Strategies for Vaccine Design for Coronavirus All Variants of Concern Using Immunoinformatics Techniques

Aims: A short sequence of viral protein/peptide could be used as a potential vaccine to treat coronavirus. Considering all variants of concern (VOC), designing a peptide vaccine for severe acute respiratory syndrome coronavirus 2 (SARS CoV-2) is a challenging task for scientists.

Materials & Methods: In this study, an epitope-containing vaccine peptide in nonstructural protein 4 (nsp4) of SARS-CoV-2 was predicted. Using a modified method for both B and T cell epitope prediction (verified by molecular docking studies), linear B and T cell epitopes of nsp4 protein were predicted. Predicted epitopes were analyzed with population coverage calculation and epitope conservancy analysis.

Findings: The short peptide sequence QRGSYTDNKAP was selected as B-cell epitope by considering the scores of surface accessibility, hydrophilicity, and beta turn for each amino acid residue. Similarly, the peptide sequences FLAHIQWVM and FLAHIQWVMFTPL were predicted as T cell epitopes for MHC-I and MHC-II molecules. These two potential epitopes could favor HLA-A*02:01 and HLA-DRB*01:01 as MHC allelic proteins with the lowest IEDB values, respectively.

No amino acid mutations were observed in GISAID (global initiative on sharing all influenza data) database for alpha, beta, gamma, and delta variants of concerns. Among seven amino acid point mutations in nsp4 protein of omicron variant, none were present in the peptide sequences of the predicted epitopes.

Conclusion: Short peptide sequences could be predicted as vaccines to prevent infections caused by coronavirus variants of concerns.

Keywords: Peptide type vaccine design, COVID-19, B and T cell epitopes, MHC allelic protein, molecular docking study.

How to cite this article
Introduction
SARS-CoV-2 is a single stranded, positive-sense, RNA virus belonging to the Coronaviridae family, which has recently caused severe acute respiratory syndrome worldwide [1]. The virus is the causative agent of coronavirus disease 2019 (COVID-19), which is a very contagious disease [2]. It is transmitted through human-to-human contact [3]. To date (June 2, 2022), 11,150,230 genome sequences have been submitted to the GISAID (global initiative on sharing all influenza data) database. Therefore, it is a challenging task for scientists to design a peptide-based vaccine which is equally effective for all variants [4-11].

Genomic epidemiological studies on this novel coronavirus have indicated that the spread of new strains from their countries of origin occurs through human-to-human contact throughout the world [12]. Phylogenetic studies on SARS-CoV-2 strains have shown that different strains identified in different countries are phylogenetically associated with subtle differences in their genome sequence [13-18]. Thus, when choosing a target viral protein for vaccine design, mutations in amino acid sequences of different SARS-CoV-2 strains must be considered [19-27].

Different viral proteins of SARS-CoV-2 have been studied for the development of proposed vaccines against COVID-19, including surface glycoprotein [28,29], spike protein [30], envelop protein [31,32,33], and nucleocapsid phosphoprotein [34].

In the GISAID database, hCoV-19/Wuhan/WIV04/2019 (WIV04) is considered as a reference genome sequence because of its consensus with previously submitted sequences of the beta coronavirus responsible for COVID-19. This genome sequence has been isolated by the Wuhan Institute of Virology from a bronchoalveolar lavage fluid sample (BALF) collected from a human seafood market retailer [23].

RNA extraction data and metagenomic next-generation sequencing data reveal that there are 12 viral proteins encoded in the viral genome sequence [25]. Since the spike

---

Figure 1) Positions of point mutations in the spike protein of different strains of SARS-Cov-2

Infection Epidemiology and Microbiology  Summer 2022, Volume 8, Issue 3
glycoprotein (S glycoprotein) is responsible for the attachment of the virion to the host cell by interacting with a few host cell receptors, this protein sequence and mRNA sequence have been considered as target sequences for COVID vaccines by different pharmaceutical companies [20]. But as shown in the spike protein image (Figure 1) based on UCGC data, there are a large number of mutations in the spike protein sequence of coronavirus variants of concerns (VOCs), including alpha, beta, gamma, delta, and omicron strains. In contrast, similar data obtained from the UCGC database show that so far only seven amino acid mutations have been observed in nonstructural protein 4 (nsp4) of omicron strains, and no mutation has been observed in nsp4 protein of alpha, beta, gamma, and delta strains (Figure 2). Thus, nsp4 is considered as a target viral protein for peptide-based vaccine design.

The existing knowledge about this viral infection inside the human host reveals that the human host cell membrane rearrangement is required for the viral replication. During the replication process, viral genome duplication and viral mRNA transcription occur in the host cell using the human cell machinery. This viral genome consists of a leader sequence, which codes proteins during RNA replication. Nonstructural protein 4 (nsp4), encoded in the leader sequence, is responsible for the formation of a double-membrane vesicle during the viral replication process. But compared to other viral proteins such as papain-like protease (nsp3) [26], nsp2 [24], host translation inhibitor (nsp1), and 3-C-like proteinase (3CL-PRO) (nsp5), so far no experimentally verified structure has been reported for nonstructural protein 4 (nsp4). In this study, nsp4 was thoroughly examined to obtained information about epitopes that could aid in the vaccine development against the SARS-CoV-2 virus. Peptide-based vaccination knowledge was used to identify B-cell and T-cell epitopes that could trigger specific types of immune responses.

Figure 2) Positions of point mutations in nsp4 of different strains of SARS-Cov-2
in human body. Identification of these epitopes will help provide low-cost, high-quality protection against all VOCs of SARS-CoV-2. Potentially active immunogenic T and B cell epitopes found in this viral protein were modelled using immunoinformatics webtools in order to develop a viable peptide vaccine for coronavirus infection.

**Materials and Methods**

**Sequence retrieval and prediction of secondary structure of nonstructural protein 4 (nsp4):** Due to the lack of experimental X-ray crystallographic structure of nsp4 viral protein, its primary structure was first retrieved. The protein three-dimensional structure was predicted using homology modeling method. The quality of the predicted structure was checked by creating a Ramachandran plot. The antigenicity of the predicted protein was calculated using Vaxign-ML software. For this viral protein, a signal peptide and transmembrane regions were also detected using TMHMM probability calculation software [35].

**Prediction of potential B cell epitopes:** B cell epitopes are antigens that interact with B lymphocytes to trigger an immune response. The linear type of B cell epitope is expected to be among the two types of B cell epitopes. Various physicochemical features of peptide chains are evaluated to determine the sites of linear epitopes of an antigenic protein, including hydrophilicity, flexibility, accessibility, turns, exposed surface, polarity, and antigenic propensity. Two methods are used to predict B cell epitopes, considering the primary structure and antigen structure. Thus, various IEDB (Immune Epitope Database) (www.iedb.org) tools were used to predict linear or continuous B cell epitopes of nonstructural protein 4 in vivo, including classical propensity scale methods such as antigenicity scale of Kolaskar and Tongaonkar (1990) [37], surface accessibility prediction scale of Emini et al. (1985) [38], hydrophilicity prediction scale of Parker and colleagues (1986) [39], flexibility prediction scale of Karplus and Schulz (1985) [40], BepiPred linear epitope prediction method [41], and beta turn prediction scale of Chou and Fasman (1978) [42]. The most likely B cell epitope of the antigenic protein was discovered using graphical data and prediction scores. The BepiPred prediction approach combines a hidden Markov model and a propensity scale method to predict score and identify antigenic protein epitope in B cells [43].

**Prediction of the three-dimensional structure of the most potential B cell epitope:** PEP-FOLD uses a hidden Markov model derived structural alphabet for de novo modelling of 3D conformations of peptides between 9 and 25 amino acids in aqueous solution [49]. Updates to PEP-FOLD allow the modelling of linear and disulfide-bonded cyclic peptides of 9 to 36 amino acids using benchmarks [51].

**Molecular docking study of the predicted B cell epitope:** Molecular docking of the predicted B cell epitope with immune globulin G (IgG) molecule was executed with the help of ClusPro docking server using FFT algorithm [52].

**Prediction of MHC I and MHC II restricted T cell epitopes:** The goal of T-cell epitope prediction is to find the smallest peptides within an antigenic protein, which could trigger an immune response in CD4 or CD8 T-cells [44]. Immunogenicity refers to the ability to excite T-cells. Antigens contain a variety of peptides, and T-cell epitope prediction approaches seek to find those that are immunogenic. Three basic processes are required for T-cell epitope immunogenicity: (i) antigen processing, (ii) peptide interaction with major histocompatibility complex (MHC) molecules, and (iii) acceptance of MHC bound peptide by a related T-cell receptor (TCR). MHC bound processed antigenic peptides could form a ternary complex with the TCR molecule. MHC-peptide binding
is the most selective activity among these three activities for defining T-cell epitopes. Prediction scores are calculated by scanning the amino acid sequence of antigenic protein for the following parameters, such as binding of MHC-I and MHC-II molecules, antigen processing with proteasome complex and TAP (transporter associated with antigen processing), and MHC-I immunogenicity [47].

**Prediction of the three-dimensional structure of T cell epitopes**: PEP-FOLD uses a hidden Markov model-derived structural alphabet for de novo modelling of 3D conformations of peptides between 9 and 25 amino acids in aqueous solution [50]. Updates to PEP-FOLD allow the modelling of linear and disulfide-bonded cyclic peptides of 9 to 36 amino acids using benchmarks [51].

**Molecular docking study of the predicted T cell epitopes with ternary complex**: Molecular docking of the predicted T cell epitopes with the ternary complex was executed with the help of ClusPro docking server using FFT algorithm [52].

**Analysis of the predicted T cell epitopes**: Researchers could estimate the fraction of people predicted to respond to a particular set of peptides using the population coverage tool and determine the conservancy of a peptide inside a protein using the conservancy tool [49]. The population coverage tool informs vaccine researchers of their vaccine’s efficacy at regional and global levels, while the conservancy analysis tool finds conserved regions of proteins or antigens that could be used as vaccine targets.

**Study of predicted epitopes for immune reactivity based on experimental results**: The Immune Epitope Database (IEDB) is a freely available database. This database indexes experimental data on antibody and T cell epitopes studied in humans, non-human primates, and other animal species in relation to allergy, infectious diseases, transplantation, and autoimmunity [46].

### Results

**Sequence retrieval and prediction of secondary structure of nonstructural protein 4 (nsp4)**: Due to the lack of experimental X-ray crystallographic structure of nsp4 viral protein, its primary structure was first retrieved.

The protein three-dimensional structure was predicted using homology modeling method.

The quality of the predicted structure was checked by creating a Ramachandran plot.
The antigenicity of the predicted protein was calculated using Vaxign-ML software\textsuperscript{[35]}. Antigenicity prediction score of nsp4 using Vaxign-ML software was 89.6 \textsuperscript{[36]}.

For this viral protein, a signal peptide and transmembrane regions were also detected using TMHMM probability calculation software.

\textbf{Prediction of potential B cell epitopes}

\textbf{a) Prediction of linear B cell epitopes using antigen sequence properties:}

There is a correlation between the position of continuous epitopes and properties of polypeptide chains such as hydrophilicity, flexibility, accessibility, beta turns, exposed surface molecules, polarity, and antigenic propensity. Due to advances in empirical methods, the sites of continuous epitopes may now be predicted based on specific characteristics of the protein sequence. Propensity scales calculated for each of the 20 amino acids provide the foundation for all prediction analyses. Based on their relative tendency to possess the feature specified by the scale, 20 values of each scale are allocated to each amino acid residue.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.png}
\caption{Predicted signal peptide and transmembrane regions of nsp4}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{Result of BepiPred-2.0 prediction}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure5.png}
\caption{Predicted epitopes \textsuperscript{[43]}}
\end{figure}

Using a random forest algorithm based on epitopes and non-epitope amino acids identified in homology modeling-based structures, the BepiPred-2.0 service predicts B-cell epitopes in a protein sequence. After that, a sequential prediction is carried out. The BepiPred-2.0 prediction graph is colored yellow to indicate that residues with scores over the cutoff (the default value is 0.5) are projected to be a component of an epitope (where Y-axes represent residue scores, and X-axes represent residue positions in the sequence). The predicted epitopes are enlisted in Figure 5.
is commonly used to predict beta turns based on the primary structure of the vial protein sequence. The results of beta turns predicted for each amino acid residue are shown in Figure 6.

**Figure 6** Results of beta-turn prediction

c) **Results of Emini surface accessibility prediction:** The method described by Emini et al. (1985) was used to predict surface accessibility\[38\]. In this method, instead of using an add-on inside the window, the computation is based on a product’s surface accessibility scale. The accessibility profile is calculated using the following formula:

\[ S_n = (n+4+i) (0.37)^{-6} \]

Where \( S_n \) is the surface probability, and \( i \) ranges from 1 to 6. A higher probability of detection on the protein surface is indicated by hexapeptide sequences with \( S_n \) values larger than 1.0.

**Figure 7** Results of surface accessibility prediction

d) **Results of Parker hydrophilicity prediction:** Hydrophilicity prediction was performed according to the method previously described by Parker and colleagues (1986)\[39\]. In this method, using high-performance liquid chromatography (HPLC) on a reversed-phase column, a hydrophilicity scale is established based on the peptide retention duration in this procedure. For the analysis of the epitope region, a window of seven residues is used. Each of the seven residues receives the relevant scale value, and the fourth residue in the segment (i+3) receives the arithmetical mean of the seven residue values.

**Figure 8** Results of Parker hydrophilicity prediction

e) **Results of Karplus & Schulz flexibility prediction:** Flexibility prediction in this study was performed according to the method described by Karplus and Schulz (1985)\[40\]. The flexibility scale used in this method is built using the mobility of protein segments and the known temperature B factors of the \( \alpha \)-carbons in 31 proteins with known structures. There are three scales for describing flexibility instead of just one scale, and the calculation based on a flexibility scale is comparable to the classical calculation except that the first amino acid of the six-amino acid window length serves as the center.

**Figure 9. Results of Karplus & Schulz flexibility prediction**
Considering the above four parameters for predicting B cell epitopes (Table 1), the short peptide sequence 74RGGSYNTNDKA84 was predicted as a potential B cell epitope of nsp4 viral protein. The scores of the above-mentioned physicochemical parameters for the predicted B cell epitope are shown in Table 1.

**Prediction of the three-dimensional structure of the most potential B cell epitope:**
The following structure was predicted for the potential B cell epitope using PEPProteasome P-FOLD software peptides [41].

**Prediction of MHC I and MHC II restricted T cell epitopes**

**a) T cell epitopes – MHC I-binding prediction:** T-cell epitope prediction tools predict IC50 values for peptide binding to specific MHC molecules. Peptide binding to MHC is necessary but not sufficient for recognition by T cells. Structural components of viral particles known as antigens are what T-cells detect during an adaptive immune response instead of pathogens like viruses or bacteria as a whole. Only when these viral antigens are recognized by antigen-presenting cells (APCs), specialized T-cell receptors on their cell surfaces could identify them. Major histocompatibility complex (MHC) molecules are bound to these APCs. Viral proteins contain T-cell epitopes, which are recognized by two different MHC molecules. They are referred to as MHC molecules of classes I (MHC-I) and II (MHC-II). Both CD8 and CD4 T-cells, which are two different subsets of T-cells, are capable of recognizing these two types of MHC molecules. Consequently, T-cell epitopes are referred to as MHC-I and MHC-II T-cell epitopes. The most decisive phase in the selection of T cell epitopes is MHC-antigenic peptide binding. Thus, the basic method for determining possible T cell epitopes is to forecast MHC-epitope binding. Table 2 lists the predicted linear T-cell epitopes of the coronavirus nonstructural protein 4 (nsp4) for MHC-I binding along with their interacting MHC-I alleles and IC50 values. The projected output in units of IC50 nM is provided for a number of epitopes. Therefore, a lower IC50 value denotes a higher affinity of that specific interacting MHC-I allele. Peptides with IC50 values of 50, 500, and 5000 nM are regarded to have high, intermediate, and poor affinity, respectively.
**Table 1** The physicochemical parameters for amino acids presented in the predicted B cell epitope

<table>
<thead>
<tr>
<th>B cell epitope</th>
<th>Chou and Fashman beta turn score for each residue (threshold=0.950)</th>
<th>Emin surface accessibility score for each residue (threshold=1.000)</th>
<th>Karplus and Schulz flexibility score for each residue (threshold=0.977)</th>
<th>Parker hydrophilicity score for each residue (threshold=0.266)</th>
</tr>
</thead>
<tbody>
<tr>
<td>74</td>
<td>Q 1.149</td>
<td>1.201</td>
<td>1.068</td>
<td>1.271</td>
</tr>
<tr>
<td></td>
<td>R 1.216</td>
<td>1.858</td>
<td>1.088</td>
<td>3.629</td>
</tr>
<tr>
<td></td>
<td>G 1.293</td>
<td>2.173</td>
<td>1.096</td>
<td>4.671</td>
</tr>
<tr>
<td></td>
<td>G 1.226</td>
<td>1.811</td>
<td>1.084</td>
<td>4.486</td>
</tr>
<tr>
<td></td>
<td>S 1.309</td>
<td>1.487</td>
<td>1.065</td>
<td>4.629</td>
</tr>
<tr>
<td></td>
<td>Y 1.381</td>
<td>2.509</td>
<td>1.052</td>
<td>5.457</td>
</tr>
<tr>
<td></td>
<td>T 1.303</td>
<td>5.07</td>
<td>1.051</td>
<td>5.457</td>
</tr>
<tr>
<td></td>
<td>N 1.174</td>
<td>3.822</td>
<td>1.05</td>
<td>4.943</td>
</tr>
<tr>
<td></td>
<td>D 1.14</td>
<td>1.307</td>
<td>1.035</td>
<td>4.214</td>
</tr>
<tr>
<td></td>
<td>K 1.194</td>
<td>1.401</td>
<td>1.015</td>
<td>4.786</td>
</tr>
<tr>
<td>84</td>
<td>A 1.141</td>
<td>0.718</td>
<td>0.98</td>
<td>2.729</td>
</tr>
</tbody>
</table>

**Table 2** MHC I binding prediction score

<table>
<thead>
<tr>
<th>Epitope sequence no</th>
<th>Peptide start</th>
<th>Peptide end</th>
<th>Peptide</th>
<th>IC50</th>
<th>Allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>359</td>
<td>367</td>
<td>FLAHIQWMV</td>
<td>2.01</td>
<td>HLA-A*02:01</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>359</td>
<td>367</td>
<td>FLAHIQWMV</td>
<td>2.44</td>
<td>HLA-A*02:03</td>
</tr>
<tr>
<td>1</td>
<td>359</td>
<td>367</td>
<td>FLAHIQWMV</td>
<td>2.78</td>
<td>HLA-A*02:06</td>
</tr>
<tr>
<td>1</td>
<td>359</td>
<td>367</td>
<td>FLAHIQWMV</td>
<td>16.29</td>
<td>HLA-A*68:02</td>
</tr>
<tr>
<td>2</td>
<td>121</td>
<td>129</td>
<td>FLPRVFSAV</td>
<td>2.47</td>
<td>HLA-A*02:03</td>
</tr>
<tr>
<td>2</td>
<td>121</td>
<td>129</td>
<td>FLPRVFSAV</td>
<td>4.04</td>
<td>HLA-A*02:06</td>
</tr>
<tr>
<td>2</td>
<td>121</td>
<td>129</td>
<td>FLPRVFSAV</td>
<td>5.52</td>
<td>HLA-A*02:01</td>
</tr>
</tbody>
</table>
b) Combined prediction of proteasomal cleavage and TAP transport for MHC I-restricted epitopes: The combined proteasomal cleavage/TAP transport/MHC-I binding prediction tool combines prediction scores of proteasomal processing, TAP transport, and MHC binding in order to produce an overall score for each peptide's intrinsic potential as a T-cell epitope. Proteasome cleavage scores could be considered as logarithms of the total amount of cleavage site usage liberating the peptide C-terminus frequencies [47]. Similarly, the TAP score calculates an effective -log (IC_{50}) value for TAP binding of a peptide or its N-terminally extended precursors. It has been demonstrated that a peptide's strong affinity corresponds to its high transport rate. Proteasomal cleavage and TAP transport predictions are used to generate a processing score. The MHC score depicts the binding score of class I MHC molecules in the form of -log (IC_{50}) values. A peptide could bind to several MHC molecules in the ER (endoplasmic reticulum); thus, this prediction provides a quantity proportional to the amount of peptides there. This makes it possible to predict T-cell epitope candidates regardless of MHC constraint. Predictions for proteasomal cleavage, TAP transport, and MHC binding are combined to form an overall score. It makes a quantitative prediction based on the amount of peptides that MHC molecules provide on the cell surface [47].

Using the MHC-II binding prediction tool in the IEDB Analysis Resource, CD4+ T-cell receptor responses against nonstructural protein 4 (nsp4) peptides were projected and enlisted as CD4+ T cell epitopes in Table 4.

c) T cell epitopes - Immunogenicity prediction: The immunogenicity prediction tool in the IEDB Analysis Resource (iedb.org) makes predictions about the relative ability of a peptide-MHC complex to elicit an immune response. This tool applies amino acid properties as well as their position within the peptide to forecast the immunogenicity of a class I MHC-peptide complex. The predicted epitope FLAHIQWMVFTPLV obtained a high immunogenicity score, indicating the stimulation of an immunological response in the human body.

Prediction of the three-dimensional structure of T cell epitopes: The T-cell epitope FLAHIQWMVFTPLV (MHC-I restricted) was selected based on its interaction with a large number of alleles and the lowest IC_{50} value with the HLA-A*02:01 MHC-I allele. The predicted three-dimensional structure of this T cell epitope is shown in Figure 11.

Using the MHC-II binding prediction tool in the IEDB Analysis Resource, CD4+ T-cell receptor responses against nonstructural protein 4 (nsp4) peptides were projected and enlisted as CD4+ T cell epitopes in Table 4.

Figure 11) Three-dimensional structure of the predicted T cell epitope

Molecular docking study of the predicted T cell epitopes with ternary complex: Molecular docking study showed that class II MHC molecule (HLA-DRB4*01:01)-restricted T-cell epitope FLAHIQWMVFTPLV binds to T-cell receptor protein (TCR), and a ternary complex (class II MHC molecule-T cell epitope- TCR) is formed during antigen presentation of T-cell epitope. With the help of molecular docking study, the
binding energy of this ternary complex was calculated. The binding energy of this ternary complex was -986 Kcal/mole.

In Figure 12, the T-cell epitope is shown as a red helix along with the HLA-DRB4*01:01 molecule and TCR in green and yellow, respectively.

**Analysis of the predicted T cell epitopes:**
The predicted T cell epitopes were analyzed with two analysis tools such as population coverage and conservation across antigens.

---

**Table 3** Predicted score for Proteasomal cleavage and TAP transport for MHC class I restricted epitope

<table>
<thead>
<tr>
<th>Allele</th>
<th>Start</th>
<th>End</th>
<th>Peptide length</th>
<th>Peptide</th>
<th>Proteasome Score</th>
<th>TAP Score</th>
<th>MHC Score</th>
<th>Processing Score</th>
<th>Total Score</th>
<th>MHC IC50 [nM]</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-A*02:01</td>
<td>9</td>
<td>17</td>
<td>9</td>
<td>FLAHIQWMV</td>
<td>1.18</td>
<td>0.13</td>
<td>-0.28</td>
<td>1.31</td>
<td>1.04</td>
<td>1.9</td>
</tr>
<tr>
<td>HLA-A*02:06</td>
<td>9</td>
<td>17</td>
<td>9</td>
<td>FLAHIQWMV</td>
<td>1.18</td>
<td>0.13</td>
<td>-0.38</td>
<td>1.31</td>
<td>0.93</td>
<td>2.4</td>
</tr>
<tr>
<td>HLA-A*68:02</td>
<td>9</td>
<td>17</td>
<td>9</td>
<td>FLAHIQWMV</td>
<td>1.18</td>
<td>0.13</td>
<td>-0.93</td>
<td>1.31</td>
<td>0.38</td>
<td>8.6</td>
</tr>
</tbody>
</table>

**Table 4** MHC II binding score

<table>
<thead>
<tr>
<th>Allele</th>
<th>Start</th>
<th>End</th>
<th>Length</th>
<th>Method used</th>
<th>Peptide</th>
<th>Percentile Rank</th>
<th>Adjusted rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-DRB4*01:01</td>
<td>359</td>
<td>373</td>
<td>15</td>
<td>Consensus (comb. lib./smm/nn)</td>
<td>FLAHIQWMVMFTPLV</td>
<td>2.10</td>
<td>2.10</td>
</tr>
</tbody>
</table>

**Table 5** Immunogenicity prediction score for T cell epitopes

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Length</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLAHIQWMVMFTPLV</td>
<td>70</td>
<td>1.5946</td>
</tr>
</tbody>
</table>

**Table 6** Population coverage of Class I MHC epitope

<table>
<thead>
<tr>
<th>Population/area</th>
<th>Class I MHC restricted epitope coverage</th>
<th>average hit</th>
<th>pc90</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>29.66%</td>
<td>0.31</td>
<td>0.14</td>
</tr>
<tr>
<td>England</td>
<td>50.35%</td>
<td>0.51</td>
<td>0.2</td>
</tr>
<tr>
<td>India</td>
<td>15.9%</td>
<td>0.16</td>
<td>0.12</td>
</tr>
<tr>
<td>North America</td>
<td>47.79%</td>
<td>0.49</td>
<td>0.19</td>
</tr>
<tr>
<td>World</td>
<td>43.26%</td>
<td>0.45</td>
<td>0.18</td>
</tr>
<tr>
<td>Average</td>
<td>37.39%</td>
<td>0.38</td>
<td>0.17</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>12.9</td>
<td>0.13</td>
<td>0.03</td>
</tr>
</tbody>
</table>
Figure 12) T cell epitope with ternary complex

T cells distinguish a complex from a definite major histocompatibility complex (MHC) molecule and an exact pathogen-derived epitope. A given epitope elicits a response only in persons who could express particular MHC molecules capable of binding to that particular epitope. MHC molecules are tremendously polymorphic, and over a thousand different human MHC (HLA) alleles have been identified so far. Choosing several peptides with different HLA binding specificities provides enhanced coverage of the patient population targeted by peptide-based vaccines or diagnostics. The topic of population coverage with regard to MHC polymorphism is very complicated due to the fact that different HLA allele types are present at dramatically different frequencies in different ethnic classes of people HLA-A. The IEDB population coverage tool was used to compute the population coverage of the predicted epitopes. This tool is useful to estimate the fraction of individuals predicted to respond to a specified epitope set according to HLA genotypic frequencies. These frequencies are found in the Allele Frequency database for 115 countries and 21 different ethnicities in 16 different geographical areas. The results of population coverage of class I MHC restricted epitope with the selected MHC-I alleles are shown in Table 6. Similarly, the results of population coverage of class II MHC bound epitope are shown in Table 7.

Table 7) Population coverage of Class II MHC epitope

<table>
<thead>
<tr>
<th>population/area</th>
<th>class II MHC restricted T cell epitope coverage</th>
<th>average hit</th>
<th>pc90</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>15.54%</td>
<td>0.16</td>
<td>0.12</td>
</tr>
<tr>
<td>England</td>
<td>41.62%</td>
<td>0.44</td>
<td>0.17</td>
</tr>
<tr>
<td>India</td>
<td>24.56%</td>
<td>0.25</td>
<td>0.13</td>
</tr>
<tr>
<td>North America</td>
<td>34.63%</td>
<td>0.36</td>
<td>0.15</td>
</tr>
<tr>
<td>World</td>
<td>28.79%</td>
<td>0.30</td>
<td>0.14</td>
</tr>
<tr>
<td>Average</td>
<td>29.03</td>
<td>0.30</td>
<td>0.14</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>8.85</td>
<td>0.10</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Results of epitope conservancy analysis

The epitope conservancy analysis results showed that FLAHIQWMV epitope matched 12.50% of the protein sequence with 100% identity.

Study of predicted epitopes for immune reactivity based on experimental results

FLAHIQWMV is a linear peptide epitope (epitope ID 1311559) that has been studied as part of replicase polyprotein 1ab of SARS-CoV-2. This epitope has been studied for immune reactivity in six publications, eight T cell assays, and two MHC ligand assays.
WFSQRGGSY is a linear peptide epitope (epitope ID 72433) that has been studied as part of replicase polyprotein 1ab of SARS-CoV-1 and tested in six MHC ligand assays.

Discussion

Considering the emergency situation of SARS-CoV-2 infection pandemic, fast design and development of an effective vaccine is the most argent step to prevent infections caused by all variants of concern (VOCs). Because by applying the vaccine, the mortality rate and the severity of the infection could be controlled. Computational vaccine design technology has been successfully applied for eradication and treatment of yellow fever virus, human papillomavirus (HPV) virus, nipah virus (NiV), etc. However, for some other common viruses such as hepatitis C virus, dengue virus, human immunodeficiency virus, and all variants of coronavirus, no vaccine has been invented yet due to the lack of definite information about the life cycle of these viruses. Therefore, first of all, several computational techniques could be used for epitope mapping, which is the preliminary step for vaccine design to prevent infections caused by all coronavirus VOCs. Several computational studies have been performed to treat SARS-CoV-2 [63-66].

In this study, several immuninformatics and molecular docking methods were integrated to recognize potential epitopes of nonstructural protein 4 in coronavirus alpha, beta, gamma, delta, omicron, and lambda variants. Two potent T cell epitopes with the ability to bind to MHC-I molecules were projected. For MHC-I and MHC-II molecules, both 9- and 15-mer peptide structures were predicted using the IEDB suggested prediction method and modeled using the PEP-FOLD web server. The percentile rank and IC_{50} values were also analyzed by the SMM/ANN method [44, 45] covering all MHC class I super-types. The most effective epitopes along with their IC_{50} values are presented in Table 2.

Based on MHC-I binding prediction scores, the peptide with the lowest percentile rank and IC_{50} value is nominated due to its highest affinity with that interacting MHC-I allele [73]. The T-cell epitope FLAHIQWMV was considered as the most effective epitope based on its interaction with a large number of alleles and the lowest IC_{50} value with the HLA-A*02:01 MHC-I allele. For this epitope, the MHC-I processing score with that specific allele comprised a proteasome score of 1.18, a TAP score of 0.13, and a MHC IC_{50} value of 1.9 nm [47, 74]. This means that this specific epitope has the highest affinity to the HLA-A*02:01 MHC-I molecule during antigen presentation. Moreover, this epitope showed the highest population coverage for Indian and global populations when interacting with the selected MHC-I allele. Henceforth, this epitope is considered as the epitope of choice for CD8+ T cells.

Likewise, the presence of wider peptide binding grooves in MHC-II molecule than in MHC-I as well as 15-mer epitopes along with their IC_{50} values were investigated by smm/nn/sturnilo [61, 62] method and listed in Table 3. In the MHC-II binding prediction method, a 15-mer T-cell epitope sequence...

Table 8) Result of Epitope Conservancy Analysis

<table>
<thead>
<tr>
<th>Epitope #</th>
<th>Epitope name</th>
<th>Epitope sequence</th>
<th>Epitope length</th>
<th>Percent of protein sequence matches at identity &lt;= 100%</th>
<th>Minimum identity</th>
<th>Maximum identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ws-separated-0</td>
<td>FLAHIQWMV</td>
<td>9</td>
<td>12.50% (1/8)</td>
<td>0.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Infection Epidemiology and Microbiology  Summer 2022, Volume 8, Issue 3
(\text{FLAHIQWMVMFTPLV})^{373} \) in non-structural protein 4 displayed a percentile rank of 2.10 when interacting with the HLA-DRB4*01:01 MHC-II allele. This result approves that this peptide could be selected as a MHC-II T-cell epitope in the protein of interest.

For B-cell epitope identification, the prediction scores related to Emini surface accessibility \[^{38}\], Parker hydrophilicity \[^{39}\], Chou and Fasman beta turn \[^{42}\], and Karplus and Schulz flexibility \[^{40}\] were calculated for each residue of peptide \(74\text{QRGGSYTNDKA} \) (starting from sequence position 74 and ending at position 84 of the viral antigenic protein), predicting this epitope as the most potent B-cell epitope available.

An important factor in vaccine design is the distribution of selected HLA allelic protein \[^{75}\]. This distribution differs among human populations in different geographic regions of the world. The T-cell epitope predicted in this study, \(\text{FLAHIQWMV}^{369}\), was bind to the MHC-I HLA-A*02:01 allele, which is present among 29.66, 15.9, and 43.26\% of Chinese, Indian, and world populations. Therefore, it may be concluded that the predicted T-cell epitope must be specifically restricted with the dominant MHC molecule (present in the target populations in India, China, and the whole world) to be effective against coronavirus.

In the present scenario, very little knowledge is available to scientists about the experimental epitopes of SARS-CoV-2 proteins and their interactions with the human immune system. Many research studies have focused on computer-aided vaccine design based on structural proteins of SARS-CoV-2 virus, specifically the spike protein \[^{76, 77}\]. However, Crooke et al. (2020) \[^{64}\] identified ten unique proteins in the novel coronavirus in their study on immunoinformatics-based vaccine design. Along with the four structural proteins, five nonstructural proteins namely NSP3, NSP6, NSP7, NSP8, and NSP10 were also selected in their study based on predicted antigenicity score using Vaxigen 2.0 server \[^{35}\]. In their study, HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*24:02, HLA-B*07:02, HLA-B*08:01, HLA-B*27:05, HLA-B*40:01, HLA-B*58:01, and HLA-B*15:01 were selected as MHC-I allelic subtypes. In the present study, interacting MHC-I alleles were HLA-A*01:01, HLA-A*02:03, HLA-A*26:01, HLA-A*11:01, and HLA-A*03:01 for \(76\text{PTDTYTSVY}^{84}\) T-cell epitope. Furthermore, their predicted epitopes provided 74\% global population coverage. One epitope, MMISAGFSL, was projected to bind to HLA-A*02:01 with high affinity (IC\(_{50}\) = 6.9 nM) \[^{64}\]. Whereas in the current study, MHC-I T-cell epitope \(\text{FLAHIQWMV}^{369}\) was predicted to interact with the HLA-A*02:01 MHC-I allele. In their research work, 36 peptides of MHC II-binding T-cell epitopes were predicted to collectively provide 99\% population coverage with binding affinity to HLA-DRB5 allele (IC\(_{50}\) value 18 nM).

However, our predicted MHC-II T-cell epitope \(\text{FLAHIQWMVMFTPLV}^{373}\) in the viral protein nsp4 interacted with HLA-DRB4*01:01.

Crooke et al. (2020) \[^{64}\] suggested HLA-B*15:01 as the MHC-I allele with a relatively high binding affinity (average IC\(_{50}\) = 67.7 nM) in their molecular docking study with PDBID 3C9N. Considering the same criteria, our selected MHC-I allelic protein was HLA-A*02:01. Moreover, their candidate vaccine peptides possessed several terminal amino acid residues such as Phe, Tyr, and Leu. These residues could fit into the hydrophobic pocket of the HLA groove. In our molecular docking study on T-cell epitope FLAHIQWMV, the terminal amino acid was valine.

Computational biology-based methods
are used for designing multi-epitope vaccines against SARS-CoV-2 [63]. In their study, Crooke et al. (2020) identified seven conformational and several linear B cell epitopes. Similarly, 16 potential structural epitopes were detected in the spike protein structure of this virus [64]. In our research work, one linear B-cell epitope was predicted. This B-cell epitope, \( \text{QRGGSYTNDKA} \), was exposed to the viral protein surface, as compared to those identified in previous works [63, 64].

Conclusion

Only seven amino acid mutations have been observed in nonstructural protein 4 (nsp4) of omicron strains, and no mutation has been observed in nsp4 protein of alpha, beta, gamma, and delta strains. For these two specific positions, 74-84 and 359-373, no mutation has been observed in different VOCs. Thus, it could be concluded that these two epitopes could be used in peptide-based vaccine design for all SARS-CoV-2 variants of concerns (VOCs).

References

19. Zumla A, Chan JF, Azhar EI, Hui DS, Yuen KY.
34. Peters B, Sette A. Generating quantitative
71. Adeyemi O, Bircan A, Cici D, Islek B, Kilinc Z,