# Immuno-Informatics Quest against COVID-19/SARS-COV-2: Determining Putative T-Cell Epitopes for Vaccine Prediction

**Article** *in* Infectious Disorders - Drug Targets(Formerly Current Drug Targets - Infectious Disorders) · September 2020

DOI: 10.2174/1871526520666200921154149

CITATIONS
11

READS
73

**4 authors:**

Nahid Akhtar
Lovely Professional University
**28** PUBLICATIONS **169** CITATIONS
SEE PROFILE

Amit Joshi
Lovely Professional University
**34** PUBLICATIONS **139** CITATIONS
SEE PROFILE

Bhupender Singh
Lovely Professional University
**14** PUBLICATIONS **99** CITATIONS
SEE PROFILE

Vikas Kaushik
Lovely Professional University
**56** PUBLICATIONS **212** CITATIONS
SEE PROFILE

*Send Orders for Reprints to reprints@benthamscience.net*

*Infectious Disorders - Drug Targets*, 2020, *20*, 1-00 

**RESEARCH ARTICLE**

# Immuno-Informatics Quest against COVID-19/SARS-COV-2: Determining Putative T-Cell Epitopes for Vaccine Prediction

Nahid Akhtar[1], Amit Joshi[1], Bhupender Singh[1] and Vikas Kaushik[1,*]

[1]*School of Bioengineering and Biosciences, Lovely Professional University, Phagwara, Punjab, India*

**Abstract:** ***Background*:** Since December 2019, a novel coronavirus, SARS-CoV-2, has caused global public health issues after being reported for the first time in Wuhan province of China. So far, there have been approximately 14.8 million confirmed cases and 0.614 million deaths due to the SARS-CoV-2 infection globally, and still, numbers are increasing. Although the virus has caused a global public health concern, no effective treatment has been developed.

***Objective*:** One of the strategies to combat the COVID-19 disease caused by SARS-CoV-2 is the development of vaccines that can make humans immune to these infections. Considering this approach, in this study, an attempt has been made to design epitope-based vaccine for combatting COVID-19 disease by analyzing the complete proteome of the virus by using immuno-informatics tools.

***Methods*:** The protein sequence of the SARS-CoV-2 was retrieved and the individual proteins were checked for their allergic potential. Then, from non-allergen proteins, antigenic epitopes were identified that could bind with MHCII molecules. The epitopes were modeled and docked to predict the interaction with MHCII molecules. The stability of the epitope-MHCII complex was further analyzed by performing a molecular dynamics simulation study. The selected vaccine candidates were also analyzed for their global population coverage and conservancy among SARS-related coronavirus species.

***Results*:** The study has predicted 5 peptide molecules that can act as potential candidates for epitope-based vaccine development. Among the 5 selected epitopes, the peptide LRARSVSPK can be the most potent epitope because of its high geometric shape complementarity score, low ACE and very high response towards it by the world population (81.81% global population coverage). Further, molecular dynamic simulation analysis indicated the formation of a stable epitope-MHCII complex. The epitope LRARSVSPK was also found to be highly conserved among the SARS-CoV--2 isolated from different countries.

***Conclusion*:** The study has predicted T-cell epitopes that can elicit a robust immune response in the global human population and act as potential vaccine candidates. However, the ability of these epitopes to act as vaccine candidate needs to be validated in wet lab studies.

**ARTICLE HISTORY**

Received: April 28, 2020
Revised: July 25, 2020
Accepted: July 27, 2020

*DOI:*
*10.2174/1871526520666200921154149*

**Keywords:** Vaccines, COVID-19, coronavirus, Epitope, MHC class II, docking, simulation.

## 1. INTRODUCTION

Coronaviruses are single-stranded RNA viruses with envelope belonging to the family *Coronaviridae* and genus *Betacoronavirus* [1, 2]. They cause upper respiratory infections in humans and animals which are usually benign but the infections caused by Severe Acute Respiratory Syndrome (SARS) coronavirus and Middle East Respiratory Syndrome (MERS) coronavirus have shown that these viruses can be fatal too [3]. But since, December 2019, a novel coronavirus, SARS-CoV-2 has caused global public health issues after being reported for the first time in Wuhan, Hubei province of China. The virus was initially believed to be hosted by bats from where it transmitted to humans *via* pangolins sold at a seafood market in China [4, 5]. However, this claim regarding the origin of SARS-CoV-2 from pangolins at wet markets in China seems unlikely, according to the recent studies [6]. A recent study suggests that the virus was being circulated cryptically among humans for years before being discovered [6]. Among humans, the virus can transmit *via* human-to-human contact and droplets of the diseased [7]. As of July 20, 2020, there have been 14,348,858 confirmed cases and 603,691 deaths due to the COVID-2019 ("Situation reports," WHO.). While most of these cases have been initially report-

* Address correspondence to this author at the School of Bioengineering and Biosciences, Lovely Professional University, Phagwara, Punjab, India; E-mail: vikas.14664@lpu.co.in

1871-5265/20 $65.00+.00 

ed in China, the virus has started to infect people across the globe. So far, the infections due to the virus have been reported in 179 countries ("Situation reports," WHO). Despite the continuous effort, no effective treatment has been developed to cure the infection [2]. Antiviral drugs that inhibit neuraminidase like paramivir and oseltamivir have shown to be ineffective [9]. Hence, it is imperative to develop new strategies for the treatment of COVID-19 disease. Recently, chloroquine phosphate has been found to be effective against COVID-19 associated pneumonia in clinical trials [10]. However, there are studies that are concerned regarding the potential harm of this therapy, such as cardiac arrhythmia, and suggest conducting more randomized control trials before incorporating chloroquine and its derivatives in treatment guidelines globally [11]. Dexamethasone has also been found to reduce harmful effects and induce rapid recovery among patients [12]. Still, many drugs and vaccine candidates are under different stages of preclinical and clinical trials.

One of the strategies to combat the COVID-19 disease is the development of vaccines that can make humans immune to these infections. The use of *in silico* Immunoinformatics can help to predict and develop vaccines in a quick and cost--efficient manner that can elicit robust immune response [13, 14]. These *in silico* approaches can also help to cope with the genetic variations, which are a major limitation in the development of effective vaccines [14]. Similarly, a novel Immunoinformatics approach to design vaccine called epitope-based vaccine design has been making progress [14]. These epitope-based vaccines can have advantages like stability, safety, specificity and convenience in synthesis and storage [15]. Hence, in this study, an attempt has been made to design epitope-based vaccines against the COVID-19 disease. Previously, epitope-based vaccines were identified for Zika and Nipah virus also [16, 17]. Considering this approach, the protein sequence of the SARS-CoV-2 was obtained and checked for its allergic potential. Then, the non-allergen SARS-CoV-2 proteins were selected for the prediction of antigenic and non-toxic epitopes. Then, the epitopes were modeled to generate their quaternary structure and docked with their corresponding class II MHC alleles. With this approach, epitopes were designed that could bind strongly with MHCII molecules and enhance immunity in humans. The population coverage analysis was also done to predict the percentage of people that will respond to these epitope-based vaccines. Further analysis was done to confirm the conservation of the selected epitope among different serotypes and SARS virus that emerged in 2002. The discovery of a conserved epitope can help to design a vaccine that will be effective against all the types of acute respiratory syndromes caused by coronaviruses.

## 2. METHODS

### 2.1. Protein Sequence Retrieval

The protein sequence with accession number MN908947 was retrieved from NCBI GenBank. The protein sequence was obtained from a virus that was isolated from the bronchoalveolar lavage fluid of a seafood market worker experiencing severe respiratory problems. The genome of the virus is 29903 bp long ssRNA and has 10 coding sequences for protein [18]. The different proteins encoded by the virus and their corresponding protein ID have been listed in table **1**. All the 10 proteins were analyzed in this study.

**Table 1. Prediction of allergen/non-allergen proteins of Coronavirus.**

| Serial Number | Protein ID | Protein Product | AlgPred Prediction |
|---|---|---|---|
| 1 | QHD43415.1 | ORF1ab polyprotein | Allergen |
| 2 | QHD43416.1 | Surface glycoprotein | Potential allergen |
| 3 | QHD43417.1 | ORF3a protein | Non-allergen |
| 4 | QHD43418.1 | Envelope protein | Non-allergen |
| 5 | QHD43419.1 | Membrane glycoprotein | Non-allergen |
| 6 | QHD43420.1 | ORF6 protein | Allergen |
| 7 | QHD43421.1 | ORF7a protein | Non-allergen |
| 8 | QHD43422.1 | ORF8 protein | Allergen |
| 9 | QHD43423.2 | Nucleocapsidphosphoprotein | Non-allergen |
| 10 | QHD42199.1 | ORF10 protein | Allergen |

### 2.2. Allergenic Protein Prediction

AlgPred server was used to predict the allergenic potential of the retrieved protein sequences [19]. The protein sequences were uploaded in FASTA format and SVM (Support Vector Machine) module based on amino acid composition was chosen as a prediction approach. For further analysis, only non-allergen proteins were selected.

### 2.3. T cell Epitope Prediction

The non-allergen proteins were then selected to predict the peptide sequences that could bind with MHCII molecules. For the prediction of these epitopes, NETMHC2 2.3 server was used [20]. This server predicts the binding of various peptides to different MHC class II molecules, thus helping in the prediction of epitopes that could initiate the immune response by activating T helper cells [20]. The protein sequences were uploaded in FASTA format and HLA--DR were selected as loci. HLA-DR are antigen-presenting molecules located on the antigen-presenting cells [21]. Then, the predicted epitopes were subjected to the Vaxijen server to identify epitopes that could bind with MHCII molecules with high confidence level [13]. Vaxijen server is used to predict the antigen molecules on the basis of the physicochemical properties of the protein. Epitopes with Vaxijen Score $\geq$1.1 were considered to be of high confidence level and only these epitopes were selected for further analysis [17]. Epitopes with a Vaxijen score above 1 are considered very antigenic [22].

### 2.4. Toxicity Prediction of the Epitopes

The epitopes should be non-toxic. Hence, to confirm the non-toxicity of the selected epitopes, ToxinPred server was used [23].ToxinPred allows predicting the toxicity, designing peptides to increase or decrease their toxicity and identi-

fying toxic regions of proteins by using machine learning and the quantitative matrix method [23]. To predict the toxicity of the peptides, the SVM (Swiss-Prot) based method of the server was used.

### 2.5. Molecular Modeling of Epitopes and Human Leukocyte Antigen (HLA) Alleles

The tertiary structure of the epitopes was predicted using the PEPstrMOD server [24]. This server is used to model the tertiary structure of peptides having 7-25 residues based on the secondary structure information, beta turn information and molecular dynamics simulation [24]. For modeling the 3D structure of the HLA alleles, proteins with PDB ID 4AH2, 3C5J and 6CQL were used as a template for the alleles DRB1_0101, DRB1_0701 and DRB1_1301, respectively. Homology modeling using Swiss-model was used to design the 3D structures of these HLA alleles [25].

### 2.6. Molecular Docking of Epitopes and HLA Alleles

The docking of the HLA alleles and epitopes was performed using the PatchDock server [26]. This server allows the prediction of protein-protein interaction by finding transformations that generate the best molecular shape complementarity [26]. The peptides and receptors were uploaded to the server in PDB format. The clustering RMSD was set at 1.5 because protein-small ligand complex type was chosen while docking.

### 2.7. Binding Affinity Prediction and Population Coverage Analysis

The binding affinity of epitopes with MHCII molecules was predicted with MHCPred server [27]. Results with IC50 value above 5000 nM are considered non-binders and with IC50 value between 0.01-5000 nM are considered binders. The population coverage analysis to identify the percentage of people that will respond to the selected epitopes was performed using the Immune Epitope Database (IEDB) Population Coverage analysis tool [28]. These studies will help to analyze the population coverage of the predicted peptides.

### 2.8. Molecular dynamics and simulation analysis for a docked complex of epitope-MHC HLA allele

Molecular dynamics study was conducted to analyze RMSD values and atomic fluctuations for all amino acids under the 100 ps time frame by deploying the MDWeb server [29]. This server is based on popular simulation tools like Amber, GROMACS, and NAMD, *etc*,, and useful in setting up force-field for evaluating coarse-grained Brownian molecular trajectory and solvation system for interacting dynamics of selected epitopes with HLA Allelic proteins.

### 2.9. Alignment of the Selected Epitope to Confirm the Peptide Conservation

Clustal omega server was used to analyze the conservation of the selected peptide sequence. Clustal omega is a tool that generates sequence alignment among different sequences [30]. The ORF7a protein sequence of the various isolates of SARS-CoV-2 from 12 different countries and of SARS virus and MERS virus was retrieved from the NCBI GenBank database and aligned against each other to analyze the conservancy. The ORF7a protein was selected because the most potent peptide LRARSVSPK for vaccine candidates is part of this protein.

## 3. RESULTS

### 3.1. Allergenic Protein Prediction

AlgPred is a server that helps to predict the allergenic proteins with very high sensitivity. The server is based on an SVM module which uses the amino acid and dipeptide composition to predict the allergen protein. Out of the ten proteins encoded by the virus, 5 were found to be non-allergen.

The viral structural proteins like envelope protein, membrane glycoprotein and nulceocapsid phosphor protein that are necessary for the production of a complete viral particle [31] were found to be non- allergenic. The result of the allergen prediction of all the 10 proteins has been summarized in Table **1**.

### 3.2. T cell Epitope Prediction

For the prediction of epitopes, only the non-allergen proteins were selected. The epitopes predicted by the NETMHCII 2.3 server were in large number, so these epitopes were sorted on the basis of their binding level. Only those peptides were selected that could bind strongly with the HLA DRB molecules. Then, these strong binding peptides were subjected to the Vaxijen server to predict epitopes with a high level of confidence and further sort the epitopes. Altogether 17 epitopes were predicted that could bind MHCII molecules with high confidence (Vaxijen Score $\geq$1.1). Most of the epitopes (12) were found to be binding with HLA allele DRB1_1301. The predicted epitopes and their corresponding MHCII alleles along with the Vaxijen score have been summarized in Table **2**.

### 3.3. Toxicity Prediction of the Epitopes

It is imperative that the predicted epitopes are non-toxic. Out of all the epitopes that were predicted to be antigens, only two were found to be toxic. These toxic peptides were not considered for further analysis. The toxicity of the different epitopes has been summarized in Table **3**.

### 3.4. Modeling of the Epitopes, HLA Alleles and Their Molecular Docking

The quaternary structures of the epitopes were modeled by the PEPstrMOD server. For modeling the 3D structure of the HLA alleles, proteins with PDB ID4AH2, 3C5J and 6CQL were used as template for the alleles DRB1_0101, DRB1_0701 and DRB1_1301, respectively. The modeled HLA alleles have been represented in (Fig. **1**). All the 15 non-toxic epitopes were docked with their corresponding HLA alleles as predicted by NETMHCII 2.3. The highest geometric

**Table 2. Epitopes predicted by NETMHCII 2.3 server and their Vaxijen score.**

| Protein ID | Peptide | Allele | Binding Affinity (nM) | Vaxijen | Antigen/Non-antigen |
|---|---|---|---|---|---|
| QHD43417.1 | LALSKGVHF | DRB1_0101 | 678.7 | 1.1343 | Antigen |
| | YFLQSINFV | DRB1_0101 | 862.9 | 1.1339 | Antigen |
| | LALSKGVHF | DRB1_0701 | 40.9 | 1.1343 | Antigen |
| | ITLKKRWQL | DRB1_1301 | 32.8 | 1.9347 | Antigen |
| | LQSINFVRI | DRB1_1301 | 50.3 | 1.4795 | Antigen |
| | WKCRSKNPL | DRB1_1301 | 54.5 | 1.2717 | Antigen |
| | LALSKGVHF | DRB1_1301 | 54.1 | 1.1343 | Antigen |
| | WLCWKCRSR | DRB1_1301 | 64.4 | 2.4498 | Antigen |
| | IITLKKRWQ | DRB1_1301 | 68.3 | 1.7866 | Antigen |
| | FTIGTVTLK | DRB1_1301 | 68.4 | 2.0317 | Antigen |
| QHD43418.1 | CNIVNVSLV | DRB1_0101 | 1579.5 | 1.1201 | Antigen |
| QHD43419.1 | LVIGAVILR | DRB1_1301 | 24.4 | 1.1027 | Antigen |
| QHD43421.1 | VYQLRARSV | DRB1_0101 | 206.0 | 1.3108 | Antigen |
| | TLCFTLKRK | DRB1_1301 | 19.7 | 2.9229 | Antigen |
| | LRARSVSPK | DRB1_1301 | 32.6 | 1.371.7 | Antigen |
| | LCFTLKRKT | DRB1_1301 | 41.2 | 3.0945 | Antigen |
| | ITLCFTLKR | DRB1_1301 | 59.0 | 2.0208 | Antigen |

**Table 3. Toxicity of different epitope/peptides.**

| Peptide | SVM Score | Toxin/Non-toxin |
|---|---|---|
| LALSKGVHF | -1.30 | Non-toxin |
| YFLQSINFV | -1.00 | Non-toxin |
| ITLKKRWQL | -1.54 | Non-toxin |
| LQSINFVRI | -1.25 | Non-toxin |
| WKCRSKNPL | 0.16 | Toxin |
| WLCWKCRSR | 0.05 | Toxin |
| IITLKKRWQ | -1.27 | Non-toxin |
| FTIGTVTLK | -1.36 | Non-toxin |
| CNIVNVSLV | -0.75 | Non-toxin |
| LVIGAVILR | -0.58 | Non-toxin |
| VYQLRARSV | -1.07 | Non-toxin |
| TLCFTLKRK | -1.24 | Non-toxin |
| LRARSVSPK | -0.80 | Non-toxin |
| LCFTLKRKT | -1.12 | Non-toxin |
| ITLCFTLKR | -1.32 | Non-toxin |

**Table 4. Geometric shape complementarity score, interface area and the ACE of the different peptides predicted by PatchDock.**

| Serial number | Peptide | Allele | Score | Area | Atomic Contact Energy (ACE) |
|---|---|---|---|---|---|
| 1 | LALSKGVHF | DRB1_0101 | 8656 | 1184.4 | -112.06 |
| 2 | YFLQSINFV | DRB1_0101 | 8910 | 1029.7 | -308.58 |
| 3 | LALSKGVHF | DRB1_0701 | 8742 | 1100.5 | -0.73 |
| 4 | ITLKKRWQL | DRB1_1301 | 11166 | 1546.5 | -38.44 |
| 5 | LQSINFVRI | DRB1_1301 | 10316 | 1361.9 | -244.33 |
| 6 | LALSKGVHF | DRB1_1301 | 9358 | 1120.7 | -89.80 |
| 7 | IITLKKRWQ | DRB1_1301 | 10384 | 1394.4 | -52.54 |
| 8 | FTIGTVTLK | DRB1_1301 | 9850 | 1209.4 | -269.64 |
| 9 | CNIVNVSLV | DRB1_0101 | 7534 | 981.2 | -184.89 |
| 10 | LVIGAVILR | DRB1_1301 | 9942 | 1272.7 | -24.48 |
| 11 | VYQLRARSV | DRB1_0101 | 7482 | 994.3 | -118.98 |
| 12 | TLCFTLKRK | DRB1_1301 | 10102 | 1394.3 | -38.23 |
| 13 | LRARSVSPK | DRB1_1301 | 9220 | 1258.5 | -212.81 |
| 14 | LCFTLKRKT | DRB1_1301 | 9250 | 1232.1 | -56.58 |
| 15 | ITLCFTLKR | DRB1_1301 | 10284 | 1238.4 | -166.16 |

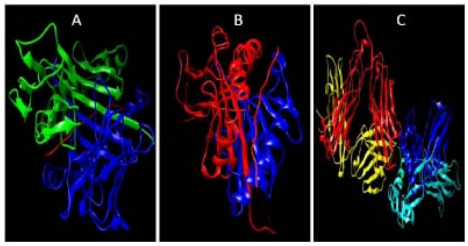


**Fig. (1). A.** Template structure of 3C5J used to model HLA-DRB1*07:01 **B.** Template structure of 4AH2 used to model HLA-DRB1*01:01 **C**. Template structure of 6CQL used to model HLA-DRB1*13:01. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

shape complementarity score (11166 was obtained while docking peptide ITLKKRWQL with DRB1_1301. The lowest ACE (-308.58) was obtained for the docking of epitope YFLQSINFV with DRB1_0101. The docking results of all the peptides that were predicted to be non-allergen, non-toxic and antigen are summarized in Table **4**. The comparison of the geometric shape complementarity score, interface area and the ACE has been shown in Fig. (**2**), Fig. (**3**) and Fig. (**4**) respectively.

### 3.5. Binding Affinity Prediction and Population Coverage Analysis

For further analysis, 5 epitopes (YFLQSINFV, LQSINFVRI, FTIGTVTLK, LRARSVSPK and ITLCFTLKR) with the highest geometric shape complementarity score and lowest atomic contact energy were selected. The complexes with the highest geometric shape complementarity score had the least steric hindrances and wide interface areas [26]; the lower atomic contact energy implies that the complex is more stable and favorable due to the low desolvation energy [32]. The docking of the five selected epitopes with their respective HLA alleles has been shown in Fig. (**5**). The binding affinity of these peptides with different HLA alleles was determined and the results are shown in Table **5**. All the epitopes showed the ability to bind with the class II MHC molecules except the peptide ITLCFTLKR which did not bind with HLA-DRB1*07:01. Worldwide, 38.05% people will respond to the peptides YFLQSINFV, LQSINFVRI, and FTIGTVTLK. Against the peptide LRARSVSPK, 81.81% people will respond and against the peptide ITLCFTLKR, only 22.06% people will respond. The summary of the population coverage analysis for different areas and ethnicities has been represented in (Fig. **6-8**).

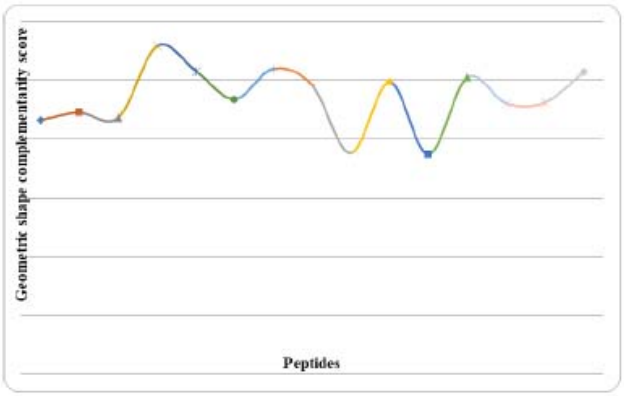


**Fig. (2).** Comparison of different epitopes and their geometric shape complementarity score. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

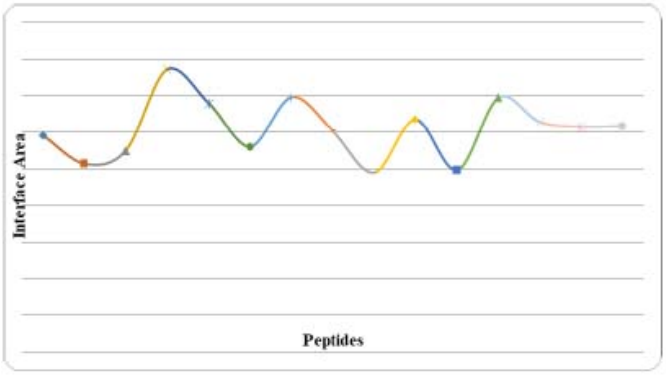


**Fig. (3).** Comparison of different epitopes and the interface area of the formed complexes. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

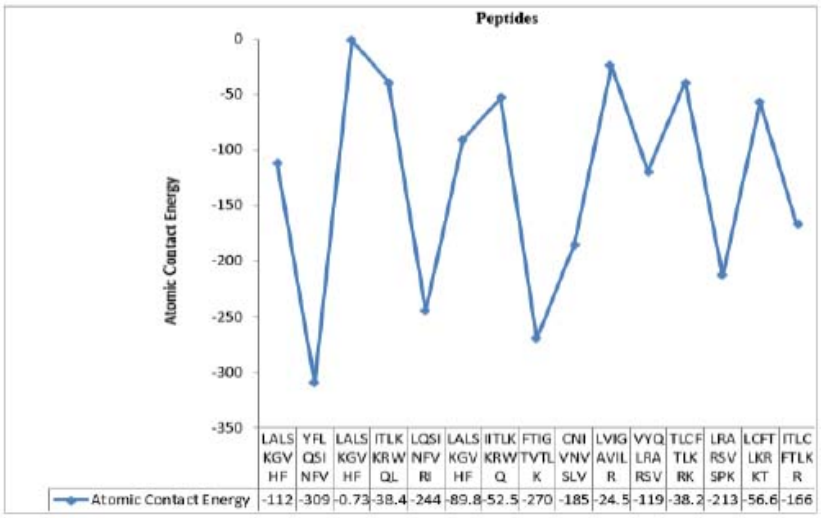


**Fig. (4).** Comparison of the atomic contact energy between the epitopes and their receptors. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

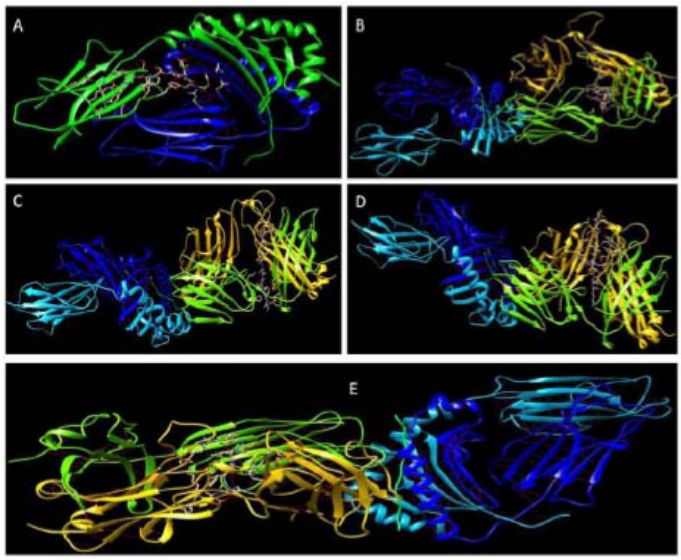


**Fig. (5). A**. Docking of epitope YFLQSINFV with 4AH2 receptor **B**. Docking of epitope LQSINFVRI with 6CQL receptor **C.** Docking of epitope FTIGTVTLK with 6CQL receptor **D.** Docking of epitope LRARSVSPK with 6CQL receptor **E.** Docking of epitope ITLCFTLKR with 6CQL receptor. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

**Table 5. The binding affinity of epitopes with different HLA alleles predicted by the MHC Pred tool.**

| EPITOPE | Number of HLA Binders | HLA with Predicted IC50 Value (nM) |
|---|---|---|
| YFLQSINFV | 3 | HLA-DRB1*01:01 (3.97), HLA-DRB1*04:01 (371.54), HLA-DRB1*07:01 (554.50) |
| LQSINFVRI | 3 | HLA-DRB1*01:01 (98.63), HLA-DRB1*04:01 (1713.96), HLA-DR-B1*07:01 (143.22) |
| FTIGTVTLK | 3 | HLA-DRB1*01:01 (25.82) HLA-DRB1*04:01 (135.21) HLA-DRB1*07:01 (986.28) |
| LRARSVSPK | 3 | HLA-DRB1*01:01 (234.42), HLA DRB1*04:01 (1101.54), HLA-DR-B1*07:01 (1224.62) |
| ITLCFTLKR | 2 | HLA-DRB1*01:01 (16.07), HLA-DRB1*04:01 (319.89), HLA-DRB1*07:01 (non-binder) |

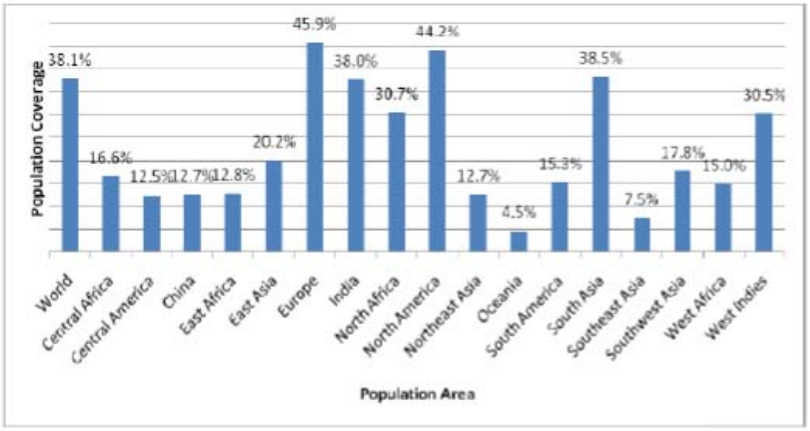


**Fig. (6).** Population Coverage of peptides YFLQSINFV, LQSINFVRI, and FTIGTVTLK. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

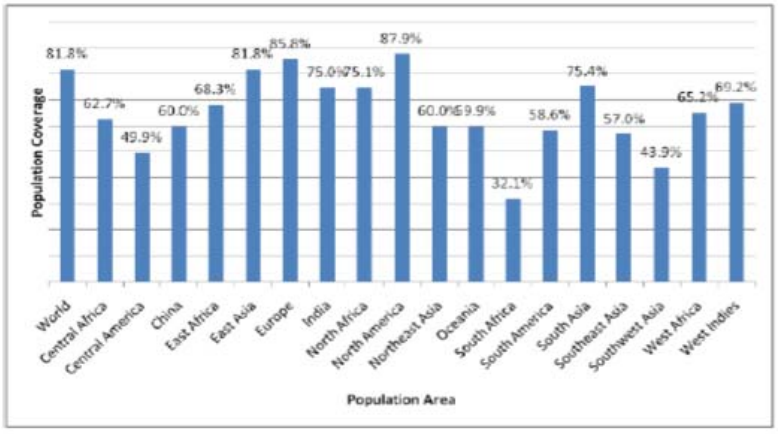


**Fig. (7).** Population Coverage of peptide LRARSVSPK. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

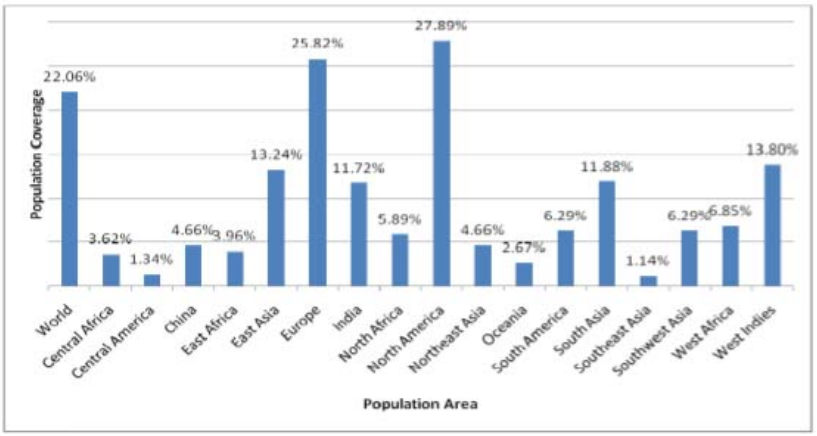


**Fig. (8).** Population Coverage of peptide ITLCFTLKR. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

### 3.6. Molecular Simulation Studies

(Fig. **9**) indicates the stability of the peptide MHC complex as evaluated with the help of RMSD and B- factor.

### 3.7. Alignment of the Selected Epitope to Confirm the Peptide Conservation

It is imperative that the vaccine candidate is effective against various genetic variants of the virus. Hence, to confirm the conservancy of the most potent epitope LRARSVSPK, the sequence of protein-encoding ORF7a from various members of the SARS-related coronavirus species was aligned using CLUSTAL omega server. The results show that the epitope is conserved among the 12 isolates of the SARS-CoV-2 from different countries and SARS virus that caused an epidemic in 2002-2004. However, it was not found to be conserved in the MERS virus. The country of isolation and the protein ID of ORF7a protein have been listed in Table **6**.

## 4. DISCUSSION

As new cases of SARS-CoV-2 infection are on rise, the infection has been declared a pandemic by the WHO; therefore, it is imperative to develop novel and effective antiviral therapies and vaccines. At the time of preparation of the manuscript, 23 vaccine candidates were found to be in different clinical trial stages and 137 candidates in preclinical stages [33]. One of the leading vaccine candidates, ChAdOx1 (Chimpanzee Adenovirus-vectored vaccine), has shown safety and ability to boost antibody production in the clinical trial stage 1/2 [33]. The ChAdOx1 vaccine consists of replication-deficient adenovirus vector containing full-length spike protein of SARS-CoV-2 along with tissue plasminogen activator leader sequence [33].

Using *in silico* Immunoinformatics approach, the viral proteins can be analyzed profoundly and epitopes can be predicted that elicit an effective immune response against several strains of viruses in an economical and rapid manner [34].

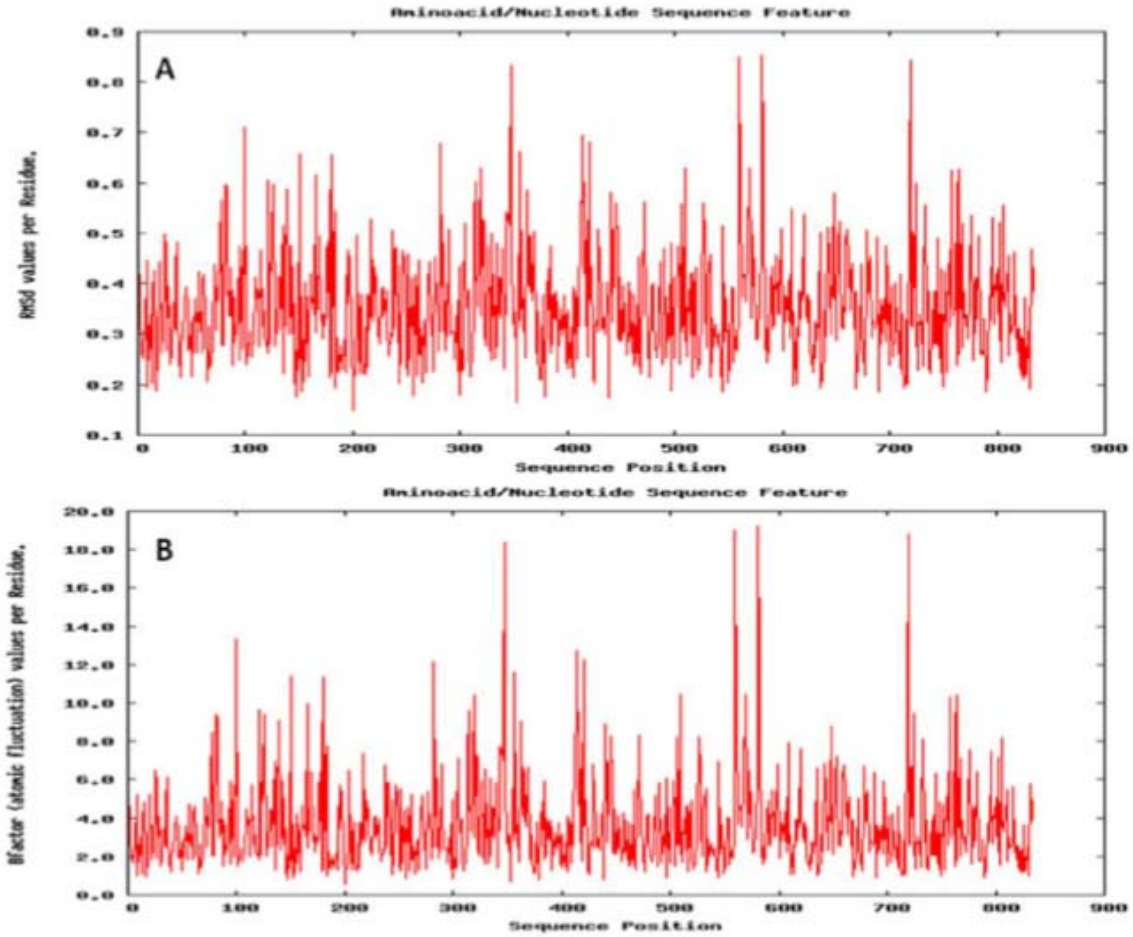


**Fig. (9). A.** RMSD Plot for LRARSVSPK - HLA complex, for each amino acid residue by molecular dynamics analysis, **B.** B-Factor (atomic fluctuation) values per amino acid residue for Epitope LRARSVSPK - HLA-A*68:01 docked complex. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

**Table 6. Protein ID of ORF7a proteins from different SARS-related coronavirus and their country of isolation.**

| Protein ID | SPECIES/COUNTRY OF ISOLATION |
|---|---|
| NP_828857.1 | SARS/Canada |
| QIK50443.1 | SARS-COV-2/Vietnam |
| QIG55999.1 | SARS-COV-2/Brazil |
| QIC53209.1 | SARS-COV-2/Sweden |
| QIB84678.1 | SARS-COV-2/Nepal |
| QIA98611.1 | SARS-COV-2/Taiwan |
| QIA98559.1 | SARS-COV-2/Italy |
| QHZ00384.1 | SARS-COV-2/South Korea |
| QHS34551.1 | SARS-COV-2/India |
| QHR84454.1 | SARS-COV-2/Australia |
| QHO60599.1 | SARS-COV-2/USA |
| BCA87366.1 | SARS-COV-2/Japan |
| QHD43421.1 | SARS-COV-2/China |
| NC_019843.3 | MERS/Saudi Arabia |

These *in silico* approaches can also help to cope with the genetic variations, which are a major limitation in the development of effective vaccines [14]. The epitope-based vaccine can also be effective against viruses that show a high degree of antigenic shift and drift [35]. Hence, considering these advantages of the epitope-based peptide vaccines, this study aimed to predict a novel, specific and effective vaccine against the SARS-CoV-2 virus. The entire proteome of the virus was analyzed and five proteins, namely ORF3a protein, membrane glycoprotein, surface glycoprotein, ORF7a protein and nucleocapsid phosphoprotein were found to be non-allergenic. These structural viral proteins play an important role in the attachment, integration and determination of antigenicity of the virus [22]. From these non-allergen proteins, 15 non-toxic peptides were predicted that could bind with MHC II molecules and activate T helper cells. The T cell epitopes were predicted because they play an important role in the initiation of the immune response [22]. The activation of T helper cells will further activate the B cells, macrophages and cytotoxic T cells which will eventually elicit a robust immune response [36]. All these epitopes had Vaxijen score above 1.1 which implies that they can be highly antigenic. Then, these highly antigenic and non-toxic epitopes were analyzed for their binding affinity and interaction with the MHCII receptors. For this purpose, molecular docking was performed and on the basis of the geometric shape complementarity score and atomic contact energy, five of the best epitopes were predicted as best vaccine candidates. The epitopes with highest geometric shape complementarity score and lowest atomic contact energy when docked with the MHC receptors were selected because the complexes with the highest geometric shape complementarity score had least steric hindrances and wide interface area [26]; the lower atomic contact energy implies that the complex is more stable and favorable due to the low desolvation energy [32]. These five best epitopes were YFLQSINFV, LQSINFVRI, FTIGTVTLK, ITLCFTLKR and LRARSVSPK. On further analysis, these epitopes were also found to bind with other MHC II receptors (HLA alleles) like HLA-DRB1*01:01, HLA-DRB1*04:01 and HLA-DRB1*07:01. Binding with more HLA alleles means that these epitopes will have more population coverage [22]. Further, to corroborate the wider population coverage of these epitopes, population coverage analysis was done. This analysis showed that the epitope LRARSVSPK will have 81.81% population coverage globally while other peptides will not achieve much significant global population coverage. Moreover, this epitope was found to be conserved among SARS-CoV-2 isolates from different countries. As, COVID-19 has been declared global pandemic, this epitope, LRARSVSPK, shows the potential to be effective vaccine candidate to protect the global population against this serious threat. However, this claim needs to be further validated by using various wet lab studies and clinical trials. Recently, epitopes for vaccine candidates against Dengue virus have also been identified by using Immunoinformatics method [37]. Dexamethasone (6mg) was also found to reduce harmful effects and induce rapid recovery among patients [12]. Still many drugs and vaccine candidates are under different stages of preclinical and clinical trials. Our study has discussed quick and fast epitope screening methods along with putative vaccine candidates based on modern *in-silico* tools to assist future research and wet lab validations. As the epitope, LRARSVSPK, was found to be non-toxic, highly antigenic and to elicit an immune response in most of the world's population, its stability upon binding with the MHC II receptor was analyzed by molecular dynamics simulation study. The molecular dynamics simulation shows the epitope-MHC II complex to be stable.

## CONCLUSION

Five different epitopes (YFLQSINFV, LQSINFVRI, FTIGTVTLK, LRARSVSPK and ITLCFTLKR) were identified that could act as potential vaccine candidates against COVID-19 based on their geometric shape complementarity score and atomic contact energy with their corresponding class II MHC molecules. Among these peptides, the peptide LRARSVSPK can be the most potent epitope because of its high geometric shape complementarity score, low ACE and very high response towards it by the world's population (81.81%). All these selected peptides were also found to be non-toxic and non-allergenic. These peptides also showed the ability to bind with the highest number of HLA alleles.

## AUTHORS' CONTRIBUTIONS

NA and VK conceived the study. NA, AJ, BS and VK performed the computational work. NA wrote the manuscript. All the authors read and approved the final manuscript.

## ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The authors did not perform any experiments on humans or animals.

## HUMAN AND ANIMAL RIGHTS

Not applicable.

## CONSENT FOR PUBLICATION

Written informed consent has been provided.

## AVAILABILITY OF DATA AND MATERIALS

Not applicable.

## FUNDING

The authors did not receive any funding for conducting this research.

## CONFLICT OF INTEREST

The authors have no conflicts of interest, financial or otherwise.

## ACKNOWLEDGEMENTS

Declared none.

## REFERENCES


[1] Habibzadeh, P.; Stoneman, E.K. The Novel Coronavirus: A Bird's Eye View. *Int. J. Occup. Environ. Med.,* **2020**, *11*(2), 65-71. [http://dx.doi.org/10.15171/ijoem.2020.1921] [PMID: 32020915]

[2] Lai, C-C.; Shih, T-P.; Ko, W-C.; Tang, H.J.; Hsueh, P.R. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int. J. Antimicrob. Agents,* **2020**, *55*(3), 105924. [http://dx.doi.org/10.1016/j.ijantimicag.2020.105924] [PMID: 32081636]

[3] Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; Cheng, Z.; Yu, T.; Xia, J.; Wei, Y.; Wu, W.; Xie, X.; Yin, W.; Li, H.; Liu, M.; Xiao, Y.; Gao, H.; Guo, L.; Xie, J.; Wang, G.; Jiang, R.; Gao, Z.; Jin, Q.; Wang, J.; Cao, B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet,* **2020**, *395*(10223), 497-506. [http://dx.doi.org/10.1016/S0140-6736(20)30183-5] [PMID: 31986264]

[4] Zu, Z.Y.; Jiang, M.D.; Xu, P.P.; Chen, W.; Ni, Q.Q.; Lu, G.M.; Zhang, L.J. Coronavirus Disease 2019 (COVID-19): A Perspective from China. *Radiology,* **2020**, *296*(2), E15-E25. [http://dx.doi.org/10.1148/radiol.2020200490] [PMID: 32083985]

[5] Lam, TT-Y; Shum, MH-H; Zhu, H-C Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China *bioRxiv,* **2020**.

[6] Chaw, S.M.; Tai, J.H.; Chen, S.L.; Hsieh, C.H.; Chang, S.Y.; Yeh, S.H.; Yang, W.S.; Chen, P.J.; Wang, H.Y. The origin and underlying driving forces of the SARS-CoV-2 outbreak. *J. Biomed. Sci.,* **2020**, *27*(1), 73. [http://dx.doi.org/10.1186/s12929-020-00665-8] [PMID: 32507105]

[7] Chang, D.; Lin, M.; Wei, L.; Xie, L.; Zhu, G.; Dela Cruz, C.S.; Sharma, L. Epidemiologic and Clinical Characteristics of Novel Coronavirus Infections Involving 13 Patients Outside Wuhan, China. *JAMA,* **2020**, *323*(11), 1092-1093. [http://dx.doi.org/10.1001/jama.2020.1623] [PMID: 32031568]

[8] *Situation reports,* https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

[9] Li, H.; Wang, Y.M.; Xu, J.Y.; Cao, B. [Potential antiviral therapeutics for 2019 Novel Coronavirus]. *Zhonghua Jie He He Hu Xi Za Zhi,* **2020**, *43*(0), E002. [PMID: 32023685]

[10] Gao, J.; Tian, Z.; Yang, X. Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. *Biosci. Trends,* **2020**, *14*(1), 72-73. [http://dx.doi.org/10.5582/bst.2020.01047] [PMID: 32074550]

[11] Pastick, KA; Okafor, EC; Wang, F Review: Hydroxychloroquine and Chloroquine for Treatment of SARS-CoV-2 (COVID-19) *Open Forum Infect. Dis.,* **2020**, *7*(4), ofaa130.

[12] Lester, M.; Shahin, A.; Pasyar, A. The use of dexamethasone in the treatment of COVID-19. *Ann. Med. Surg. (Lond.),* **2020**, ••• [http://dx.doi.org/10.1016/j.amsu.2020.07.004]

[13] Doytchinova, I.A.; Flower, D.R. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics,* **2007**, *8*, 4. [http://dx.doi.org/10.1186/1471-2105-8-4] [PMID: 17207271]

[14] Oyarzún, P.; Kobe, B. Recombinant and epitope-based vaccines on the road to the market and implications for vaccine design and production. *Hum. Vaccin. Immunother.,* **2016**, *12*(3), 763-767. [http://dx.doi.org/10.1080/21645515.2015.1094595] [PMID: 26430814]

[15] Lei, Y.; Zhao, F.; Shao, J.; Li, Y.; Li, S.; Chang, H.; Zhang, Y. Application of built-in adjuvants for epitope-based vaccines. *PeerJ,* **2019**, *6*, e6185. [http://dx.doi.org/10.7717/peerj.6185] [PMID: 30656066]

[16] Sharma, P.; Kaur, R.; Upadhyay, A.K. In-Silico Prediction of Peptide Based Vaccine Against Zika Virus. *Int. J. Pept. Res. Ther.,* **2019**, *26*, 85-91. [http://dx.doi.org/10.1007/s10989-019-09818-2]

[17] Kaushik, V. In Silico Identification of Epitope-Based Peptide Vaccine for Nipah Virus. *Int. J. Pept. Res. Ther.,* **2020**, *26*, 1147-1153. [http://dx.doi.org/10.1007/s10989-019-09917-0]

[18] Wu, F.; Zhao, S.; Yu, B.; Chen, Y.M.; Wang, W.; Song, Z.G.; Hu, Y.; Tao, Z.W.; Tian, J.H.; Pei, Y.Y.; Yuan, M.L.; Zhang, Y.L.; Dai, F.H.; Liu, Y.; Wang, Q.M.; Zheng, J.J.; Xu, L.; Holmes, E.C.; Zhang, Y.Z. A new coronavirus associated with human respiratory disease in China. *Nature,* **2020**, *579*(7798), 265-269. [http://dx.doi.org/10.1038/s41586-020-2008-3] [PMID: 32015508]

[19] Saha, S.; Raghava, G.P.S. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res.,* **2006**, *34*(Web Server issue), W202-9. [http://dx.doi.org/10.1093/nar/gkl343] [PMID: 16844994]

[20] Jensen, K.K.; Andreatta, M.; Marcatili, P.; Buus, S.; Greenbaum, J.A.; Yan, Z.; Sette, A.; Peters, B.; Nielsen, M. Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology,* **2018**, *154*(3), 394-406. [http://dx.doi.org/10.1111/imm.12889] [PMID: 29315598]

[21] Saraiva, D.P.; Jacinto, A.; Borralho, P.; Braga, S.; Cabral, M.G. HLA-DR in Cytotoxic T Lymphocytes Predicts Breast Cancer Patients' Response to Neoadjuvant Chemotherapy. *Front. Immunol.,* **2018**, *9*, 2605. [http://dx.doi.org/10.3389/fimmu.2018.02605] [PMID: 30555458]

[22] Adhikari, U.K.; Tayebi, M.; Rahman, M.M. Immunoinformatics Approach for Epitope-Based Peptide Vaccine Design and Active Site Prediction against Polyprotein of Emerging Oropouche Virus. *J. Immunol. Res.,* **2018**, *2018*, 6718083. [http://dx.doi.org/10.1155/2018/6718083] [PMID: 30402510]

[23] Gupta, S.; Kapoor, P.; Chaudhary, K.; Gautam, A.; Kumar, R.; Raghava, G.P. Open Source Drug Discovery Consortium. In silico approach for predicting toxicity of peptides and proteins. *PLoS One,* **2013**, *8*(9), e73957. [http://dx.doi.org/10.1371/journal.pone.0073957] [PMID: 24058508]

[24] Singh, S.; Singh, H.; Tuknait, A.; Chaudhary, K.; Singh, B.; Kumaran, S.; Raghava, G.P. PEPstrMOD: structure prediction of peptides containing natural, non-natural and modified residues. *Biol. Direct,* **2015**, *10*, 73. [http://dx.doi.org/10.1186/s13062-015-0103-4] [PMID: 26690490]

[25] Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.,* **2018**, *46*(W1), W296-W303. [http://dx.doi.org/10.1093/nar/gky427] [PMID: 29788355]

[26] Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H.J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.,* **2005**, *33*(Web Server issue), W363-7. [http://dx.doi.org/10.1093/nar/gki481] [PMID: 15980490]

[27] Guan, P.; Doytchinova, I.A.; Zygouri, C.; Flower, D.R. MHCPred: A server for quantitative prediction of peptide-MHC binding. *Nucleic Acids Res.,* **2003**, *31*(13), 3621-3624. [http://dx.doi.org/10.1093/nar/gkg510] [PMID: 12824380]

[28] Bui, H.H.; Sidney, J.; Dinh, K.; Southwood, S.; Newman, M.J.; Sette, A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics,* **2006**, *7*, 153. [http://dx.doi.org/10.1186/1471-2105-7-153] [PMID: 16545123]

[29] Hospital, A.; Andrio, P.; Fenollosa, C.; Cicin-Sain, D.; Orozco, M.; Gelpí, J.L. MDWeb and MDMoby: an integrated web-based platform for molecular dynamics simulations. *Bioinformatics,* **2012**, *28*(9), 1278-1279. [http://dx.doi.org/10.1093/bioinformatics/bts139] [PMID: 22437851]

[30] Sievers, F; Higgins, DG Clustal Omega *Curr. Protoc. Bioinforma,* **2014**. [http://dx.doi.org/10.1002/0471250953.bi0313s48]

[31] Schoeman, D.; Fielding, B.C. Coronavirus envelope protein: current knowledge. *Virol. J.,* **2019**, *16*(1), 69. [http://dx.doi.org/10.1186/s12985-019-1182-0] [PMID: 31133031]

[32] Guo, F.; Li, S.C.; Wang, L.; Zhu, D. Protein-protein binding site identification by enumerating the configurations. *BMC Bioinformatics,* **2012**, *13*, 158.

[http://dx.doi.org/10.1186/1471-2105-13-158] [PMID: 22768846]

[33] Folegatti, P.M.; Ewer, K.J.; Aley, P.K.; Angus, B.; Becker, S.; Belij-Rammerstorfer, S.; Bellamy, D.; Bibi, S.; Bittaye, M.; Clutterbuck, E.A.; Dold, C.; Faust, S.N.; Finn, A.; Flaxman, A.L.; Hallis, B.; Heath, P.; Jenkin, D.; Lazarus, R.; Makinson, R.; Minassian, A.M.; Pollock, K.M.; Ramasamy, M.; Robinson, H.; Snape, M.; Tarrant, R.; Voysey, M.; Green, C.; Douglas, A.D.; Hill, A.V.S.; Lambe, T.; Gilbert, S.C.; Pollard, A.J. Oxford COVID Vaccine Trial Group. Safety and immunogenicity of the ChAdOx1 nCoV--19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *Lancet,* **2020**, S0140-6736(20)31604-4.
[PMID: 32702298]

[34] Tahir Ul Qamar, M.; Saleem, S.; Ashfaq, U.A.; Bari, A.; Anwar, F.; Alqahtani, S. Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: an immune-informatics study. *J. Transl. Med.,* **2019**, *17*(1), 362.
[http://dx.doi.org/10.1186/s12967-019-2116-8] [PMID: 31703698]

[35] Ben-Yedidia, T.; Arnon, R. Epitope-based vaccine against influenza. *Expert Rev. Vaccines,* **2007**, *6*(6), 939-948.
[http://dx.doi.org/10.1586/14760584.6.6.939] [PMID: 18034655]

[36] Alberts, B.; Johnson, A.; Lewis, J. *Helper T Cells and Lymphocyte Activation,* 4th ed; , **2002**.

[37] G SK. Joshi A, Kaushik V. T cell epitope designing for dengue peptide vaccine using docking and molecular simulation studies. *Mol. Simul.,* **2020**, *46*(10), 787-795.
[http://dx.doi.org/10.1080/08927022.2020.1772970]