ISSN: 2588-4115; Infection Epidemiology and Microbiology. 2022;8(3):259-276 10.52547/iem.8.3.259

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

ARTICLE INFO

Article Type
Original Article

Authors
Anamika Basu, *PhD*[1*]

How to cite this article
Basu B. Strategies for Vaccine Design for Coronavirus All Variants of Concern Using Immunoinformatics Techniques. Infection Epidemiology and Microbiology. 2022;8(3): 259-276

[1] Lecturer, Girudas College, Calcutta University.

** Correspondence*
Lecturer,Girudas College, Calcutta University.
Email:anamika.biochem@gurudas.education

Article History
Received: November 10 ,2021
Accepted: July 19 ,2022
Published: September 19 ,2022

ABSTRACT

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 ^{74}QRGGSYTNDKA^{84} 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 ^{359} FLAHIQWMV^{367} and ^{359}FLAHIQWVMFTPLV^{373} 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 IC_{50} 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.

CITATION LINKS

[1] Cui, J., Li, F., & Shi, Z. L. (2019). Origin ... [2] Sohrabi, C., Alsafi, Z., O'Neill, N., ... [3] Jiang, S., Shi, Z., Shu, Y., Song, J., ... [4] Liu, N. N., Tan, J. C., Li, J., Li, S., ... [5] Lai, C. C., Shih, T. P., Ko, W. C., ... [6] Guo, Y. R., Cao, Q. D., Hong, Z. S., ... [7] Guan, Y., Zheng, B. J., He, Y. Q., Liu, X. L., ... [8] Alagaili, A. N., Briese, T., Mishra, N., ... [9] De Wit, E., Van Doremalen, N., ... [10] Zhou, P., Yang, X. L., Wang, X. ... [11] Chen, Y., Liu, Q., & Guo, D. ... [12] Chan, J. F. W., Yuan, S., Kok, K. H., To, K. ... [13] Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., ... [14] Young, B. E., Ong, S. W. X., ... [15] Rothe, C., Schunk, M., ... [16] Gonzalez-Reiche, A. S., Hernandez, M. ... [17] da Silva Candido, D. ... [18] Huang, C., Wang, Y., Li ... [19] Zumla, A., Chan, J. F., Azhar, E. I., Hui, D. S., ... [20] Lu, R., Zhao, X., Li, J., Niu, P., Yang, ... [21] Tan, W., Zhao, X., Ma, X., Wang, W., ... [22] Xu, J., Zhao, S. ... [23] Chan, J. F. W., Kok, K. H., Zhu, Z., Chu, H., ... [24] Davies, J. P., Almasy ... [25] Bianchi, M., Benvenut ... [26] Doyle, N., Neuman, B. W. ... [27] Yin, C. (2020). Genotyping coronavirus...[28] Mercatelli, D., & Giorgi, F. M... [29] da Silva, S. J. R., da Silva, C. T. A., Mendes ... [30] Lon, J. R., Bai, Y. ... [31] He, J., Huang, F., Zhang, J. ... [32] Chen, Z., Ruan, P., Wang, L ... [33] Dagur, H. S., Dhakar, S. S., & Gupta.. [34] Iva, Y. A. (2021). In-silico based vaccine development ... [35] Doytchinova, I.A. and ... [36] Fieser TM ... [37] Kolaskar AS, Tongaonkar PC. A semi-empirical method ... [38] Emini EA, Hughes JV, Perlow DS, et al. Induction of... [39] Parker JM, Guo D, Hodges RS. New... [40] Karplus PA, Schulz GE. Prediction of ... [41] Larsen, J.E., Lund, O. ... [42] Chou PY... [43] Jespersen, M. C., Peters, B., Nielsen ... [44] Nielsen M, Lundegaard C, Worning ... [45] Peters B, Sette A. 2005. Generating ... [46] Vita R, Overton JA ... [47] Tenzer S, Peters B, Bulik S, ... [48] Larsen, M.V., Lundegaard... [49] Bui HH, Sidney J, Dinh K, Southwood ... [50] Thévenet, P., Shen ... [51] Shen, Y., Maupetit, J.... [52] Kozakov, D., Hall, D.R., Xia, B., Porter, K.A., Padhorny, D., Yueh, C., Beglov, D. and ... [53] Kozakov D, Brenke R, Comeau SR, Vajda S. PIPER: an ... [54] Hunter, S., Apweiler, R., Attwood, T. K., Bairoch ... [55] Mulder, N. J., Apweiler, R. ... [56] Healy, M. D. ... [57] Desai, D. V., & Kulkarni-Kale, U. (2014). T-cell epitope... [58] Lafuente, E. M., & Reche, P. A. (2009). Prediction of ... [59] Lundegaard, C., Lund, O., Buus, ... [60] Karosiene, E., Lundegaard ... [61] Nielsen, M., Lundegaard ... [62] Sturniolo, T., Bono, E., Ding, J., Raddrizzani, L. ... [63] Kar, T., Narsaria, U., Basak, S. De, D., Castiglion, F. ... [64] Crooke, S. N., Ovsyannikova, I. G., Kennedy... [65] Bhattacharya, M., Sharma, A. R., Patra, P., Ghosh ... [66] Wang, D., Mai, J., Zhou, W., Yu ... [67] Tosta, S. F. D. O. ... [68] Jabbar, B., Rafique, S., ... [69] Ojha, R., Pareek, A., Pandey, R. ... [70] Forster, P., Forster, L. ... [71] Adebali, O., Bircan, A., Circi, D.... [72] Devendran, R., Kumar, M. ... [73] Bui, H. H., Sidney, J., Peters, B. ... [74] Tenzer, S., Peters, B., Bulik, S. ... [75] Peters, B., Bulik, S., Tampe, R. ... [76] Alter, I., Gragert, L., Fingerson, S. ... [77] Peele, K. A., Srihansa... [78] Samad, A., Ahammad, F., Nain, Z ... .

[DOI: 10.52547/iem8.3.259]

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]. Phylodynamic 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

[DOI: 10.52547/iem8.3.259]

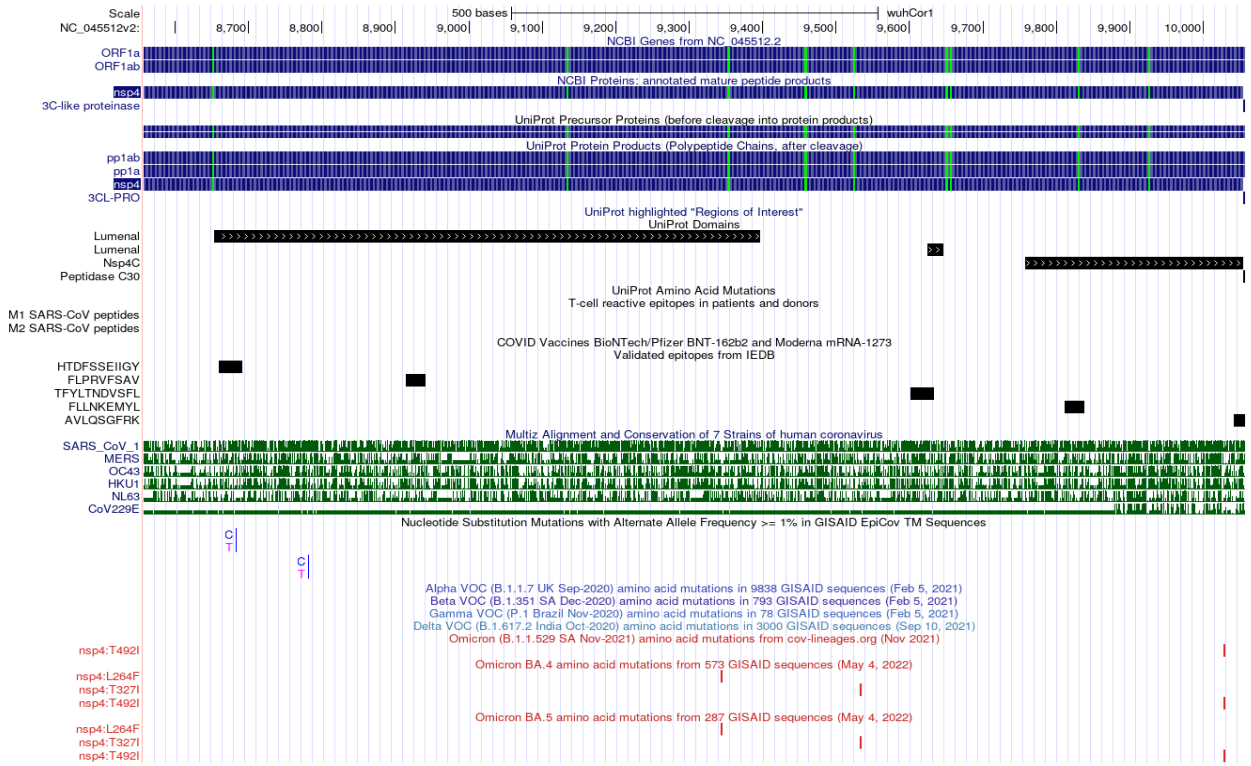


Figure 2) Positions of point mutations in nsp4 of different strains of SARS-Cov-2

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

[DOI: 10.52547/iem8.3.259]

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[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 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

[DOI: 10.52547/iem8.3.259]

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.

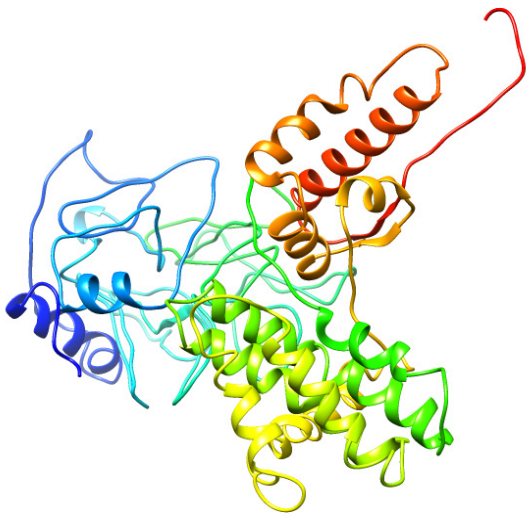
Study of predicted epitopes for immune reactivity based on experimental results: The Immune Epitope Database (IEDB) is a freely available database. This database indexes experimental data on antibody and T cell epitopes studied in humans, non-human primates, and other animal species in relation to allergy,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

Results

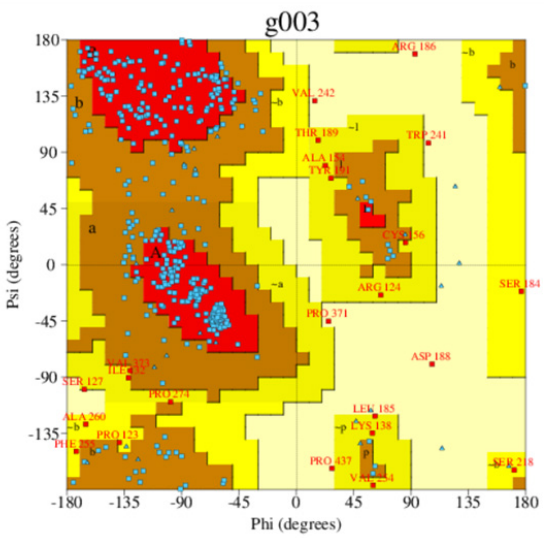
Sequence retrieval and prediction of secondary structure of nonstructural protein 4 (nsp4): Due to the lack of experimental X-ray crystallographic structure of nsp4 viral protein, its primary structure was first retrieved.

```
>YP_009742611.1 nsp4 [Severe acute respiratory syndrome coronavirus 2]
KIVNNWLKQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIASTDTCFA
NKHADFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTTNGDFLHFLPRVFS
AVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPVPYCYDTNVLEGSVAYESLRPDTRY
VLMDGSIIQFPNTYLEGSVRVVTTFDSEYCRHGTCERSEAGVCVSTSGRWVLNNDYYRSLP
GVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCLAYYFMRFRRAFGEYSHVVA
FNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFLAHIQWMVMFTPLVPFWITIAY
IICISTKHFYWFFSNYLKRRVVFNGVSFSTFEEAALCTFLLNKEMYLKLRSDVLLPLTQYNRY
LALYNKYKYFSGAMDTTSYREAACCHLAKALNDFSNSGSDVLYQPPQTSITSAVLQ
```

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



The quality of the predicted structure was checked by creating a Ramachandran plot.



[DOI: 10.52547/iem8.3.259]

PROCHECK statistics

1. Ramachandran Plot statistics

		No. of residues	%-tage
Most favoured regions	[A,B,L]	332	73.6%**
Additional allowed regions	[a,b,l,p]	100	22.2%
Generously allowed regions	[~a,~b,~l,~p]	15	3.3%
Disallowed regions	[XX]	4	0.9%*
Non-glycine and non-proline residues		451	100.0%
End-residues (excl. Gly and Pro)		2	
Glycine residues		27	
Proline residues		20	
Total number of residues		500	

Based on an analysis of **118** structures of resolution of at least **2.0** Angstroms and *R*-factor no greater than **20.0** a good quality model would be expected to have over **90%** in the most favoured regions [A,B,L].

The antigenicity of the predicted protein was calculated using Vaxign-ML software[35]. Antigenicity prediction score of nsp4 using Vaxign-ML software was 89.6 [36].

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

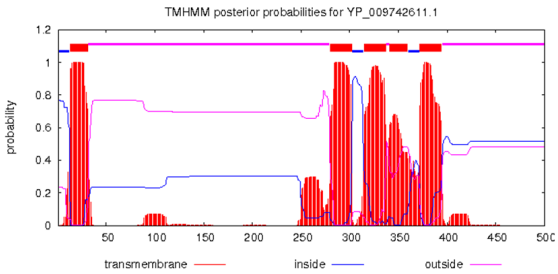


Figure 3) Predicted signal peptide and transmembrane regions of nsp4

Prediction of potential B cell epitopes

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.

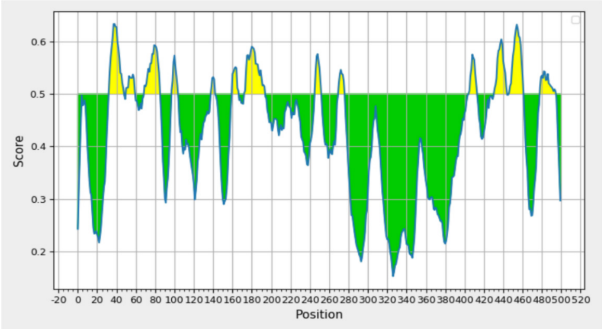


Figure 4) Result of BepiPred-2.0 prediction

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.

Predicted peptides:

No.	Start	End	Peptide	Length
1	33	48	MSKHTDFSSEIIGYKA	16
2	51	60	GGVTRDIAST	10
3	70	86	DFDTWFSQRGGSYTNDK	17
4	98	104	EVGFVVP	7
5	139	143	LIEYT	5
6	161	167	DASGKPV	7
7	173	194	TNVLEGSVAYESLRPDTRYVLM	22
8	246	252	DYYRSLP	7
9	271	276	LIQPIG	6
10	405	413	NGVSFSTFE	9
11	432	444	SDVLLPLTQYNRY	13
12	447	461	LYNKYKYFSGAMDTT	15
13	479	495	SNSGSDVLYQPPQTSIT	17

Figure 5) Predicted epitopes [43]

b) Results of Chou & Fasman beta-turn prediction: The Chou and Fasman (1978) scale

[DOI: 10.52547/iem8.3.259]

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.

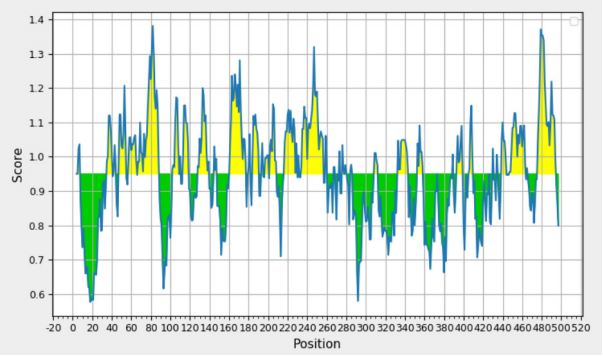


Figure 6) Results of beta-turn prediction

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:

Sn = (n+4+i) (0.37)-6

Where Sn 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 Sn values larger than 1.0.

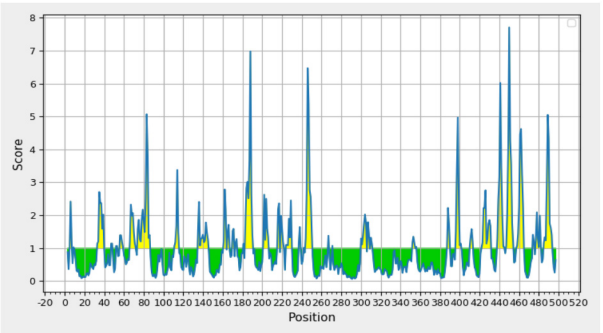


Figure 7) Results of surface accessibility prediction

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.

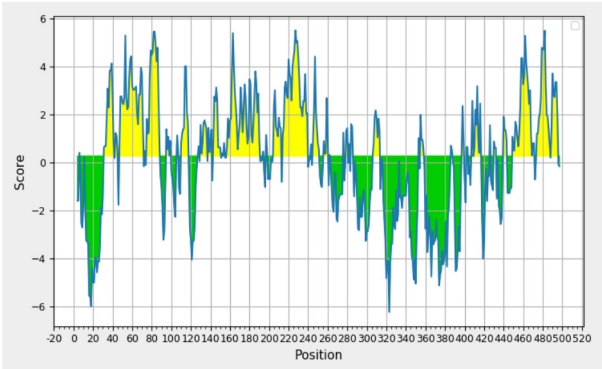


Figure 8) Results of Parker hydrophilicity prediction

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 a-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.

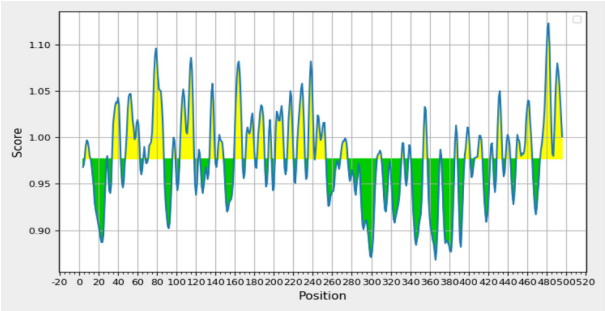


Figure 9. Results of Karplus & Schulz flexibility prediction

[DOI: 10.52547/iem8.3.259]

Considering the above four parameters for predicting B cell epitopes (Table 1), the short peptide sequence ^{74}QRGGSYTNDKA^{84} 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 PEProteasome P-FOLD software peptides [41].

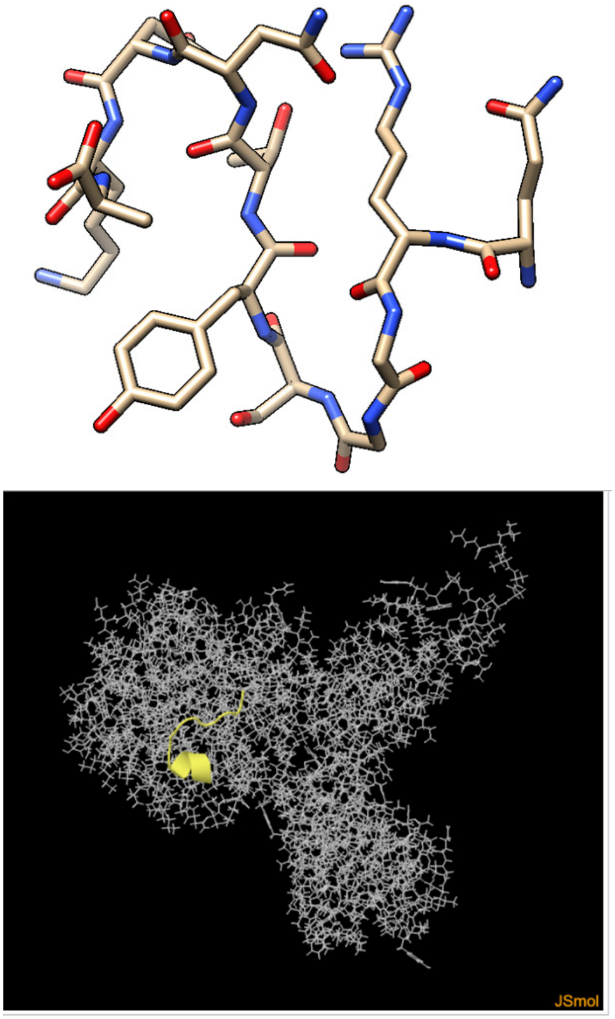


Figure 10) Three-dimensional structure of the predicted B cell epitope and its location in the three-dimensional structure of nsp4 protein

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.

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 IC_{50} 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 IC_{50} values. The projected output in units of IC_{50} nM is provided for a number of epitopes. Therefore, a lower IC_{50} value denotes a higher affinity of that specific interacting MHC-I allele. Peptides with IC_{50} values of 50, 500, and 5000 nM are regarded to have high, intermediate, and poor affinity, respectively.

[DOI: 10.52547/iem8.3.259]

Table 1) The physicochemical parameters for amino acids presented in the predicted B cell epitope

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

Table 2) MHC I binding prediction score

Epitope sequence no	Peptide start	Peptide end	Peptide	IC50	Allele
	359	367	FLAHIQWMV	2.01	HLA-A*02:01
1	359	367	FLAHIQWMV	2.44	HLA-A*02:03
1	359	367	FLAHIQWMV	2.78	HLA-A*02:06
1	359	367	FLAHIQWMV	16.29	HLA-A*68:02
2	121	129	FLPRVFSAV	2.47	HLA-A*02:03
2	121	129	FLPRVFSAV	4.04	HLA-A*02:06
2	121	129	FLPRVFSAV	5.52	HLA-A*02:01

[DOI: 10.52547/iem8.3.259]

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 FLAHIQWMVMFTPLV 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 ^{359}FLAHIQWMV^{369} (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.

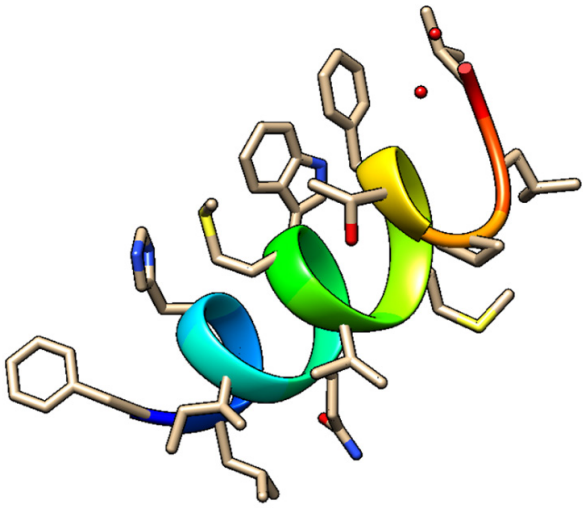


Figure 11) Three-dimensional structure of the predicted T cell epitope

Molecular docking study of the predicted T cell epitopes with ternary complex: Molecular docking study showed that class II MHCmolecule(HLA-DRB4*01:01)-restricted T-cell epitope FLAHIQWMVMFTPLV 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

[DOI: 10.52547/iem8.3.259]

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

Allele	Start	End	Peptide Length	Peptide	Proteasome Score	TAP Score	MHC Score	Processing Score	Total Score	MHC IC50 [nM]
HLA-A*02:01	9	17	9	FLAHIQWMV	1.18	0.13	-0.28	1.31	1.04	1.9
HLA-A*02:06	9	17	9	FLAHIQWMV	1.18	0.13	-0.38	1.31	0.93	2.4
HLA-A*68:02	9	17	9	FLAHIQWMV	1.18	0.13	-0.93	1.31	0.38	8.6

Table 4) MHC II binding score

Allele	Start	End	Length	Method used	Peptide	Percentile Rank	Adjusted rank
HLA-DRB4*01:01	359	373	15	Consensus (comb. lib./smm/nn)	FLAHIQWMVMFTPLV	2.10	2.10

Table 5) Immunogenicity prediction score for T cell epitopes

Class I Immunogenicity

Masking: **default**
Masked variables: **[1, 2, 'cterm']**

Peptide	Length	Score
FYLTNDVSFLAHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFEEAALCTF	70	1.5946
SASIVAGGIVAIVVTCLAYYFMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLT	70	1.18638
FDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTNGDFLHFLPRVFSAVGNICYTPSKLI	70	1.15968
RVVTTFDSEYCRHGTCERSEAGVCVSTSGRWVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDI	70	0.78289
KIVNNWLKQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIASTDTCFANKHAD	70	0.69645
EYTDFATSACVLAAECTIFKDASGKPVPYCYDTNVLEGSVAYESLRPDTRYVLMDGSIIQFPNTYLEGSV	70	0.19461
QTSITSAVLQ	10	0.02211
LLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHLAKALNDFSNSGSDVLYQPP	70	-1.68192

binding energy of this ternary complex was calculated. The binding energy of this ternary complex was -986 Kcal/mole.

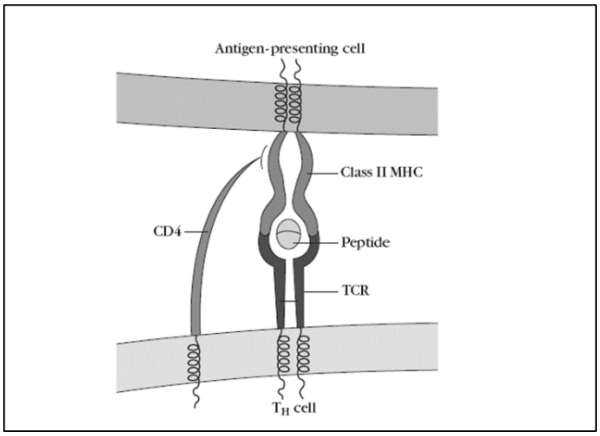
In Figure 12, the T-cell epitope is shown as a red helix along with the HLA-DRB4*01:01 molecule and TCR in green and yellow, respectively.

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 6) Population coverage of Class I MHC epitope

population/area	Class I MHC restricted epitope		
	coverage	average_hit	pc90
China	29.66%	0.31	0.14
England	50.35%	0.51	0.2
India	15.9%	0.16	0.12
North America	47.79%	0.49	0.19
World	43.26%	0.45	0.18
Average	37.39	0.38	0.17
Standard deviation	12.9	0.13	0.03

[DOI: 10.52547/iem8.3.259]



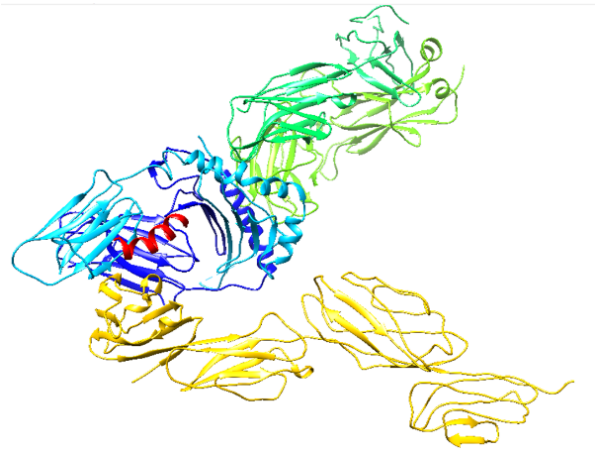


Figure 12) T cell epitope with ternary complex

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 [49] 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

population/area	Class II MHC restricted T cell epitope		
	coverage	average_hit	pc90
China	15.54%	0.16	0.12
England	41.62%	0.44	0.17
India	24.56%	0.25	0.13
North America	34.63%	0.36	0.15
World	28.79%	0.3	0.14
Average	29.03	0.3	0.14
Standard deviation	8.85	0.1	0.02

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.

[DOI: 10.52547/iem8.3.259]

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 immunoinformatics 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 ^{359}FLAHIQWMV^{369} 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

Epitope #	Epitope name	Epitope sequence	Epitope length	Percent of protein sequence matches at identity <= 100%	Minimum identity	Maximum identity
1	ws-separated-0	FLAHIQWMV	9	12.50% (1/8)	0.00%	100.00%

[DOI: 10.52547/iem8.3.259]

(^{359}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}QRGGSYTNDKA^{84} (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, ^{359}FLAHIQWMV^{369}, was bound 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}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 ^{359}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 ^{359}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

[DOI: 10.52547/iem8.3.259]

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, ^{47}QRGGSYTNDKA^{84}, 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

1. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol. 2019;17(3):181-92.
2. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). Int J Surg. 2020;76:71-6.
3. Jiang S, Shi Z, Shu Y, Song J, Gao GF, Tan W, et al. A distinct name is needed for the new coronavirus. Lancet. 2020;395(10228):948-9.
4. Liu NN, Tan JC, Li J, Li S, Cai Y, Wang H. COVID-19 pandemic: Experiences in China and implications for its prevention and treatment worldwide. Curr Cancer Drug Targets. 2020;20(6):410-6.
5. Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): The epidemic and the challenges. Int J Antimicrob Agents. 2020;55(3):105924.
6. Guo YR, Cao QD, Hong ZS, Tan YY, Chen SD, Jin HJ, et al. The origin, transmission, and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak–an update on the status. Mil Med Res. 2020;7(1):1-10.
7. Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science. 2003;302(5643):276-8.
8. Alagaili AN, Briese T, Mishra N, Kapoor V, Sameroff SC, De Wit E, et al. Middle East respiratory syndrome coronavirus infection in dromedary camels in Saudi Arabia. MBio. 2014;5(2):e00884-14.
9. De Wit E, Van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: Recent insights into emerging coronaviruses. Nat Rev Microbiol. 2016;14(8):523-34.
10. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579(7798):270-3.
11. Chen Y, Liu Q, Guo D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. J Med Virol. 2020;92(4):418-23.
12. Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. Lancet. 2020;395(10223):514-23.
13. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. Lancet. 2020;395(10223):507-13.
14. Young BE, Ong SW, Kalimuddin S, Low JG, Tan SY, Loh J, et al. Epidemiologic features and clinical course of patients infected with SARS-CoV-2 in Singapore. Jama. 2020;323(15):1488-94.
15. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, et al. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. New Eng J Med. 2020;382(10):970-1.
16. Gonzalez-Reiche AS, Hernandez MM, Sullivan MJ, Ciferri B, Alshammary H, Obla A, et al. Introductions and early spread of SARS-CoV-2 in the New York City area. Science. 2020;369(6501):297-301.
17. Da Silva Candido, D., Claro, I. M., de Jesus, J. G., de Souza, W. M., Moreira, F. R. R., Dellicour, S., ... & Manuli, E. R. (2020). Evolution and epidemic spread of SARS-CoV-2 in Brazil. medRxiv.
18. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395(10223):497-506.
19. Zumla A, Chan JF, Azhar EI, Hui DS, Yuen KY.

[DOI: 10.52547/iem8.3.259]

Coronaviruses—drug discovery and therapeutic options. Nat Rev Drug Discov. 2016;15(5):327-47.
20. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterization and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. Lancet. 2020;395(10224):565-74.
21. Tan W, Zhao X, Ma X, Wang W, Niu P, Xu W, et al. A novel coronavirus genome identified in a cluster of pneumonia cases—Wuhan, China 2019– 2020. China CDC Weekly. 2020;2(4):61-2.
22. Xu J, Zhao S, Teng T, Abdalla AE, Zhu W, Xie L, et al. Systematic comparison of two animal-to-human transmitted human coronaviruses: SARS-CoV-2 and SARS-CoV. Viruses. 2020;12(2):244.
23. Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect. 2020;9(1):221-36.
24. Davies JP, Almasy KM, McDonald EF, Plate L. Comparative multiplexed interactomics of SARS-CoV-2 and homologous coronavirus non-structural proteins identifies unique and shared host-cell dependencies. bioRxiv. 2020.
25. Bianchi M, Benvenuto D, Giovanetti M, Angeletti S, Ciccozzi M, Pascarella S. Sars-CoV-2 envelope and membrane proteins: Structural differences linked to virus characteristics? BioMed Res Int. 2020;2020.
26. Doyle N, Neuman BW, Simpson J, Hawes PC, Mantell J, Verkade P, et al. Infectious bronchitis virus nonstructural protein 4 alone induces membrane pairing. Viruses. 2018;10(9):477.
27. Yin C. Genotyping coronavirus SARS-CoV-2: Methods and implications. Genomics. 2020;112(5):3588-96.
28. Mercatelli D, Giorgi FM. Geographic and genomic distribution of SARS-CoV-2 mutations. Front Microbiol. 2020;11:1800.
29. Da Silva, S. J. R., da Silva, C. T. A., Mendes, R. P. G., & Pena, L. (2020). Role of Nonstructural Proteins in the Pathogenesis of SARS-CoV-2. Journal of Medical Virology.
30. Lon, J. R., Bai, Y., Zhong, B., Cai, F., & Du, H. (2020). Prediction and evolution of B cell epitopes of surface protein in SARS-CoV-2. Virology journal, 17(1), 1-9.
31. He, J., Huang, F., Zhang, J., Chen, Q., Zheng, Z., Zhou, Q., ... & Chen, J. (2021). Vaccine design based on 16 epitopes of SARS-CoV-2 spike protein. Journal of medical virology, 93(4), 2115-2131.
32. Chen, Z., Ruan, P., Wang, L., Nie, X., Ma, X., & Tan, Y. (2021). T and B cell Epitope analysis of SARS-CoV-2 S protein based on immunoinformatics and experimental research. Journal of cellular and molecular medicine, 25(2), 1274-1289.
33. Dagur, H. S., Dhakar, S. S., & Gupta, A. (2020). Epitope-based vaccine design against novel coronavirus SARS-CoV-2 envelope protein. EJMO, 4(3), 201-208.
34. Iva, Y. A. (2021). In-silico based vaccine development: targeting nucleocapsid protein of SARS-CoV-2 (Doctoral dissertation, Brac University).
35. Doytchinova IA, Flower DR. VaxiJen: A server for prediction of protective antigens, tumor antigens, and subunit vaccines. BMC Bioinform. 2007;8(1):1-7.
36. Fieser TM, Tainer JA, Geysen HM, Houghten RA, Lerner RA. Influence of protein flexibility and peptide conformation on reactivity of monoclonal anti-peptide antibodies with a protein a-helix. Proc Natl Acad Sci. 1987;84(23):8568–72.
37. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett. 1990;276(1–2):172–4.
38. Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J Virol. 1985;55(3):836–9.
39. Parker JM, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: Correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry. 1986;25(19):5425–32.
40. Karplus PA, Schulz GE. Prediction of chain flexibility in proteins. Naturwissenschaften. 1985; 72(4):212–3.
41. Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immunome Res. 2006;2(1):1-7.
42. Chou PY, Fasman GD. Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol Relat Areas Mol Biol. 1978;47:45–148.
43. 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(W1):W24-9.
44. Nielsen M, Lundegaard C, Worning P, Lauemoller SL, Lamberth K, Buus S, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci. 2003;12(5):1007-17.
45. Peters B, Sette A. Generating quantitative

[DOI: 10.52547/iem8.3.259]

models describing the sequence specificity of biological processes with the stabilized matrix method. BMC Bioinform. 2005;6(1):1-9.
46. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, et al. The immune epitope database (IEDB) 3.0. Nucleic Acids Res. 2014;43(D1):D405-12.
47. Tenzer S, Peters B, Bulik S, Schoor O, Lemmel C, Schatz MM, et al. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport, and MHC class I binding. Cell Mol Life Sci. 2005;62(9):1025-37.
48. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinform. 2007;8(1):1-2.
49. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinform. 2006;7(1):1-5.
50. Thévenet P, Shen Y, Maupetit J, Guyon F, Derreumaux P, Tufféry P. PEP-FOLD: An updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. Nucleic Acids Res. 2012;40(W1):W288-93.
51. Shen Y, Maupetit J, Derreumaux P, Tufféry P. Improved PEP-FOLD approach for peptide and miniprotein structure prediction. J Chem Theory Comput. 2014;10(10):4745-58.
52. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusPro web server for protein-protein docking. Nat Protoc. 2017;12(2):255-78.
53. Kozakov D, Brenke R, Comeau SR, Vajda S. PIPER: An FFT-based protein docking program with pairwise potentials. Proteins: Struct Funct Bioinform. 2006;65(2):392-406.
54. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. InterPro: The integrative protein signature database. Nucleic Acids Res. 2009;37(Suppl-1):D211-5.
55. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. New developments in the InterPro database. Nucleic Acids Res. 2007;35(Suppl-1):D224-8.
56. Healy MD. Using BLAST for performing sequence alignment. Curr Protoc Hum Genet. 2007;52(1):6-8.
57. Desai DV, Kulkarni-Kale U. T-cell epitope prediction methods: An overview. In: De R,Tomar N, editors. Immunoinformatics: Methods in molecular biology. New York, NY: Humana Press; 2014, vol. 1184, p. 333-364.
58. Lafuente EM, Reche PA. Prediction of MHC-peptide binding: A systematic and comprehensive overview. Cur Pharm Des. 2009;15(28):3209-20.
59. Lundegaard C, Lund O, Buus S, Nielsen M. Major histocompatibility complex class I binding predictions as a tool in epitope discovery. Immunology. 2010;130(3):309-18.
60. Karosiene E, Lundegaard C, Lund O, Nielsen M. NetMHCcons: A consensus method for the major histocompatibility complex class I predictions. Immunogenetics, 2012;64(3):177-86.
61. Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. BMC Bioinform. 2007;8(1):1-2.
62. Sturniolo T, Bono E, Ding J, Raddrizzani L, Tuereci O, Sahin U, et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nat Biotechnol. 1999;17(6):555-61.
63. Kar T, Narsaria U, Basak S, Deb D, Castiglion F, Mueller DM, et al. A candidate multi-epitope vaccine against SARS-CoV-2. Sci Rep. 2020;10(1):1-24.
64. 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(1):1-5.
65. Bhattacharya M, Sharma AR, Patra P, Ghosh P, Sharma G, Patra BC, et al. Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach. J Med Virol. 2020;92(6):618-31.
66. Wang D, Mai J, Zhou W, Yu W, Zhan Y, Wang N, et al. Immunoinformatic analysis of T-and B-cell epitopes for SARS-CoV-2 vaccine design. Vaccines. 2020;8(3):355.
67. Tosta SF, Passos MS, Kato R, Salgado Á, Xavier J, Jaiswal AK, et al. Multi-epitope based vaccine against yellow fever virus applying immunoinformatics approaches. J Biomol Struct Dyn. 2021;39(1):219-35.
68. Jabbar B, Rafique S, Salo-Ahen OM, Ali A, Munir M, Idrees M, et al. Antigenic peptide prediction from E6 and E7 oncoproteins of HPV Types 16 and 18 for therapeutic vaccine design using immunoinformatics and MD simulation analysis. Front Immunol. 2018;9:3000.
69. Ojha R, Pareek A, Pandey RK, Prusty D, Prajapati VK. Strategic development of a next-generation multi-epitope vaccine to prevent Nipah virus zoonotic infection. ACS Omega. 2019;4(8):13069-79.
70. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. Proc Natl Acad Sci. 2020;117(17):9241-3.
71. Adebali O, Bircan A, Circi D, Islek B, Kilinc Z,

[DOI: 10.52547/iem8.3.259]

Selcuk B, et al. Phylogenetic analysis of SARS-CoV-2 genomes in Turkey. bioRxiv. 2020.
72. Devendran R, Kumar M, Chakraborty S. Genome analysis of SARS-CoV-2 isolates occurring in India: Present scenario. Indian J Public Health. 2020;64(6):147-55.
73. Bui HH, Sidney J, Peters B, Sathiamurthy M, Sinichi A, Purton KA, et al. Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications. Immunogenetics. 2005;57(5):304-14.
74. Peters B, Bulik S, Tampe R, Van Endert PM, Holzhütter HG. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. J Immunol. 2003;171(4):1741-9.
75. Alter I, Gragert L, Fingerson S, Maiers M, Louzoun Y. HLA class I haplotype diversity is consistent with selection for frequent existing haplotypes. PLoS Comput Biol. 2017;13(8):e1005693.
76. Peele KA, Srihansa T, Krupanidhi S, Sai AV, Venkateswarulu TC. Design of multi-epitope vaccine candidate against SARS-CoV-2: An in-silico study. J Biomol Struct Dyn. 2021;39(10):3793-801.
77. Samad A, Ahammad F, Nain Z, Alam R, Imon RR, Hasan M, et al. Designing a multi-epitope vaccine against SARS-CoV-2: An immunoinformatics approach. J Biomol Struct Dyn. 2022;40(1):14-30.

[DOI: 10.52547/iem8.3.259]