



Research article

An immunoinformatics study on the spike protein of SARS-CoV-2 revealing potential epitopes as vaccine candidates



Arafat Islam Ashik, Mahedi Hasan, Atiya Tahira Tasnim, Md. Belal Chowdhury, Tanvir Hossain, Shamim Ahmed*

Department of Biochemistry and Molecular Biology, School of Life Sciences, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh

ARTICLE INFO

Keywords:

Biochemistry
Bioinformatics
Immunology
Virology
COVID-19
Spike glycoprotein
B-cell and T-cell epitope
Epitope-based vaccine design
SARS-CoV-2

ABSTRACT

Background: The pandemic situation of SARS-CoV-2 infection has sparked global concern due to the disease COVID-19 caused by it. Since the first cluster of confirmed cases in China in December 2019, the infection has been reported across the continents and inflicted upon a substantial number of populations.

Method: This study is focused on immunoinformatics analyses of the SARS-CoV-2 spike glycoprotein (S protein) which is key for the viral attachment to human host cells. Computational analyses were carried out for the prediction of B-cell and T-cell (MHC class I and II) epitopes of S protein and the analyses were extended further for the prediction of their immunogenic properties. The interaction and binding affinity of T-cell epitopes with HLA-B7 were also investigated by molecular docking.

Result: Three distinct epitopes for vaccine design were predicted from the sequence of S protein. The potential B-cell epitope was KNHTSPDVLG possessing the highest antigenicity score of 1.4039 among other B-cell epitopes. T-cell epitope for human MHC class I was VVLSFELL with an antigenicity score of 1.0909 and binding ability to 29 MHC-I alleles. The predicted T-cell epitope for human MHC class II molecule was VVIGVNNNT with a corresponding 1.3063 antigenicity score, less digesting enzymes, and 7 MHC-II alleles binding ability. All these three peptides were predicted to be highly antigenic, non-allergenic, and non-toxic. Analyses of the physiochemical properties of these predicted epitopes indicate their stable nature for plausible vaccine design. Furthermore, molecular docking investigation between the MHC class-I epitopes and human HLA-B7 reflects the stable interaction with high affinity among them.

Conclusion: The present study posits three potential epitopes of S protein of SARS-CoV-2 predicted by immunoinformatic methods based on their immunogenic properties and interactions with the host counterpart that can facilitate the development of vaccine against SARS-CoV-2. This study can act as the springboard for the future development of the COVID-19 vaccine.

1. Introduction

COVID 19, an ongoing pandemic, is the viral disease attributed to the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) [1]. The SARS-CoV-2 is a coronavirus strain under one of the (+) ssRNA virus family Coronaviridae. SARS-CoV-2 uses human ACE2 (angiotensin converting enzyme 2) protein as its receptor [2]. The origin of SARS-CoV-2 is still an unsettled issue since the link of virus's speculated origin to Wuhan, China is yet to be established. Studies supporting and arguing the connection to a particular region of China have positioned the origin of the virus up in the air [3, 4, 5]. Although SARS-CoV-2 was initially speculated to be transmitted to humans from Huanan Seafood Wholesale

Market's seafood in Wuhan, China, person-to-person transmission is now evident [4, 5, 6]. Most patients are being given systematic antiviral therapy which is only effective to alleviate the symptoms [3]. While the antiviral drugs are only effective against virus-infected individuals, vaccination provides a protective shield to the unaffected populations against the possible viral attack and gradually drives the eradication of the virus as a consequence. That is why prevention by immunization is presumably the most pragmatic solution to tackle the virus, and thus huge endeavors are being channeled on vaccine development.

The basic architectural features of SARS-CoV-2 comprise its 30kb long RNA genome and four major types of structural protein capsid. Among the whole genome of coronavirus, the 3-terminus ORFs encode the

* Corresponding author.

E-mail addresses: shamim1174-bmb@sust.edu, shamim1174@gmail.com (S. Ahmed).

structural proteins-spike proteins (S), envelope proteins (E), membrane proteins (M), and nucleocapsid proteins (N) [7]. All these structural and nonstructural accessory proteins of coronaviruses are translated from sub-genomic RNAs (sgRNAs) that are negative-stranded and synthesized by the replication-transcription complex (RTC). The distinct feature of S protein that paves the way to viral attachment with the host cell surface receptor makes it a promising target for vaccine design [8, 9, 10]. S protein is a trimeric class I fusion protein having S1 and S2 subunits. Cryo-electron microscopy structure and 3D atomic-scale map of S protein have already been developed. In order to fuse with the host cell membrane, the S1 subunit of S protein binds to the receptor that leads the transition of the S2 subunit from pre-fusion conformation to post-fusion conformation. This phenomenon happens for all coronavirus class-I fusion S protein. This transition of S protein for membrane fusion represents it as a target for Ab-mediated neutralization [11].

In the mammalian body, virus-infected cells are usually killed by cytotoxic action of CTLs which are signaled by the presentation of epitopes via HLA class-I alleles [12]. Epitopes for helper T-cells also relay signals for B-cells and macrophages. In general, T-helper cell epitopes presented by human HLA class-II alleles activate the B-cells to produce antibodies by plasma cells. Some memory B-cells derived from plasma B-cell keep the invader's memory to exhibit faster recognition and quick response during future encounters [13, 14]. Therefore, viral epitopes are the ideal candidates for vaccine construction that can trigger the immune system to develop immunity against the virus.

The early approach of immune therapeutics depended on the use of the whole antigen upon trial and error methods and some other cumbersome processes [15, 16]. The conventional ways of vaccine design were based on administering an entire disease-causing agent that consisted of the killed or live-attenuated organisms, detoxified toxins, etc. However, associated drawbacks of conventional vaccine design including difficulties in culture, killing or attenuation of the virus, identification of genetic variations that give rise to newer strains, and time-consuming nature in development have diverted the focus of researchers to other alternatives [17, 18]. To overcome these drawbacks, some modern approaches like recombinant DNA technology, polysaccharide chemistry, rational vaccinology, structural vaccinology, RNA vaccines, and epitope-based vaccine design have now been embraced by scientists. Several computational programs have also been developed to make these modern approaches fast and accurate based on genome data and protein 3D structures, simulation, and so on. Preliminary computational analyses also narrow down the investigation into a specific area thus saving both time and cost [19]. Immunoinformatics, also known as computational immunology, utilizes the advancement in computational capacity to develop various computational methods and resources to study the immune functions [20]. The epitope-based vaccine prediction for both B-cells and T-cells using immunoinformatics method can facilitate the development of anti-viral vaccine and can show effectiveness in preventing COVID-19. Moreover, epitope-based vaccine is effective in both pre-exposure and post-exposure periods which makes it more advantageous to fight COVID-19 since the repeated infection by the SARS-CoV-2 to the same individual has been reported in several cases [21, 22]. Concerning all the aspects of epitope-based vaccine design, this study is focused on the candidate epitope prediction for SARS-CoV-2 exploiting viral S protein and investigates the immunogenic properties of candidate epitopes as they are effective, safe and stable vaccine [23].

2. Materials and methods

2.1. Acquisition of sequence data and physicochemical properties analysis

The most recent protein sequence of SARS-CoV-2 S protein was acquired from NCBI (National Center for Biotechnology Information) using the accession number: QIC53213.1 in FASTA format and the updated structure (released on February 2020) of the SARS-CoV-2 S protein was retrieved from PDB (ID: 6VSB). To analyze various physical and chemical

parameters of retrieved S protein, an online tool ExPASy-ProtParam [24] was used (Supplementary Table S1).

2.2. B-cell epitope prediction

IEDB (Immune-Epitope-Database and Analysis-Resource) [25], a free accessible online resource tool, was used to predict the linear B-cell epitopes. These epitopes were predicted through the BepiPred Linear epitope prediction mode of IEDB. Based on the standard parameters including antigenicity, accessibility of surface, flexibility, hydrophilicity, and beta-turn, the epitopes were identified. These parameters were measured using Kolaskar and Tongaonkar antigenicity scale, Emini surface accessibility prediction tool, Karplus and Schulz flexibility prediction tool, Parker hydrophilicity prediction algorithms [26], Chou and Fasman beta-turn prediction algorithm respectively [27]. Prediction of protective antigens was carried out via Vaxijen v2.0 [28] server through forecasting vaxijen scores with the default threshold value 0.4 (70% accuracy). AllergenFP v.1.0 [29] server was used to examine the allergenicity of the predicted peptides. Furthermore, the server DiscoTope 2.0 [30] was used to predict discontinuous B-cell epitope from the S protein 3-D structure amino acid sequence in FASTA format, using the default threshold value of -3.7. This threshold value corresponds to a specificity of 75% and also indicates that 25% of non-epitope residues are predicted as part of epitopes. This method utilizes calculation of surface accessibility (estimated in terms of contact numbers) and a novel epitope propensity amino acid score.

2.3. T-cell epitope prediction

MHC class-I and MHC class-II alleles were predicted to obtain corresponding Cytotoxic T-lymphocyte (CTL) and T helper cell epitopes. T cell epitopes are crucial in vaccine design and prediction of them reduces the time and cost in wet-lab experiments.

2.3.1. MHC class-I alleles

A neural network-based MHC class-I binding peptide prediction server nHLAPred [31], containing the ComPred section, was used to forecast MHC class-I binding peptides or CTL epitopes of 67 MHC alleles. In this comprehensive method, threshold and stringency were maintained at 5% as well as proteasome and immune proteasome filters were applied. From this server, an amount of maximum allele binding epitopes based on their maximum antigenicity was selected. Antigenicity of corresponding epitopes as vaxijen scores was predicted by the Vaxijen v2.0 server.

2.3.2. MHC class-II alleles

An online free accessible server ProPred [32] was applied to predict the MHC class-II binding epitopes for a given antigen sequence. The method was applied for all alleles with threshold 5% and the top 10 scorers were displayed. In the same way as MHC class-I, an amount of maximum allele binding epitopes were predicted for MHC class-II from ProPred based on maximum antigenicity and allergenicity. Antigenicity was predicted from vaxijen scores of corresponding epitopes.

2.3.3. T-cell epitopes and features profiling

Physicochemical properties of selected MHC class-I and MHC class-II epitopes including maximum allele binding, antigenicity, allergenicity, digestion, mutation, toxicity, hydrophobicity, hydrophilicity, hydrophobicity, half-life, surface accessibility, polarity, charge, pI, and molecular weight were determined. An online server, peptide digestion server (<http://db.systemsbiology.net:8080/proteomicsToolkit/proteinDigest.html>), was utilized to measure the digestion of peptides by several digestive enzymes. The less the number of digestive enzymes the more the epitope stability, thus more favorable [33]. ToxinPred [34], an online tool, developed to predict and design toxic/non-toxic peptides, was used to measure mutation position, toxicity, hydrophobicity, hydrophilicity,

Table 1. B-cell epitopes of S protein predicted via IEDB analysis resource along with their start position, end position, length, vexijen scores (antigenicity), allergenicity, and toxicity.

Start	End	Peptide	Length	Vexijen score	Allergenicity	Toxicity
1157	1167	KNHTSPDVDLG	11	1.4039	Non-allergen	Non-toxin
407	420	VRQIAPGQTGKIAD	14	1.2606	Non-allergen	Non-toxin
805	816	ILPDPSPKSKRS	12	0.5322	Non-allergen	Non-toxin
249	261	LTPGDSSSGWTAG	13	0.4950	Non-allergen	Non-toxin
597	606	VITPGTNTSN	10	0.4217	Non-allergen	Non-toxin
21	31	RTQLPPAYTNS	11	0.8710	Allergen	Non-toxin
495	506	YGFQPTNGVGYQ	12	0.7136	Allergen	Non-toxin

hydropathicity, pI value, charge, and molecular weight of selected epitopes. Another tool, AllergenFP v.1.0, was utilized to measure the allergenicity of selected epitopes. A webserver, HLP [35], available for predicting half-life of peptides in intestine like environment, was

employed to measure different physicochemical properties of selected epitopes, including half-life, HPLC parameters, surface accessibility, and flexibility. An SVM based model was used for this server since the epitopes were 9 residues long.

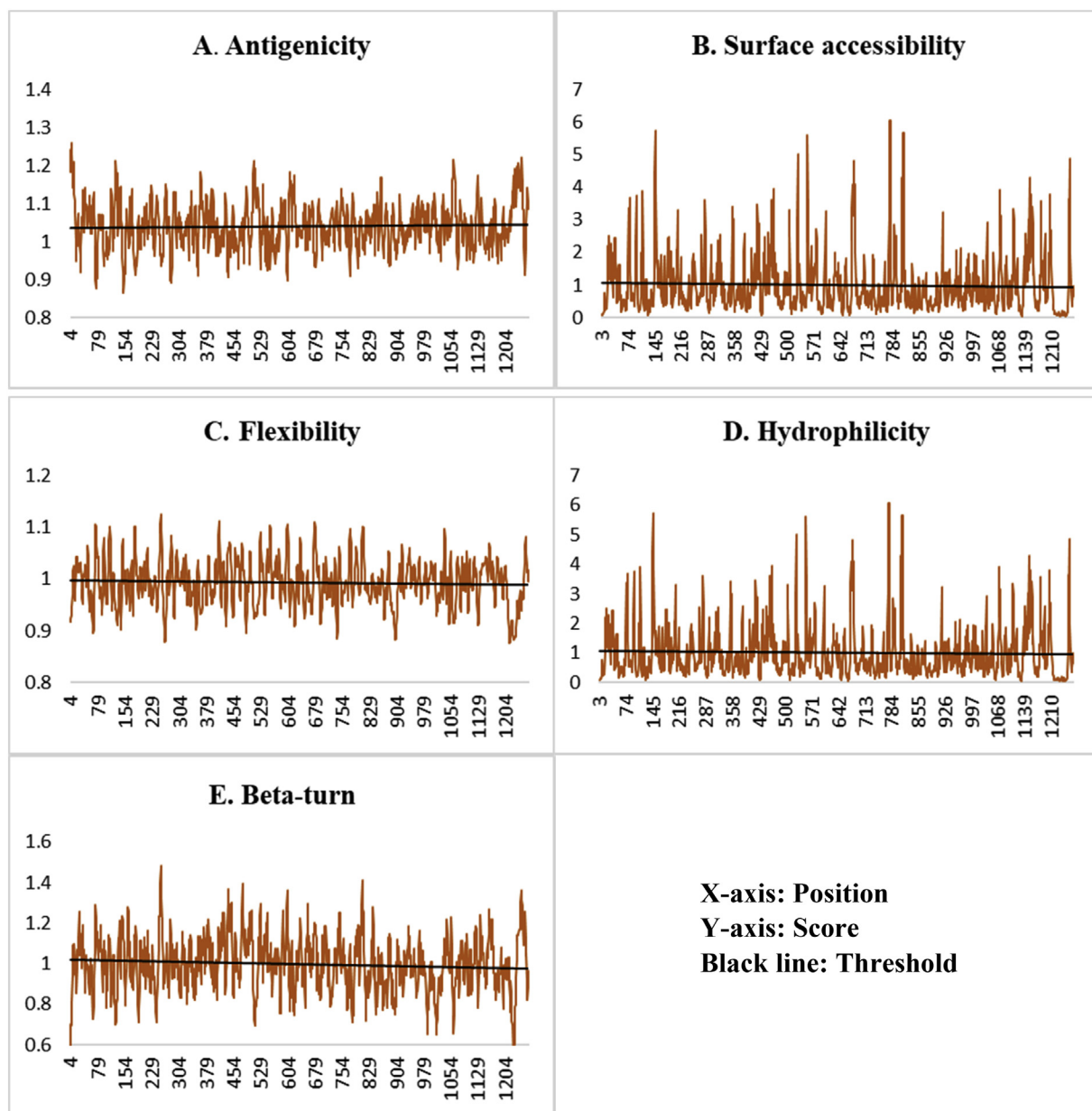


Figure 1. Prediction of B-cell epitopes of S protein by different IEDB scales (A. Kolaskar and Tongaonkar antigenicity, B. Emimi surface accessibility, C. Karplus and Schulz flexibility, D. Parker hydrophilicity, E. Chou and Fasman beta-turn). Regions above threshold (black line) are proposed to be a part of B cell epitope while regions below the threshold (black line) are not.

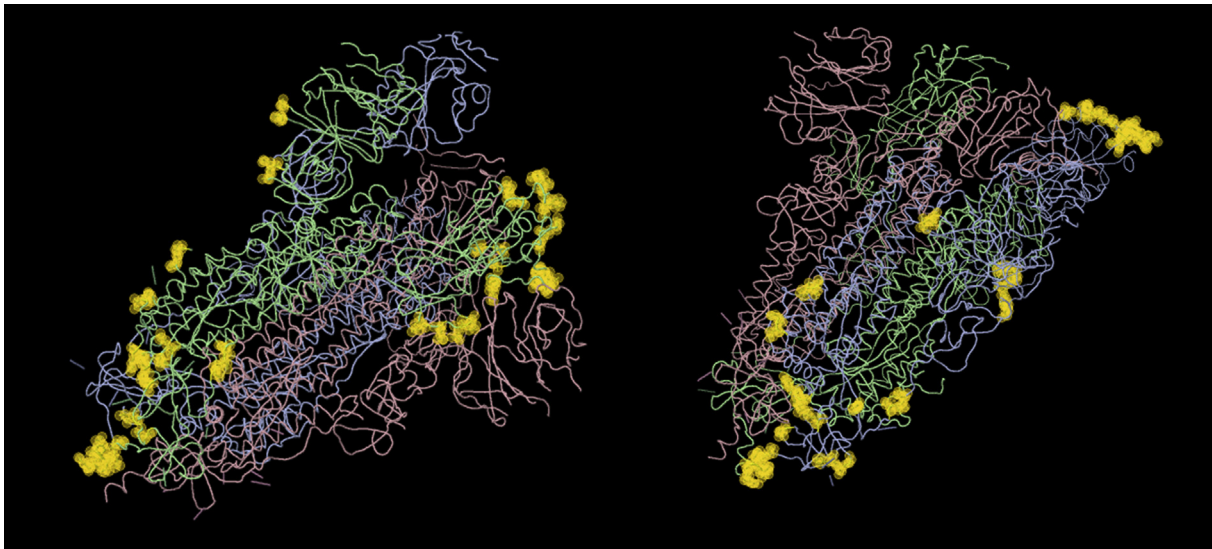


Figure 2. Discontinuous B-cell epitopes of S protein predicted by DiscoTope 2.0 server. Positions of predicted peptides are marked as yellow.

2.4. Structure determination and interaction study with HLA-B7

3D structures of all selected MHC class-I epitopes were modeled by an online protein structure prediction server namely the PEP-FOLD server [36]. The crystallized 3D model of the human HLA-B7 allele was retrieved from PDB (PDB ID: 3VCL) [37]. The interaction pattern and binding affinity between HLA-B7 and predicted MHC class-I epitopes of SARS-CoV-2 S protein were investigated using molecular docking tool AutoDock Vina 4.2 [38]. AutoDock Vina is an open-source program for molecular docking and AutoDock Vina significantly improves the average accuracy of the binding mode predictions. The PEP-FOLD server predicted five 3D models for each epitope and the best one was chosen according to its binding affinity with HLA-B7. The higher the binding affinity, the more stable the interaction. PyMOL tool was used to study the binding pattern and interactions.

3. Results

3.1. B-cell epitopes of S protein

Linear B-cell epitopes were predicted from IEDB (Bipred linear epitope prediction_B cell analysis tool), that gave several probable epitopes with different length. Among them, epitopes with 10–14 amino acid residues were selected for further studies since they can function as potential antigens. Various physicochemical properties of the initially determined epitopes were measured and filtered using several tools according to their properties such as length, vaxijen score for antigenicity, non-allergenicity, and non-toxicity (Supplementary Table S2). A total of 7 epitopes within this residue length and maximum allele binding ability were selected as the best antigenic epitopes which were KNHTSPDVLDG, VRQIAPGQTGKIAD, ILPDPSKPSKRS, LTPGDSSSGWTAG, VITPGTNTSN, RTQLPPAYTNS, YGFQPTNGVGYQ in Table 1. None of them were found to be toxic but two selected peptides showed allergenicity. So, a non-allergen and non-toxic epitope ‘KNHTSPDVLDG’ with residue length 11 and the highest vaxijen score 1.4039 was selected as a potential antigenic B-cell epitope.

Some essential parameters including antigenicity, accessibility of surface, flexibility, hydrophilicity, and beta-turn were also measured. Kolaskar and Tongaonkar antigenicity scale (with threshold 1.048) resulted in a minimum score of 0.866 and a maximum score of 1.261, and our selected peptide showed an antigenicity score of 1.0163 (Supplementary Table S3). Emini surface accessibility prediction tool (with threshold 1.00) resulted in a minimum score of 0.042 and a maximum

score of 6.051, and our selected peptide scored 1.33 in surface accessibility (Supplementary Table S4). Karplus and Schulz's flexibility prediction tool (with threshold 0.993) resulted in a minimum score of 0.876 and a maximum score of 1.125, and our selected peptide showed 1.033 in flexibility score (Supplementary Table S5). Parker hydrophilicity prediction algorithms (with threshold 1.238) resulted in a minimum score of -7.629 and a maximum score of 7.743, and our selected peptide showed 1.33 in hydrophilicity score (Supplementary Table S6). Chou and Fasman's beta-turn prediction algorithm (with threshold 0.997) resulted in a minimum score of 0.541 and a maximum score of 1.484, and our selected peptide showed a propensity score of 1.164 (Supplementary Table S7) presented in Figure 1. In the aforementioned analyses, threshold value denotes the average of all predicted residue scores, and the residues with scores above the threshold value are predicted to be part of an epitope.

Alongside the linear (continuous) B-cell epitopes, discontinuous (conformational) B-cell epitopes of S protein were also predicted and are given in Figure 2. List of the residues and other determined parameters of discontinuous epitopes predicted from both A-chain and B-chain of S protein using DiscoTope are presented in Supplementary Table S8.

3.2. T-cell epitopes of S protein

A total number of 14 T-cell epitopes of S protein were predicted using nHLAPred and ProPred for both MHC class-I and MHC class-II alleles respectively.

3.2.1. MHC class-I alleles

nHLAPred tool's ComPred section resulted in CTL epitopes for 67 MHC class-I alleles where threshold and stringency were maintained at 5%, and proteasome and immune proteasome filters were applied (Supplementary Table S9). Among all epitopes for each 67 alleles, 7 epitopes that possessed more than 13 binding alleles were selected for further study. These epitopes were chosen according to the properties such as vaxijen score for antigenicity, non-allergenicity, non-toxicity, and half-lives for better immunogenic response and stability (Supplementary Table S10). In this case, selected seven epitopes are VVLSFELL, TLDSKTQSL, FEYVSQPFL, GLTVLPPLL, WTFGAGAAL, VYDPLQPEL, GGFNFSQIL. Among all the 7 best-scored epitopes, VVLSFELL has the highest antigenicity according to its corresponding vaxijen score of 1.0909. It is found to be non-allergenic (retrieved using AllergenFP v.1.0) and can bind to a large number of alleles including HLA-A*0201, HLA-A*0206, HLA-A11, HLA-A24, HLA-A3, HLA-A*3101, HLA-A31, HLA-A*0301, HLA-A*3302, HLA-A2.1, HLA-B14, HLA-B27, HLA-B*2705,

Table 2. Positions of MHC class I allele binding peptides predicted by ProPred with their starting position in protein sequence, vaxijen score (antigenicity) and allergenicity.

Peptide	Position	MHC class I alleles	Vaxijen score	Allergenicity
VVVSFELL	510	HLA-A*0201, HLA-A*0206, HLA-A*1101, HLA-A11, HLA-A24, HLA-A*2402, HLA-A3, HLA-A*3101, HLA-A31, HLA-A*0301, HLA-A*3302, HLA-A68.1, HLA-A2.1 HLA-B14, HLA-B27, HLA-B*2705, HLA-B*3501, HLA-B*3701, HLA-B*3901, HLA-B*5101, HLA-B*5102, HLA-B*5103, HLA-B*5301, HLA-B*5401, HLA-B*51, HLA-B7, HLA-B*0702, HLA-B8, HLA-Cw*0301, H2-Db, H2-Kb, H2-Kd, H2-Ld, H-2Qa, HLA-B35, Mamu-A*01, HLA-B*2703, HLA-B*2704, HLA-A*6801, HLA-A*6802	1.0909	Non-allergen
TLDSKTQSL	109	HLA-A*0202, HLA-A*1101, HLA-A11, HLA-A*2402, HLA-A3, HLA-A*3101, HLA-A31, HLA-A*0301, HLA-A20, HLA-A2.1, HLA-B14, HLA-B*2702, HLA-B27, HLA-B*2705, HLA-B*3501, HLA-B*3801, HLA-B*3901, HLA-B*5101, HLA-B*5301, HLA-B*5401, HLA-B*51, HLA-B7, HLA-B*0702, HLA-B8, HLA-Cw*0602, H2-Db, H2-Kb, H2-Kd, H2-Ld, HLA-B35, HLA-A*0204, HLA-A*6801, HLA-A*6802	1.0685	Non-allergen
FEYVSQPFL	168	HLA-A*0202, HLA-A*0203, HLA-A*0206, HLA-B*2702, HLA-B*3701, HLA-B40, HLA-B*4403, HLA-B*5201, HLA-B60, HLA-B61, HLA-Cw*0301, H2-Dd, H2-Kk, HLA-B*2704, HLA-A*3301, HLA-B44	0.6324	Non-allergen
GLTVLPPLL	853	HLA-A2, HLA-A*0201, HLA-A*0202, HLA-A*0203, HLA-A*0205, HLA-A*1101, HLA-A11, HLA-A*2402, HLA-A3, HLA-A31, HLA-A*0301, HLA-A20, HLA-A2.1, HLA-B14, HLA-B*2702, HLA-B27, HLA-B*2705, HLA-B*3501, HLA-B*3701, HLA-B*3902, HLA-B*5101, HLA-B*5301, HLA-B*5401, HLA-B*51, HLA-B62, HLA-B7, HLA-B*0702, HLA-B8, HLA-Cw*0602, H2-Db, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, HLA-A*0204, HLA-B*2704, HLA-A*3301, HLA-A*6801, HLA-A*6802	0.6621	Non-allergen
WTFGAGAAL	882	HLA-A*0205, HLA-A24, HLA-A*3101, HLA-A68.1, HLA-B*2702, HLA-B*3801, HLA-B*3901, HLA-B40, HLA-B*5801, HLA-B61, HLA-B7, HLA-Cw*0401, HLA-A*0204, HLA-B*2704, HLA-B*2902, HLA-B44	0.4918	Non-allergen
VYDPLQPEL	1133	HLA-A*0201, HLA-A*0203, HLA-A*0205, HLA-A*1101, HLA-A11, HLA-A*2402, HLA-A3, HLA-A31, HLA-A*0301, HLA-A2.1, HLA-B14, HLA-B27, HLA-B*2705, HLA-B*3501, HLA-B*3801, HLA-B*3901, HLA-B*5101, HLA-B*5301, HLA-B*5401, HLA-B*51, HLA-B7, HLA-B*0702, HLA-B8, HLA-Cw*0602, HLA-Cw*0702, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-B*2703, HLA-B*2704, HLA-A*3301, HLA-A*6801, HLA-A*6802	0.4525	Non-allergen
GGFNFSQIL	794	HLA-B*3901, HLA-B*3902, HLA-B40, HLA-B*5102, HLA-B*5103, HLA-B*5201, HLA-B61, HLA-Cw*0301, HLA-Cw*0602, H2-Dd, H2-Kb, Mamu-A*01, HLA-B*2704	0.7967	Allergen

HLA-B*3501, HLA-B*3701, HLA-B*5101, HLA-B*5102, HLA-B*5301, HLA-B*5401, HLA-B*51, HLA-B7, HLA-B8, HLA-Cw*0301, H2-Db, H2-Kb, H2-Kd, H2-Ld, HLA-B35, Mamu-A*01. So, for MHC class-I, 'VVVSFELL' was selected as the best one as it had the highest vaxijen score. The primarily selected 7 epitopes with a list of possible alleles for

binding, starting position, antigenicity score, and allergenicity are given in [Table 2](#).

3.2.2. MHC class-II alleles

ProPred was operated to predict the MHC class-II binding epitopes with threshold 5% and top scorers 10 (Supplementary Table S11). For

Table 3. Positions of MHC class II allele binding peptides predicted by ProPred with their starting position in protein sequence, vaxijen score (antigenicity), and allergenicity.

Peptide	Position	MHC class II alleles	Vaxijen score	Allergenicity
VVIGIVNNT	1123	DRB1_0301, DRB1_0306, DRB1_0307, DRB1_0308, DRB1_0309, DRB1_0311, DRB1_1107	1.3063	Non-allergen
IGINITRFQ	230	DRB1_0402, DRB1_0804, DRB1_0806, DRB1_1102, DRB1_1114, DRB1_1121, DRB1_1301, DRB1_1304, DRB1_1322, DRB1_1323, DRB1_1327, DRB1_1328	1.3386	Non-allergen
VLSFELLNA	511	DRB1_1104, DRB1_1106, DRB1_1311, DRB1_1501, DRB1_1502, DRB1_1506	1.0433	Non-allergen
FKIYSKHTP	200	DRB1_0801, DRB1_0802, DRB1_0804, DRB1_0806, DRB1_0813, DRB1_0817, DRB1_1114, DRB1_1120, DRB1_1302, DRB1_1323, DRB1_1502	0.9886	Non-allergen
FVFLVLLPL	1	DRB1_0101, DRB1_0102, DRB1_0701, DRB1_0703, DRB1_0801, DRB1_0802, DRB1_0817, DRB1_1101, DRB1_1104, DRB1_1106, DRB1_1114, DRB1_1120, DRB1_1128, DRB1_1302, DRB1_1305, DRB1_1307, DRB1_1311, DRB1_1321, DRB1_1323, DRB1_1501, DRB1_1502, DRB1_1506, DRB5_0101, DRB5_0105	0.8601	Non-allergen
FLVLLPLVS	3	DRB1_0101, DRB1_0102, DRB1_0401, DRB1_0408, DRB1_0421, DRB1_0426, DRB1_0802, DRB1_0813, DRB1_0817, DRB1_1101, DRB1_1104, DRB1_1106, DRB1_1114, DRB1_1120, DRB1_1128, DRB1_1302, DRB1_1305, DRB1_1307, DRB1_1311, DRB1_1321, DRB1_1323, DRB5_0101, DRB5_0105	0.4266	Non-allergen
LVLLPLVSS	4	DRB1_0102, DRB1_0301, DRB1_0305, DRB1_0306, DRB1_0307, DRB1_0308, DRB1_0309, DRB1_0311, DRB1_0404, DRB1_0423, DRB1_0802, DRB1_0804, DRB1_0806, DRB1_0813, DRB1_1101, DRB1_1102, DRB1_1104, DRB1_1106, DRB1_1107, DRB1_1114, DRB1_1120, DRB1_1121, DRB1_1128, DRB1_1301, DRB1_1302, DRB1_1304, DRB1_1305, DRB1_1307, DRB1_1311, DRB1_1321, DRB1_1322, DRB1_1323, DRB1_1327, DRB1_1328	0.6523	Allergen

Table 4. Physicochemical properties of predicted MHC class I peptides.

Peptide	Half-life	Mutation	Toxicity	Hydrophobicity	Hydropathicity	Hydrophilicity	Charge	pI	MW	HPLC parameter	Surface accessibility	Flexibility	Polarity
VVVSFELL	0.865	NM	NT	0.33	2.50	-1.01	-1.00	4.00	1018.40	-3.733	32.178	3.570	53.670
TLDSKTQSL	2.795	NM	NT	-0.26	-0.70	0.27	0.00	6.19	992.22	2.967	51.278	4.090	101.290
FEYVSQPFL	0.154	NM	NT	0.07	0.32	-0.79	-1.00	4.00	1129.40	-1.200	43.100	3.800	59.580
GLTVLPPLL	3.298	NM	NT	0.28	1.68	-1.01	0.00	5.88	922.32	-2.822	34.067	3.830	6.750
WTFGAGAAL	1.741	NM	NT	0.27	1.07	-1.07	0.00	5.88	893.13	-0.611	29.467	3.570	4.560
VYDPLQPEL	0.520	NM	NT	-0.06	-0.36	-0.13	-2.00	3.67	1073.34	0.444	48.289	4.050	99.940
GGFNFSQIL	1.603	NM	NT	0.13	0.59	-0.88	0.00	5.88	982.24	-0.522	35.689	3.980	9.880

NM: No Mutation, NT: Non-Toxin, MW: Molecular Weight.

Table 5. Physicochemical properties of predicted MHC class II peptides.

Peptide	Half-life	Mutation	Toxicity	Hydrophobicity	Hydropathicity	Hydrophilicity	Charge	pI	MW	HPLC parameter	Surface accessibility	Flexibility	Polarity
VVIGIVNNT	1.071	NM	NT	0.20	1.50	-0.90	0.00	5.88	928.24	-0.244	34.044	3.990	9.110
IGINITRFQ	0.689	NM	NT	-0.03	0.41	-0.54	1.00	10.11	1061.39	-0.567	42.989	4.150	61.370
VLSFELLNA	1.384	NM	NT	0.16	1.38	-0.71	-1.00	4.00	1005.31	-1.900	36.678	3.610	56.790
FKIYSKHTP	1.136	NM	NT	-0.19	-0.90	-0.13	2.50	9.72	1120.44	0.911	55.411	3.910	157.620
FVFLVLLPL	1.735	NM	NT	0.48	3.07	-1.69	0.00	5.88	1060.53	-6.722	28.922	3.350	4.340
FLVLLPLVS	0.771	NM	NT	0.39	2.67	-1.38	0.00	5.88	1000.43	-4.978	30.756	3.550	5.660
LVLPLVSS	0.628	NM	NT	0.29	2.27	-1.07	0.00	5.88	940.33	-3.233	32.589	3.750	6.980

NM: No Mutation, NT: Non-Toxin, MW: Molecular Weight.

Table 6. Best binding affinity and interacting residues.

Peptide	Binding affinity	Interacting residues
VVLSFELL	-5.7	ARG35, PRO47, ARG48, GLU50, GLU53, TYR67, ARG239, GOL304
TLDSKTQSL	-5.8	PHE22, GLN32, ARG48, GLU69, LYS178, ASP183, ARG239, THR240
FEYVSQPFL	-7.0	ARG12, HIS13, PRO14, LYS186, THR187, HIS188, LEU206, GLY207, LEU270, LEU272, ARG273, GLU275
GLTVLPPLL	-6.0	THR187, HIS188, PRO267, LEU270, THR271, LEU272, ARG273, GLU275
WTFGAGAAL	-6.1	ARG12, GLN32, ARG48, GLY237, ARG239
VYDPLQPEL	-5.8	ARG12, PRO47, ARG48, PRO50, GLU53, TYR67, ARG239
GGFNFSQIL	-6.5	LYS178, ARG181, ASP183, LYS186, THR187, HIS188, THR240, PRO267, GLU275

MHC class-II, 7 epitopes were selected on the basis of their ability to bind more than 5 alleles and also filtered based on the similar properties investigated for the MHC class-I epitopes (Supplementary Table S12). According to these properties, the 7 best scoring epitopes are IGINITRFQ, VVIGIVNNT, VLSFELLNA, FKIYSKHTP, FVFLVLLPL, FLVLLPLVS, LVLPLVSS. The epitope 'IGINITRFQ' has a maximum vaxijen score of 1.3386 but 'VVIGIVNNT' has almost similar vaxijen score of 1.3063 which also indicates its extensive antigenicity too. It has also been found as non-allergen by AllergenFP v.1.0 and can bind a number of MHC class-II alleles including DRB1_0402, DRB1_0804, DRB1_0806, DRB1_1102, DRB1_1114, DRB1_1121, DRB1_1301, DRB1_1304, DRB1_1322, DRB1_1323, DRB1_1327, DRB1_1328 and DRB1_0301, DRB1_0306, DRB1_0307, DRB1_0308, DRB1_0309, DRB1_0311, DRB1_1107. The primarily selected 7 epitopes with a list of possible alleles for binding, starting position, antigenicity score, and allergenicity for MHC class-II binding peptides are presented in Table 3.

3.2.3. Features of T-cell epitopes

Important physicochemical properties of selected CTL epitopes for MHC class-I and T helper cell epitopes for MHC class-II were determined using several tools. The peptide digestion server was used to predict the digestion of selected peptides. The more the number of digesting enzymes the less its stability, and vice-versa. The enzymes Trypsin, Clostripain, Cyanogen Bromide, Iodoso Benzoate, Proline Endopept, Trypsin K, Trypsin R, and AspN cannot digest the selected VVLSFELL epitope for MHC class-I. The non-digesting enzymes of MHC class-I binding epitopes were also identified (Supplementary Table S13). The enzymes Trypsin, Chymotrypsin, Clostripain, Cyanogen Bromide, IodosoBenzoate, Proline Endopept, Staph Protease, Trypsin K, Trypsin R and AspN cannot digest the selected VVIGIVNNT epitope for MHC class- II. The non-digesting enzymes of MHC class-II binding epitopes were also determined (Supplementary Table S14).

ToxinPred was used to predict and design toxic/non-toxic peptides and also operated to measure mutation position, toxicity, hydrophobicity, hydrophilicity, hydropathicity, pI value, charge, and molecular weight of selected epitopes. None of these peptides were found to be toxic. HLP was used for predicting half-life, HPLC parameters, surface accessibility, and flexibility based on the SVM model. The physicochemical properties of selected epitopes including half-life, mutation,

toxicity, hydrophobicity, hydrophilicity, hydropathicity, charge, pI, molecular weight, HPLC parameters, surface accessibility, flexibility, and polarity are given in Table 4 for MHC class-I binding CTL epitopes and in Table 5 for MHC class-II binding T helper cell epitopes.

3.3. Molecular docking studies

Corresponding 3D structures of CTL epitopes for MHC class-I binding peptides were predicted via PEP-FOLD server and 5 models for every 7 epitopes were obtained. From the 5 models for each epitope, the best model was selected according to the highest binding affinity with HLA-B7 protein by molecular docking. All 7 MHC class-I epitopes rather than the best one identified by other metrics were also included in the docking study to check whether the other epitopes could also bind to HLA B7. Binding affinity scores of the best model of all 7 epitopes with HLA B7 and their corresponding interacting residues are listed in Table 6. Among the best 7 epitopes, 'FEYVSQPFL' exhibited the highest binding affinity of -7.0 kcal/mol and candidate epitope of this study 'VVLSFELL' exhibited binding affinity of -5.7 kcal/mol. The binding affinity of each epitope is listed in Table 6. PyMOL tool was used to visualize the binding pattern and interacting residues. Several numbers of residues of HLA B7 were found to interact with the candidate epitope 'VVLSFELL' and the residues are: ARG-35, PRO-47, ARG-48, GLU-50, GLU-53, TYR-67, ARG-239, GOL-304. The visualization of best interactions for all 7 epitopes is presented in Figure 3.

4. Discussion

Application of epitope-based vaccine has overcome the obstacles associated with conventional vaccines including undesired allergenic response due to large protein size, time-consuming nature of development, and it is also an effective process to treat viral infection in the post-exposure state [39, 40]. The epitope-based vaccine is a type of subunit vaccine which utilizes small molecular weight peptides known as epitope that do not cause any infection and not prone to any mutation. It is both timesaving and cost-effective to construct a vaccine by this approach that can generate strong immunogenic responses [41, 42]. Prediction of epitope-based vaccines using computational tools have been successfully

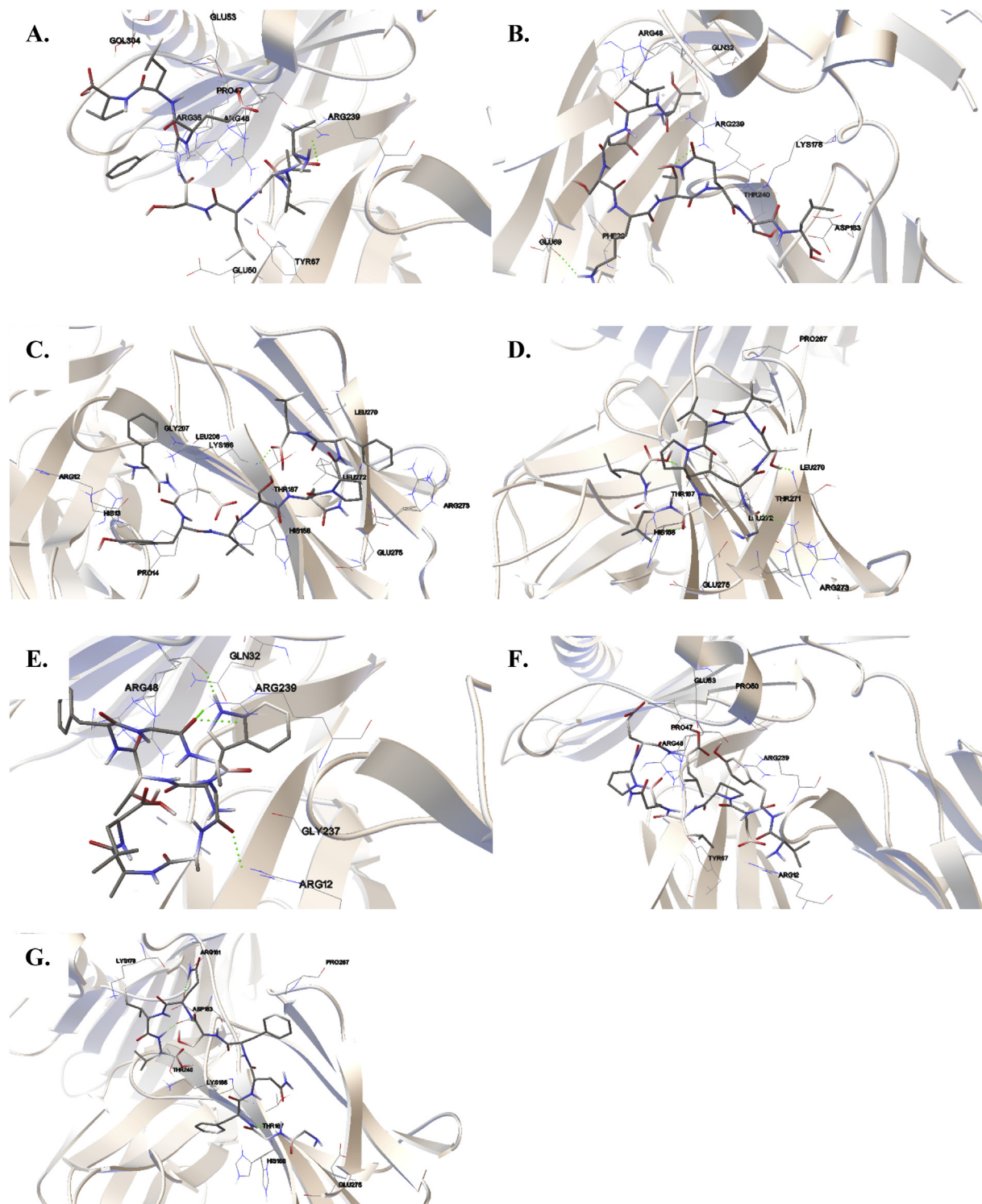


Figure 3. Representation of the interactions between HLA B7 and the predicted MHC class I alleles' binding peptides, respectively, (A) VVLSFELL, (B) TLDSKTQSL, (C) FEYVSQPFL, (D) GLTVLPPLL, (E) WTFGAGAAL, (F) VYDPLQPEL, and (G) GGFNFSQIL. The peptides are in symmetry with the information provided in Table 6. Ribbon structure represents HLA B7, wire structure represents peptide, and green dotted line represents hydrogen bond between interacting residues.

reported against the Chikungunya virus, Zika virus, Ebola virus, Herpes simplex virus, Yellow fever virus, etc. [43, 44, 45, 46, 47]. The focus of this study was to predict both B-cell and T-cell epitopes using the S protein of SARS-CoV-2 by immunoinformatics method. S protein of SARS-CoV-2 was selected for its immunogenicity along with the important role in viral attachment to the host cell surface [48, 49, 50]. Also,

epitopes of viral S protein (such as SARS-CoV S protein) are the most antigenic part that elicit an intrinsic immunogenic response by signaling immune cells to increase both humoral and adaptive immunity [51, 52, 53].

B-cell epitope mapping is the keystone in the development of diagnostics, and it is an initial step to design potent vaccines [54, 55, 56,

57]. Both continuous (linear) and discontinuous (conformational) type of B-cell epitopes are crucial to the peptide-based vaccine and disease diagnosis [58, 59]. Unlike the linear epitopes in which residues occur in a consecutive fashion in the primary sequence, residues of discontinuous epitopes happen in a distant position of primary sequence but are drawn to proximity due to peptide folding [60]. B-cell linear epitopes were predicted by the IEDB analysis-resource tool that generated epitopes comprising more than nine amino acid residues. Using several tools various physicochemical properties of the initially determined epitopes were measured and they were filtered. Among them, KNHTSPDVLG peptide was selected as a strong immunogen and stable epitope according to its highest antigenic score (1.4039), non-allergenicity, and toxicity.

Similarly, CTL epitopes for MHC class-I allele binding were predicted by nHLAPred, and for the prediction of MHC class II peptide, ProPred tool was employed. These epitopes were also filtered according to their different properties for better immunogenic response and stability. In this case, the best antigenic score of 1.0909 was found for VVVSFELL peptide for MHC class-I allele, and for MHC class-II epitope, 'IGINTRFQ' has a maximum vaxijen score of 1.3386 but 'VVIGIVNNT' has almost similar vaxijen score of 1.3063 and exhibited the highest non-digesting enzymes. According to their highest antigenicity and other properties, VVVSFELL and VVIGIVNNT were selected as the best epitopes that could be able to work as strong immunogens.

As shown in Tables 1, 2, and 3, the results of this study identified KNHTSPDVLG as B-cell epitope, VVVSFELL epitope for MHC class-I binding, and VVIGIVNNT for MHC class-II binding of T cell. These are predicted to be potential vaccine candidates due to their features and high antigenic nature to elicit an intrinsic immunogenic response.

Furthermore, the interaction between primarily selected MHC class I epitopes and human HLA B7 allele were analyzed to understand the binding affinities between them. Molecular docking of HLA B7 with each epitope modeled by PepFold denoted a stable interaction between HLA B7 and each ligand. Several significant hydrogen bonds were present in each docked complex (Figure 3) indicating stable use of these peptides as potential immunogens. Although the candidate MHC class-I epitope 'VVVSFELL' has not been found to possess the highest binding affinity score with HLA B7 among remaining epitopes, a score of -5.7 kcal/mol is significant and non-negligible, nonetheless. Moreover, this epitope is preferred due to the other properties such as antigenicity, the number of MHC binding alleles, allergenicity, etc. HLA-B7 was selected for the docking study as a representative protein of MHC class-I alleles since HLA-B is the most prevalent serotype and is available at a higher frequency in humans compared to the other MHC class-I molecules [61]. HLA-B7 also plays a major role in viral epitope binding in viral infection such as HIV and MERS-CoV [33, 62]. HLA-B7 is also one of the common and focused haplotypes in prediction of HLA binding to viral derived peptides of coronavirus [63, 64].

Physicochemical properties (peptide digestion, half-life, hydrophobicity, flexibility, toxicity, etc.), immunogenic properties (allergenicity, antigenicity), and stable HLA allele binding capability reinforce the potentiality of these peptides as effective, safe and stable vaccine candidates.

5. Conclusion

This study has presented several epitopes of the S protein of SARS-CoV-2 for both B-cell and T-cell screened for efficient immunogenic response according to their antigenic nature. This in silico method of epitope-based vaccine design can save both time and cost by filtering out a number of vaccine candidates and associated trials. This study can be used as a steppingstone and advanced further to tackle the challenge imposed by SARS-CoV-2.

Declarations

Author contribution statement

Arafat Islam Ashik: Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Mahedi Hasan: Performed the experiments; Wrote the paper.

Atiya Tahira Tasnim: Contributed reagents, materials, analysis tools or data; Wrote the paper.

Md. Belal Chowdhury: Analyzed and interpreted the data; Wrote the paper.

Tanvir Hossain: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data.

Shamim Ahmed: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interest statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at <https://doi.org/10.1016/j.heliyon.2020.e04865>.

Acknowledgements

The authors would like to acknowledge the Department of Biochemistry and Molecular Biology, Shahjalal University of Science and Technology, Sylhet for laboratory facilities and logistic supports.

References

- [1] C. Wang, P.W. Horby, F.G. Hayden, G.F. Gao, A novel coronavirus outbreak of global health concern, *Lancet* 395 (2020) 470–473.
- [2] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F. Zhan, L. Wang, T. Hu, H. Zhou, Z. Hu, W. Zhou, L. Zhao, J. Chen, Y. Meng, J. Wang, Y. Lin, J. Yuan, Z. Xie, J. Ma, W.J. Liu, D. Wang, W. Xu, E.C. Holmes, G.F. Gao, G. Wu, W. Chen, W. Shi, W. Tan, Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, *Lancet* 395 (2020) 565–574.
- [3] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, *Lancet* 395 (2020) 497–506.
- [4] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K.S.M. Leung, E.H.Y. Lau, J.Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T.T.Y. Lam, J.T. Wu, G.F. Gao, B.J. Cowling, B. Yang, G.M. Leung, Z. Feng, Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia, *N. Engl. J. Med.* 382 (2020) 1199–1207.
- [5] J.F.-W. Chan, S. Yuan, K.-H. Kok, K.K.-W. To, H. Chu, J. Yang, F. Xing, J. Liu, C.C.-Y. Yip, R.W.-S. Poon, H.-W. Tsoi, S.K.-F. Lo, K.-H. Chan, V.K.-M. Poon, W.-M. Chan, J.D. Ip, J.-P. Cai, V.C.-C. Cheng, H. Chen, C.K.-M. Hui, K.-Y. Yuen, A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster, *Lancet* 395 (2020) 514–523.
- [6] W.-B. Yu, G.-D. Tang, L. Zhang, R.T. Corlett, Decoding the evolution and transmissions of the novel pneumonia coronavirus (SARS-CoV-2/HCoV-19) using whole genomic data, *Zool. Res.* 41 (2020) 247–257.
- [7] M. Prajapat, P. Sarma, N. Shekhar, P. Avti, S. Sinha, H. Kaur, S. Kumar, A. Bhattacharyya, H. Kumar, S. Bansal, B. Medhi, Drug targets for corona virus: a systematic review, *Indian J. Pharmacol.* 52 (2020) 56.

- [8] Y. Chen, Q. Liu, D. Guo, Emerging coronaviruses: genome structure, replication, and pathogenesis, *J. Med. Virol.* 92 (2020) 418–423.
- [9] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T.S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche, M.A. Müller, C. Drosten, S. Pöhlmann, SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven Protease inhibitor, *Cell* 181 (2020) 271–280, e8.
- [10] D. Schoeman, B.C. Fielding, Coronavirus envelope protein: current knowledge, *Virol. J.* 16 (2019) 69.
- [11] D. Wrapp, N. Wang, K.S. Corbett, J.A. Goldsmith, C.-L. Hsieh, O. Abiona, B.S. Graham, J.S. McLellan, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, *Science* 367 (2020) 1260–1263.
- [12] M.H. Andersen, D. Schrama, P. thor Straten, J.C. Becker, Cytotoxic T cells, *J. Invest. Dermatol.* 126 (2006) 32–41.
- [13] R.L. Coffman, B.W.P. Seymour, D.A. LeBman, D.D. Hiraki, J.A. Christiansen, B. Shrader, H.M. Cherwinski, H.F.J. Savelkoul, F.D. Finkelman, M.W. Bond, T.R. Mosmann, The role of helper T cell products in mouse B cell differentiation and isotype regulation, *Immunol. Rev.* 102 (1988) 5–28.
- [14] D.C. Parker, T cell-dependent B cell activation, *Annu. Rev. Immunol.* 11 (1993) 331–360.
- [15] T. Ahmad, The development of immunization trials, *Edorium J. Biotechnol.* 1 (2015).
- [16] A. Arora, N. Mendoza, A.M. Tremaine, S.K. Tyring, Clinical and basic immunodermatology, *Vaccines* 12 (2008) 597–603.
- [17] R. Arnon, T. Ben-Yedidia, Old and new vaccine approaches, *Int. Immunopharmacol.* 3 (2003) 1195–1204.
- [18] S.K. Grimm, M.E. Ackerman, Vaccine design: emerging concepts and renewed optimism, *Curr. Opin. Biotechnol.* 24 (2013) 1078–1088.
- [19] A. Sajid Sunita, Y. Singh, P. Shukla, Computational tools for modern vaccine development, *Hum. Vaccines Immunother.* 16 (2020) 723–735.
- [20] V. Brusic, N. Petrovsky, Immunoinformatics and its relevance to understanding human immune disease, *Expert Rev. Clin. Immunol.* 1 (2005) 145–157.
- [21] M. Bhattacharya, A.R. Sharma, P. Patra, P. Ghosh, G. Sharma, B.C. Patra, S.-S. Lee, C. Chakraborty, Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-CoV-2): immunoinformatics approach, *J. Med. Virol.* 92 (2020) 618–631.
- [22] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, Y. Zhao, Y. Li, X. Wang, Z. Peng, Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China, *J. Am. Med. Assoc.* 323 (2020) 1061–1069.
- [23] M. Kamthania, D.K. Sharma, Screening and structure-based modeling of T-cell epitopes of Nipah virus proteome: an immunoinformatic approach for designing peptide-based vaccine, *3 Biotech* 5 (2015) 877–882.
- [24] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein identification and analysis tools on the ExPASy server, in: J.M. Walker (Ed.), *Proteomics Protoc. Handb.*, Humana Press, Totowa, NJ, 2005, pp. 571–607.
- [25] R. Vita, S. Mahajan, J.A. Overton, S.K. Dhandra, S. Martini, J.R. Cantrell, D.K. Wheeler, A. Sette, B. Peters, The immune epitope database (IEDB): 2018 update, *Nucleic Acids Res.* 47 (2019) D339–D343.
- [26] M.C. Jespersen, B. Peters, M. Nielsen, P. Marcatili, BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes, *Nucleic Acids Res.* 45 (2017) W24–W29.
- [27] P. Koehl, M. Levitt, Structure-based conformational preferences of amino acids, *Proc. Natl. Acad. Sci.* 96 (1999) 12524–12529.
- [28] I.A. Doytchinova, D.R. Flower, VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines, *BMC Bioinf.* 8 (2007) 4.
- [29] I. Dimitrov, L. Naneva, I. Doytchinova, I. Bangov, AllergenFP: allergenicity prediction by descriptor fingerprints, *Bioinformatics* 30 (2014) 846–851.
- [30] J.V. Kringelum, C. Lundegaard, O. Lund, M. Nielsen, Reliable B cell epitope predictions: impacts of method development and improved benchmarking, *PLoS Comput. Biol.* 8 (2012), e1002829.
- [31] H.-P. Adams, J.A. Koziol, Prediction of binding to MHC class I molecules, *J. Immunol. Methods* 185 (1995) 181–190.
- [32] H. Singh, G.P. Raghava, ProPred: prediction of HLA-DR binding sites, *Bioinforma. Oxf. Engl.* 17 (2001) 1236–1237.
- [33] M. Tahir ul Qamar, S. Saleem, U.A. Ashfaq, A. Bari, F. Anwar, S. Alqahtani, Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: an immune-informatics study, *J. Transl. Med.* 17 (2019) 362.
- [34] S. Gupta, P. Kapoor, K. Chaudhary, A. Gautam, R. Kumar, O.S.D.D. Consortium, G.P.S. Raghava, In silico approach for predicting toxicity of peptides and proteins, *PLoS One* 8 (2013), e73957.
- [35] A. Sharma, D. Singla, M. Rashid, G.P.S. Raghava, Designing of peptides with desired half-life in intestine-like environment, *BMC Bioinf.* 15 (2014) 282.
- [36] J. Maupetit, P. Derreumaux, P. Tufféry, A fast method for large-scale De Novo peptide and miniprotein structure prediction, *J. Comput. Chem.* 31 (2010) 726–738.
- [37] R.M. Brennan, J. Petersen, M.A. Neller, J.J. Miles, J.M. Burrows, C. Smith, J. McCluskey, R. Khanna, J. Rossjohn, S.R. Burrows, The impact of a large and frequent deletion in the human TCR β locus on antiviral immunity, *J. Immunol.* 188 (2012) 2742–2748.
- [38] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility, *J. Comput. Chem.* 30 (2009) 2785–2791.
- [39] V. Chauhan, T. Rungta, K. Goyal, M.P. Singh, Designing a multi-epitope based vaccine to combat Kaposi Sarcoma utilizing immunoinformatics approach, *Sci. Rep.* 9 (2019) 1–15.
- [40] O. Finco, R. Rappuoli, Designing vaccines for the twenty-first century society, *Front. Immunol.* 5 (2014).
- [41] E. Raoufi, M. Hemmati, S. Eftekhari, K. Khaksaran, Z. Mahmodi, M.M. Farajollahi, M. Mohsenzadegan, Epitope prediction by novel immunoinformatics approach: a state-of-the-art review, *Int. J. Pept. Res. Therapeut.* 26 (2020) 1155–1163.
- [42] A. Sette, J. Fikes, Epitope-based vaccines: an update on epitope identification, vaccine design and delivery, *Curr. Opin. Immunol.* 15 (2003) 461–470.
- [43] L.C. Chong, A.M. Khan, Vaccine target discovery, *Encycl. Bioinforma. Comput. Biol.* (2019) 241–251.
- [44] S.F. de O. Tosta, M.S. Passos, R. Kato, Á. Salgado, J. Xavier, A.K. Jaiswal, S.C. Soares, V. Azevedo, M. Giovanetti, S. Tiwari, L.C.J. Alcantara, Multi-epitope based vaccine against yellow fever virus applying immunoinformatics approaches, *J. Biomol. Struct. Dyn.* (2019) 1–17.
- [45] R. Rappuoli, Reverse vaccinology, *Curr. Opin. Microbiol.* 3 (5) (2000) 445–450.
- [46] M. Hasan, S. Islam, S. Chakraborty, A.H. Mustafa, K.F. Azim, Z.F. Joy, M.N. Hossain, S.H. Foysal, M.N. Hasan, Contriving a chimeric polyvalent vaccine to prevent infections caused by herpes simplex virus (type-1 and type-2): an exploratory immunoinformatic approach, *J. Biomol. Struct. Dyn.* (2019) 1–18.
- [47] M.A. Khan, M.U. Hossain, S.M. Rakib-Uz-Zaman, M.N. Morshed, Epitope-based peptide vaccine design and target site depiction against Ebola viruses: an immunoinformatics study, *Scand. J. Immunol.* 82 (2015) 25–34.
- [48] A. Grifoni, J. Sidney, Y. Zhang, R.H. Scheuermann, B. Peters, A. Sette, A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2, *Cell Host Microbe* 27 (2020) 671–680, e2.
- [49] A.C. Walls, Y.-J. Park, M.A. Tortorici, A. Wall, A.T. McGuire, D. Velesler, Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein, *Cell* 181 (2020) 281–292, e6.
- [50] J. Li, C. Luo, Y. Deng, Y. Han, L. Tang, J. Wang, J. Ji, J. Ye, F. Jiang, Z. Xu, W. Tong, W. Wei, Q. Zhang, S. Li, W. Li, H. Li, Y. Li, W. Dong, J. Wang, S. Bi, H. Yang, The structural characterization and antigenicity of the S protein of SARS-CoV, *Dev. Reprod. Biol.* 1 (2003) 108–117.
- [51] L. Du, Y. He, Y. Zhou, S. Liu, B.-J. Zheng, S. Jiang, The spike protein of SARS-CoV — a target for vaccine and therapeutic development, *Nat. Rev. Microbiol.* 7 (2009) 226–236.
- [52] Y. He, Y. Zhou, H. Wu, B. Luo, J. Chen, W. Li, S. Jiang, Identification of immunodominant sites on the spike protein of severe acute respiratory syndrome (SARS) coronavirus: implication for developing SARS diagnostics and vaccines, *J. Immunol.* 173 (2004) 4050–4057.
- [53] J. Ignjatovic, S. Sapats, Identification of previously unknown antigenic epitopes on the S and N proteins of avian infectious bronchitis virus, *Arch. Virol.* 150 (2005) 1813–1831.
- [54] S.R. Weiss, S. Navas-Martin, Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus, *Microbiol. Mol. Biol. Rev.* 69 (2005) 635–664.
- [55] B. Hjelm, Epitope Mapping of Antibodies towards Human Protein Targets, PhD Thesis, KTH Royal Institute of Technology, 2011.
- [56] M.B. Irving, O. Pan, J.K. Scott, Random-peptide libraries and antigen-fragment libraries for epitope mapping and the development of vaccines and diagnostics, *Curr. Opin. Chem. Biol.* 5 (2001) 314–324.
- [57] X. Yang, X. Yu, An introduction to epitope prediction methods and software, *Rev. Med. Virol.* 19 (2009) 77–96.
- [58] A. Pomés, Relevant B cell epitopes in allergic disease, *Int. Arch. Allergy Immunol.* 152 (2010) 1–11.
- [59] J.M. Gershoni, A. Roitburd-Berman, D.D. Siman-Tov, N.T. Freund, Y. Weiss, Epitope mapping, *BioDrugs* 21 (2007) 145–156.
- [60] H.R. Ansari, G.P. Raghava, Identification of conformational B-cell Epitopes in an antigen from its primary sequence, *Immunorev.* 6 (2010) 6.
- [61] S.G.E. Marsh, E.D. Albert, W.F. Bodmer, R.E. Bontrop, B. Dupont, H.A. Erlich, M. Fernández-Viña, D.E. Geraghty, R. Holdsworth, C.K. Hurley, M. Lau, K.W. Lee, B. Mach, M. Maiers, W.R. Mayr, C.R. Müller, P. Parham, E.W. Petersdorf, T. Sasazuki, J.L. Strominger, A. Svejgaard, P.I. Terasaki, J.M. Tiercy, J. Trowsdale, Nomenclature for factors of the HLA system, 2010, *Tissue Antigens* 75 (2010) 291.
- [62] A.S. De Groot, D.S. Rivera, J.A. McMurry, S. Buus, W. Martin, Identification of immunogenic HLA-B7 “Achilles’ heel” epitopes within highly conserved regions of HIV, *Vaccine* 26 (2008) 3059–3071.
- [63] C.H. Lee, H. Koohy, In Silico Identification of Vaccine Targets for 2019-nCoV, 2020. F1000Research 9.
- [64] A. Sanchez-Mazas, HLA studies in the context of coronavirus outbreaks, *Swiss Med. Wkly.* 150 (2020).