted from ORF9b, N, and M/M.ext proteins. Note that ORF9b, N, and M/M.ext epitopes are depicted in pink, green, and blue, respectively.

terms of the relative translational level (36). “LSPRWYFYF” (N protein₁₀₄₋₁₁₂) and “QIGYYRRATRRIRGG” (N protein₈₃₋₉₇) are predicted to be potent CD8+ and CD4+ T-cell epitopes, respectively. Notably, the predicted “LSPRWYFYF” (N protein₁₀₄₋₁₁₂) epitope [covers three HLA class I supergroups (S1, S3, and S6)] (**Figure 6A**) overlaps (underlined region) with a previously reported “SPRWYFYFL” immunodominant epitope

(**Figure 6B**) (13, 16, 19, 20, 32, 33). Similarly, “QIGYYRRATRRIRGG” (N protein₈₃₋₉₇) has partial overlap with the previously reported epitope (14, 19, 32, 34). The linear B-cell epitopes “KPRQKRTAT” (N protein₂₅₇₋₂₆₅) and “RGPEQTQGNF” (N protein₂₇₇₋₂₈₆) are considered to be potent epitopes for the vaccine design. Among them, “KPRQKRTAT” (N protein₂₅₇₋₂₆₅) has partial overlap with the conformational B-

cell epitope of the N protein and has been reported in previous immunoinformatics studies (33, 34).

Aforementioned potent epitopes identified from N protein (four epitopes), M.ext/M protein (five epitopes), and ORF9b protein (four epitopes) may be more appropriate for the vaccine design, since these proteins occupy the top 3 positions among the SARS-CoV-2 proteins in terms of the relative translational level (as revealed from the ribosome profiling) (36). Considering this point, “SELVIGAVI” (M₁₃₈₋₁₄₆, M.ext₈₉₋₉₇), “GTITVEELK” (M₆₋₁₄, M.ext₁₉₋₂₇), “LSPRWYFY” (N₁₀₄₋₁₁₂), “LPDEFVVVT” (ORF9b₈₇₋₉₅), “RTLSTYKLGASQRVA” (M.ext₁₈₇₋₂₀₁, M₁₇₄₋₁₈₈), “KVYPIILRLGSPSL” (ORF9b₄₀₋₅₄), “QIGYYRRATRRIRGG” (N₈₃₋₉₇), “LNSLEDKA” (ORF9b₆₁₋₆₈), “AVGRDQNNVGP” (ORF9b₂₉₋₃₉), “ASQRVAG” (M.ext₁₉₆₋₂₀₂, M₁₈₃₋₁₈₉), “RIGNYKLNTHSSSSDNI” (M.ext₂₁₃₋₂₃₀, M₂₀₀₋₂₁₇), “KPRQKRTAT” (N₂₅₇₋₂₆₅), and “RGPEQTQGNF” (N₂₇₇₋₂₈₆) epitopes predicted in this study can be used for multi-epitope-based SARS-CoV-2 vaccine design, wherein each epitope is connected through a suitable linker region (Figure 6C). The multi-epitope vaccine is designed in such a way to cover all the HLA class I supergroups (Figure 6A).

Thus, the potent epitopes from the proteins expressed by the canonical ORFs and uORFs of SARS-CoV-2 can be used in the design of multi-epitope vaccine against SARS-CoV-2.

CONCLUSIONS

To facilitate the design of a multi-epitope vaccine against SARS-CoV-2, an immunoinformatics analysis has been carried out here to identify the potential linear B-cell, conformational B-cell, CD4+ T-cell, and CD8+ T-cell epitopes present in the 23 uORF proteins. Using stringent criteria, nine linear B-cell, one conformational B-cell, 17 CD4+ T-cell, and 13 CD8+ T-cell uORF epitopes are shortlisted. Notably, the current study has identified ORF9b epitopes as promising candidates for the multi-epitope vaccine design. “KVYPIILRL” [ORF9b₄₀₋₄₈] CD8+ T-cell (MHC I/HLA class I) epitope is the most promising epitope not only based on the antigenicity, immunogenicity, and IC50 but also based on its highest HLA class I allele coverage, *viz.*, it covers six out of seven HLA class I supergroups. Furthermore, this region has an overlap with the four potent CD4+ T-cell (MHC II/HLA class II) epitopes. Among the shortlisted uORF epitopes, eight linear B-cell, one conformational B-cell, 10 CD4+ T-cell, and 12 CD8+ T-cell epitopes are finally suggested as potent epitopes based on the mutational analysis. Similar immunoinformatics analysis is also extended for 26 canonical

REFERENCES

- Salvatori G, Luberto L, Maffei M, Aurisicchio L, Roscilli G, Palombo F, et al. SARS-CoV-2 SPIKE PROTEIN: An Optimal Immunological Target for Vaccines. *J Transl Med* (2020) 18:222. doi: 10.1186/s12967-020-02392-y
- Dong Y, Dai T, Wei Y, Zhang L, Zheng M, Zhou F. A Systematic Review of SARS-CoV-2 Vaccine Candidates. *Signal Transduct Target Ther* (2020) 5:237. doi: 10.1038/s41392-020-00352-y

ORF proteins. Considering the high expression level of N protein (encoded by canonical ORF), M/M.ext protein (encoded by canonical/uORF), and ORF9b protein (encoded by uORF), 13 potent epitopes from these proteins are finally considered for the proposed multivalent vaccine design.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

PU carried out the immunoinformatics analysis. LP wrote codes for mutation data analysis and plotting. CS wrote scripts to generate the plots and did plotting. LP, CS, and PU analyzed the data. PU and TR wrote the manuscript. PU independently devised the immunoinformatics analysis protocol. TR designed and supervised the project. All authors contributed to the article and approved the submitted version.

FUNDING

LP and CS thank MHRD for fellowship. PU thanks CSIR for fellowship.

ACKNOWLEDGMENTS

The authors wish to acknowledge all the researchers who have deposited the SARS-CoV-2 genome sequences (used in this study) to GISAID and GISAID for providing the sequences. The authors also thank Ms. Sruthi Sundaresan for proofreading the manuscript and IIT Hyderabad for computational resources.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2021.692937/full#supplementary-material>

- Amanat F, Krammer F. SARS-CoV-2 Vaccines: Status Report. *Immunity* (2020) 52:583–9. doi: 10.1016/j.immuni.2020.03.007
- Abdool Karim SS, de Oliveira T. New SARS-CoV-2 Variants - Clinical, Public Health, and Vaccine Implications. *N Engl J Med* (2021) 384:1866–8. doi: 10.1056/NEJMc2100362
- McCarthy KR, Rennick LJ, Nambulli S, Robinson-McCarthy LR, Bain WG, Haidar G, et al. Recurrent Deletions in the SARS-CoV-2 Spike Glycoprotein Drive Antibody Escape. *Science* (2021) 371:1139–42. doi: 10.1126/science.abc6950

6. Altmann DM, Boyton RJ, Beale R. Immunity to SARS-Cov-2 Variants of Concern. *Science* (2021) 371:1103–4. doi: 10.1126/science.abg7404
7. Garcia-Beltran WF, Lam EC, St Denis K, Nitido AD, Garcia ZH, Hauser BM, et al. Multiple SARS-CoV-2 Variants Escape Neutralization by Vaccine-Induced Humoral Immunity. *Cell* (2021) 184:2372–83.e9. doi: 10.1016/j.cell.2021.03.013
8. Williams TC, Burgers WA. SARS-CoV-2 Evolution and Vaccines: Cause for Concern? *Lancet Respir Med* (2021) 9:333–5. doi: 10.1016/S2213-2600(21)00075-8
9. Patro LPP, Sathyaseelan C, Uttamrao PP, Rathinavelan T. Global Variation in SARS-Cov-2 Proteome and its Implication in Pre-Lockdown Emergence and Dissemination of 5 Dominant SARS-CoV-2 Clades. *Infection Genet Evol* (2021) 93:104973. doi: 10.1016/j.meegid.2021.104973
10. Patro LPP, Sathyaseelan C, Uttamrao PP, Rathinavelan T. The Evolving Proteome of SARS-CoV-2 Predominantly Uses Mutation Combination Strategy for Survival. *Comput Struct Biotechnol J* (2021) 19:3864–75. doi: 10.1016/j.csbj.2021.05.054
11. Arashkia A, Jalilvand S, Mohajel N, Afchangi A, Azadmanesh K, Salehi-Vaziri M, et al. Severe Acute Respiratory Syndrome-Coronavirus-2 Spike (s) Protein Based Vaccine Candidates: State of the Art and Future Prospects. *Rev Med Virol* (2020) 31:Se2183. doi: 10.1002/rmv.2183
12. Kao DJ, Hodges RS. Advantages of a Synthetic Peptide Immunogen Over a Protein Immunogen in the Development of an Anti-Pilus Vaccine for *Pseudomonas Aeruginosa*. *Chem Biol Drug Des* (2009) 74:33–42. doi: 10.1111/j.1747-0285.2009.00825.x
13. Le Bert N, Tan AT, Kunasegaran K, Tham CYL, Hafezi M, Chia A, et al. SARS-CoV-2-Specific T Cell Immunity in Cases of COVID-19 and SARS, and Uninfected Controls. *Nature* (2020) 584:457–62. doi: 10.1038/s41586-020-2550-z
14. Mateus J, Grifoni A, Tarke A, Sidney J, Ramirez SI, Dan JM, et al. Selective and Cross-Reactive SARS-CoV-2 T Cell Epitopes in Unexposed Humans. *Science* (2020) 370:89–94. doi: 10.1126/science.abd3871
15. Amrun SN, Lee CY, Lee B, Fong SW, Young BE, Chee RS, et al. Linear B-Cell Epitopes in the Spike and Nucleocapsid Proteins as Markers of SARS-CoV-2 Exposure and Disease Severity. *EBioMedicine* (2020) 58:102911. doi: 10.1016/j.ebiom.2020.102911
16. Sekine T, Perez-Potti A, Rivera-Ballesteros O, Stralin K, Gorin JB, Olsson A, et al. Robust T Cell Immunity in Convalescent Individuals With Asymptomatic or Mild COVID-19. *Cell* (2020) 183:158–68.e114. doi: 10.1016/j.cell.2020.08.017
17. Qi H, Ma ML, Jiang HW, Ling JY, Chen LY, Zhang HN, et al. Systematic Profiling of SARS-CoV-2-Specific Igg Epitopes at Amino Acid Resolution. *Cell Mol Immunol* (2021) 18:1067–9. doi: 10.1038/s41423-021-00654-3
18. Shomuradova AS, Vagida MS, Sheetikov SA, Zornikova KV, Kiryukhin D, Titov A, et al. SARS-CoV-2 Epitopes are Recognized by a Public and Diverse Repertoire of Human T Cell Receptors. *Immunity* (2020) 53:1245–57.e1245. doi: 10.1016/j.immuni.2020.11.004
19. Peng Y, Mentzer AJ, Liu G, Yao X, Yin Z, Dong D, et al. Broad and Strong Memory CD4(+) and CD8(+) T Cells Induced by SARS-CoV-2 in UK Convalescent Individuals Following COVID-19. *Nat Immunol* (2020) 21:1336–45. doi: 10.1038/s41590-020-0782-6
20. Lineburg KE, Grant EJ, Swaminathan S, Chatzileontiadou DSM, Szeto C, Sloane H, et al. CD8(+) T Cells Specific for an Immunodominant SARS-CoV-2 Nucleocapsid Epitope Cross-React With Selective Seasonal Coronaviruses. *Immunity* (2021) 54:1055–65.e1055. doi: 10.1016/j.immuni.2021.04.006
21. Shrock E, Fujimura E, Kula T, Timms RT, Lee IH, Leng Y, et al. Viral Epitope Profiling of COVID-19 Patients Reveals Cross-Reactivity and Correlates of Severity. *Science* (2020) 370:eabd4250. doi: 10.1126/science.abd4250
22. Li Y, Lai DY, Zhang HN, Jiang HW, Tian X, Ma ML, et al. Linear Epitopes of SARS-CoV-2 Spike Protein Elicit Neutralizing Antibodies in COVID-19 Patients. *Cell Mol Immunol* (2020) 17:1095–7. doi: 10.1038/s41423-020-00523-5
23. Musico A, Frigerio R, Mussida A, Barzon L, Sinigaglia A, Riccetti S, et al. SARS-CoV-2 Epitope Mapping on Microarrays Highlights Strong Immune-Response to N Protein Region. *Vaccines (Basel)* (2021) 9:35. doi: 10.3390/vaccines9010035
24. Ng KT, Mohd-Ismail NK, Tan YJ. Spike S2 Subunit: The Dark Horse in the Race for Prophylactic and Therapeutic Interventions Against SARS-CoV-2. *Vaccines (Basel)* (2021) 9:178. doi: 10.3390/vaccines9020178
25. Li Y, Lai DY, Lei Q, Xu ZW, Wang F, Hou H, et al. Systematic Evaluation of Igg Responses to SARS-CoV-2 Spike Protein-Derived Peptides for Monitoring COVID-19 Patients. *Cell Mol Immunol* (2021) 18:621–31. doi: 10.1038/s41423-020-00612-5
26. Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A, et al. And Bioinformatic Approach can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host Microbe* (2020) 27:671–80.e672. doi: 10.1016/j.chom.2020.03.002
27. Chen HZ, Tang LL, Yu XL, Zhou J, Chang YF, Wu X. Bioinformatics Analysis of Epitope-Based Vaccine Design Against the Novel SARS-CoV-2. *Infect Dis Poverty* (2020) 9:88. doi: 10.1186/s40249-020-00713-3
28. Kar T, Narsaria U, Basak S, Deb D, Castiglione F, Mueller DM, et al. A Candidate Multi-Epitope Vaccine Against SARS-CoV-2. *Sci Rep* (2020) 10:10895. doi: 10.1038/s41598-020-67749-1
29. Li W, Li L, Sun T, He Y, Liu G, Xiao Z, et al. Spike Protein-Based Epitopes Predicted Against SARS-CoV-2 Through Literature Mining. *Med Nov Technol Devices* (2020) 8:100048. doi: 10.1016/j.medntd.2020.100048
30. He J, Huang F, Zhang J, Chen Q, Zheng Z, Zhou Q, et al. Vaccine Design Based on 16 Epitopes of SARS-CoV-2 Spike Protein. *J Med Virol* (2021) 93:2115–31. doi: 10.1002/jmv.26596
31. Crooke SN, Ovsyannikova IG, Kennedy RB, Poland GA. Immunoinformatic Identification of B Cell and T Cell Epitopes in the SARS-CoV-2 Proteome. *Sci Rep* (2020) 10:14179. doi: 10.1038/s41598-020-70864-8
32. Behmar E, Soleymani B, Najafi A, Barzegari E. Immunoinformatic Design of a COVID-19 Subunit Vaccine Using Entire Structural Immunogenic Epitopes of SARS-CoV-2. *Sci Rep* (2020) 10:20864. doi: 10.1038/s41598-020-77547-4
33. Oliveira SC, de Magalhaes MTQ, Homan EJ. Immunoinformatic Analysis of SARS-CoV-2 Nucleocapsid Protein and Identification of COVID-19 Vaccine Targets. *Front Immunol* (2020) 11:587615. doi: 10.3389/fimmu.2020.587615
34. Dong R, Chu Z, Yu F, Zha Y. Contriving Multi-Epitope Subunit of Vaccine for COVID-19: Immunoinformatics Approaches. *Front Immunol* (2020) 11:1784. doi: 10.3389/fimmu.2020.01784
35. Lee E, Sandgren K, Duette G, Stylianou VV, Khanna R, Eden JS, et al. Identification of SARS-CoV-2 Nucleocapsid and Spike T-Cell Epitopes for Assessing T-Cell Immunity. *J Virol* (2021) 95:02002–20. doi: 10.1128/JVI.02002-20
36. Finkel Y, Mizrahi O, Nachshon A, Weingarten-Gabbay S, Morgenstern D, Yahalom-Ronen Y, et al. The Coding Capacity of SARS-CoV-2. *Nature* (2021) 589:125–30. doi: 10.1038/s41586-020-2739-1
37. Jiang HW, Zhang HN, Meng QF, Xie J, Li Y, Chen H, et al. SARS-CoV-2 Orf9b Suppresses Type I Interferon Responses by Targeting TOM70. *Cell Mol Immunol* (2020) 17:998–1000. doi: 10.1038/s41423-020-0514-8
38. Paul S, Sidney J, Sette A, Peters B. Tepitool: A Pipeline for Computational Prediction of T Cell Epitope Candidates. *Curr Protoc Immunol* (2016) 114:18 19 11–18 19 24. doi: 10.1002/cpim.12
39. Calis JJ, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, et al. Properties of MHC Class I Presented Peptides That Enhance Immunogenicity. *PLoS Comput Biol* (2013) 9:e1003266. doi: 10.1371/journal.pcbi.1003266
40. Paul S, Lindestam Arlehamn CS, Scriba TJ, Dillon MB, Oseroff C, Hinz D, et al. Development and Validation of a Broad Scheme for Prediction of HLA Class II Restricted T Cell Epitopes. *J Immunol Methods* (2015) 422:28–34. doi: 10.1016/j.jim.2015.03.022
41. Dhanda SK, Karosiene E, Edwards L, Grifoni A, Paul S, Andreatta M, et al. Predicting HLA CD4 Immunogenicity in Human Populations. *Front Immunol* (2018) 9:1369. doi: 10.3389/fimmu.2018.01369
42. Paul S, Weiskopf D, Angelo MA, Sidney J, Peters B, Sette A. HLA Class I Alleles are Associated With Peptide-Binding Repertoires of Different Size, Affinity, and Immunogenicity. *J Immunol* (2013) 191:5831–9. doi: 10.4049/jimmunol.1302101
43. Doytchinova IA, Flower DR. Vaxijen: A Server for Prediction of Protective Antigens, Tumour Antigens and Subunit Vaccines. *BMC Bioinf* (2007) 8:4. doi: 10.1186/1471-2105-8-4
44. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting Population Coverage of T-Cell Epitope-Based Diagnostics and Vaccines. *BMC Bioinf* (2006) 7:153. doi: 10.1186/1471-2105-7-153
45. Mukherjee S, Warwicker J, Chandra N. Deciphering Complex Patterns of Class-I HLA-Peptide Cross-Reactivity via Hierarchical Grouping. *Immunol Cell Biol* (2015) 93:522–32. doi: 10.1038/icb.2015.3

46. Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA Class I Supertypes: A Revised and Updated Classification. *BMC Immunol* (2008) 9:1. doi: 10.1186/1471-2172-9-1
47. Jespersen MC, Peters B, Nielsen M, Marcatili P. Bepipred-2.0: Improving Sequence-Based B-Cell Epitope Prediction Using Conformational Epitopes. *Nucleic Acids Res* (2017) 45:W24–9. doi: 10.1093/nar/gkx346
48. Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B Cell Epitope Predictions: Impacts of Method Development and Improved Benchmarking *Plos Comput Biol* (2012) 8:e1002829. doi: 10.1371/journal.pcbi.1002829
49. Rigsby RE, Parker AB. Using the Pymol Application to Reinforce Visual Understanding of Protein Structure. *Biochem Mol Biol Educ* (2016) 44:433–7. doi: 10.1002/bmb.20966
50. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The Immune Epitope Database (IEDB): 2018 Update. *Nucleic Acids Res* (2019) 47: D339–43. doi: 10.1093/nar/gky1006
51. Li Y, Xu Z, Lei Q, Lai DY, Hou H, Jiang HW, et al. Antibody Landscape Against SARS-Cov-2 Reveals Significant Differences Between non-Structural/ Accessory and Structural Proteins. *Cell Rep* (2021) 109391:109391. doi: 10.1016/j.celrep.2021.109391
52. Prachar M, Justesen S, Steen-Jensen DB, Thorgrimsen S, Jurgons E, Winther O, et al. Identification and Validation of 174 COVID-19 Vaccine Candidate

Epitopes Reveals Low Performance of Common Epitope Prediction Tools. *Sci Rep* (2020) 10:20465. doi: 10.1038/s41598-020-77466-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Uttamrao, Sathyaseelan, Patro and Rathinavelan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.