International Journal of Research in Medical Sciences
Bency BJ et al. Int J Res Med Sci. 2020 Jul;8(7):2385-2394
www.msjonline.org
pISSN 2320-6071 | eISSN 2320-6012
DOI: http://dx.doi.org/10.18203/2320-6012.ijrms20202875

Original Research Article

Novel epitope based peptides for vaccine against SARS-CoV-2 virus: immunoinformatics with docking approach

Jesvin Bency B.[1]*, Mary Helen P. A.[2]

[1]Department of Biotechnology, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli, Tamil Nadu, India
[2]Department of Biotechnology, Malankara Catholic College, Kanyakumari, Tamil Nadu, India

Received: 19 April 2020
Accepted: 28 May 2020

***Correspondence:**
Jesvin Bency B.,
E-mail: bencyboaz@gmail.com

ABSTRACT

Background: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative viral strain for the contagious pandemic respiratory illness in humans which is a public health emergency of international concern. There is a desperate need for vaccines and antiviral strategies to combat the rapid spread of SARS-CoV-2 infection.
Methods: The present study based on computational methods has identified novel conserved cytotoxic T-lymphocyte epitopes as well as linear and discontinuous B-cell epitopes on the SARS-CoV-2 spike (S) protein. The predicted MHC class I and class II binding peptides were further checked for their antigenic scores, allergenicity, toxicity, digesting enzymes and mutation.
Results: A total of fourteen linear B-cell epitopes where GQSKRVDFC displayed the highest antigenicity-score and sixteen highly antigenic 100% conserved T-cell epitopes including the most potential vaccine candidates MHC class-I peptide KIADYNYKL and MHC class-II peptide VVFLHVTYV were identified. Furthermore, the potential peptide QGFSALEPL with high antigenicity score attached to larger number of human leukocyte antigen alleles. Docking analyses of the allele HLA-B*5201 predicted to be immunogenic to several of the selected epitopes revealed that the peptides engaged in strong binding with the HLA-B*5201 allele.
Conclusions: Collectively, this research provides novel candidates for epitope-based peptide vaccine design against SARS-CoV-2 infection.

Keywords: Epitope-based peptides, S protein, SARS-CoV-2 virus, Vaccine

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection is usually characterized by fever (83-99%), cough (59-82%), fatigue (44-70%), anorexia (40-84%), shortness of breath (31-40%), sputum production (28-33%) and myalgias (11-35%).[1] The threat of coronavirus pandemic has become very real as the global cases surpass 21,58,594 including 1,45,533 related deaths as of April 17, 2020. The causal agent was not identified until 7 Jan 2020; a new type of coronavirus was isolated by Chinese authority.[2] The genome of SARS-CoV-2 is single, positive-stranded RNA that is 29,811 nucleotides long.[3] No licensed vaccine or specific treatment is available till date. With the absence of effective treatment drugs and with the severity of the disease with high mortality rates, vaccination is considered to be one of the indispensable, potentially urgent strategies to tackle this novel virus infection. Earlier data suggest that the S protein is a major neutralization determinant in the SARS vaccine which can induce potent neutralizing antibodies to block SARS-CoV entry.[4] Hence, S protein is

considered to be a promising target for effective SARS-CoV-2 vaccine design. Further, the high mutation rate of the target protein may result in the escape of neutralizing antibodies against the virus.[5] Therefore, highly conserved targets that elicit both neutralizing antibodies as well as cellular immunity against the infection are essential for an effective vaccine development.[6]

Several past studies have proven that epitope-based vaccines could efficiently provoke protective immune responses.[7] Epitopes are antigenic determinants that represent the minimal immunogenic region of an antigenic protein and precisely elicit specific immune responses.[8] Peptide vaccines have several advantages over other types of vaccines such as conventional vaccines and newly developed DNA or cellular vaccines. Easy synthesis at less cost, relative safety and increased stability has been demonstrated through clinical studies. In addition, peptide vaccines have no limitation in target diseases and allergies.[9]

Although the genome sequences of several pathogens are available in several databases, their immunity based research associated with infection protection is still naive. The current study was based on immunoinformatics computational approach to screen the dominant immunogenic epitopes on SARS-CoV-2 spike proteins based on their sequences that could elicit protective humoral and cellular immune responses.

METHODS

Retrieval of SARS-CoV-2 S protein sequence and its 3D structure

Primary sequence of SARS-CoV-2 S protein (accession number: QII57161.1) was retrieved from NCBI database. The 3D structure of SARS-CoV-2 S protein (PDB ID: 6VXX) was obtained from Protein-Data-Bank.

Analysis of physical and chemical properties of SARS-CoV-2 S protein

The chemical and physical properties of SARS-CoV-2 S protein including GRAVY (Grand average of hydropathicity), molecular weight, half-life, amino acid atomic composition and stability index were analysed using Protparam tool. The secondary structure of SARS-CoV-2 S protein was analysed through PSIPRED workbench. Allergenicity of the query sequence was tested using AllerTOP v2.0 and antigenicity testing was carried out through Vaxijen v2.0. The transmembrane topology of S protein was examined with TMHMM tool. Prevalence of disulphide-bonds were examined through using DIANNA v1.1.

Linear B-cell epitope prediction

IEDB and BCPRED tools were used for B-cell epitope prediction. Criteria were set to 75% specificity. Epitopes that were visible on outer surface were tested for its antigenicity using Vaxijen 2.0.

Hydrophilicity, accessibility of surface and flexibility analysis were performed using Bepipred linear epitope prediction with Parker hydrophilicity prediction algorithms, Kolaskar and Tongaonkar antigenicity scale, Emini surface accessibility prediction and Karplus and Schulz flexibility prediction. Beta turns in polyprotein was forecasted using Chou and Fasman beta-turn prediction algorithm.

Discontinuous B-cell epitope prediction

Since discontinuous epitopes have higher dominant attributes over linear epitopes, discontinuous B-cell epitopes were also identified using DiscoTope with parameters set to ≥0.5 which indicates 90% specificity and 23% sensitivity.

Positional affirmation of predicted B-cell epitopes on S protein

The positions of predicted epitopes on 3D structure of S protein was observed via PepSurf and Pymol.

T-cell epitope prediction

With Propred-1 and Propred tools, Cytotoxic T-lymphocyte (CTL) epitopes of the S protein that bind to MHC class-I and MHC class-II alleles were predicted. For Propred-1 proteasome and immuno-proteasome filters with a threshold value of 5% were kept on.

Allergenicity and toxicity extrapolation

Features including digestion, mutation, toxicity, allergenicity, hydro and physiochemical of selected T-cell epitopes with higher antigenicity scores were checked using Vaxijen, Protein digest, AllergenFP 1.0, Aller Hunter and ToxinPred. For further analysis, only nontoxic and non-allergic epitopes were used.

Epitope conservancy analysis

SARS-CoV-2 S protein sequence of isolates from 10 different countries including Australia (QHR84449.1), Finland (QHU79173.2), India (QIA98583.1), Sweden (QIC53204.1), China-Wuhan (YP_009724390.1), China-Yunnan (QIA20044.1) USA (QIJ96493.1), Colombia (QIS30054.1), Italy (QIA98554.1), Japan(BCA87361.1), Vietnam (QIK50438.1) were subjected to multiple-sequence-alignment using Clustal Omega to analyse the identity and conservation of chosen epitopes.

The aligned files (.aln) were additionally utilized to make phylogenetic tree via MEGA7 software. The chosen epitopes were checked for their variability and conservation with IEDB conservation-analysis tool.

Structural modelling of epitopes

3D structures of three peptide epitopes (NFGAISSVL, QGFSALEPL, GGFNFSQIL) were modelled via PEPFOLD 3.5 server at RPBS MOBYL portal.

Molecular docking

The structure of the allele HLA-B*5201 that was predicted to be immunogenic to all the three of the selected epitopes was retrieved from Protein databank (PDB ID: 3W39). The structured peptide models were docked against HLA-B*5201 using the web server HPEPDOCK to analyse the binding potential with their docking scores. The position of the amino acid residues involved in the binding was found using Pymol software.

RESULTS

Structural analyses

SARS-CoV-2 S protein computed via Protparam demonstrates that it contained 1255 amino acids (aa) with a molecular weight of 139125.14 kDa. Theoretical isoelectric point (PI) of the subject protein was 5.56. Total number of negatively charged residues (Asp + Glu) were 115 and positively charged residues (Arg + Lys) were 99. The instability index (II) computed was found to be 32.42. This classifies the protein as stable. Aliphatic-index 82.73, which devotes a thought of proportional volume hold by aliphatic side chain and GRAVY value for protein sequence is -0.045. Half-life of protein was computed as 30 h for mammalian-reticulocytes, > 20 h for yeast, > 10 h for *Escherichia coli*. Total number of Carbon (C), Oxygen (O), Nitrogen (N), Hydrogen (H) and Sulfur (S) could be represented by the formula $C_{6252}H_{9593}N_{1609}O_{1871}S_{59}$.

PSIPRED analysis revealed the beta sheets, helixes and loops present in S protein as shown in Additional file 1: Figure S1. 3Dstructure of SARS-CoV-2 S protein is shown in Additional file 1: Figure S2. Furthermore in the target protein, DiANNA1.1 tool calculated 780 disulphides bond (S–S) positions. Antigenicity analysis of full-length protein showed antigenicity 0.4646 for S protein showing it as an expected antigen. TMHMM tool was used to check the transmembrane protein topology and it revealed that residues from 1 to 1213 were exposed on the surface, while residues from 1214 to 1236 were in the helix and residues from 1237 to 1273 were buried within the core-region of the S protein. The sequence was also confirmed to be a non-allergen.

Recognition of linear B-cell epitopes

A total of 99 B-cell epitopes were predicted among which 14 epitopes (Table 1) exposed on the surface of S protein with higher antigenicity scores were selected for further studies. Vaxijen 2.0 was used to compute antigenicity scores and TMHMM server was utilized to check their surface availability. Among these selected epitopes, 'GQSKRVDFC' predicted at position 1035 showed the highest antigenicity score.

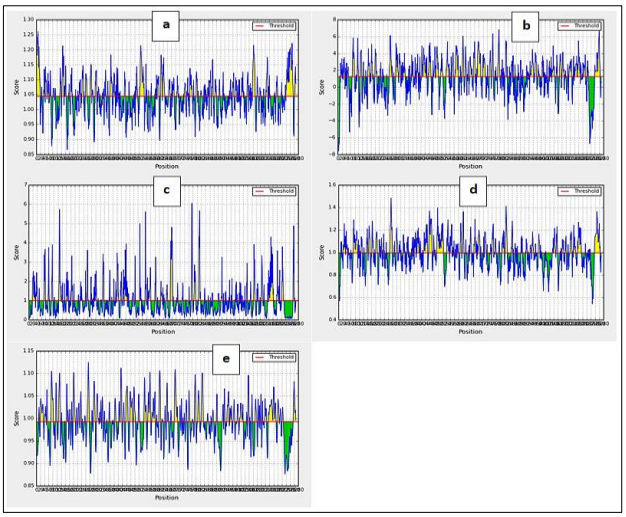

Figure 1: a) Prediction of antigenic determinants on SARS-CoV-2 spike proteins using Kolaskar and Tongaonkar antigenicity scale; b) Hydrophilicity prediction using Parker hydrophilicity; c) Surface accessibility analyses using Emini surface accessibility scale d beta turns analyses in structural polyprotein using Chou and Fasman beta turn prediction, e) flexibility analyses using Karplus and Schulz flexibility scale. X axis represents score, y axis the position.

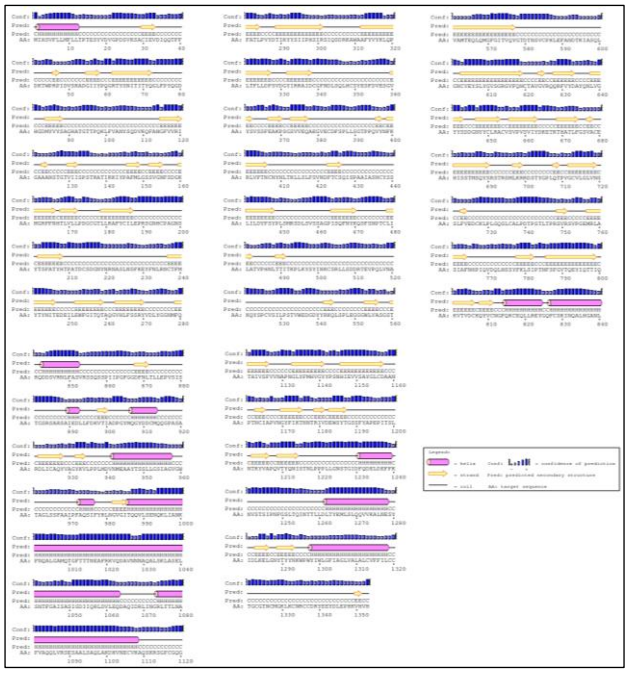

Figure S1: PSIPRED analysis of SARS-CoV-2 S protein. Helixes are cylindrical and colored pink, beta-strands are shown as arrows and coloured yellow, and random coil regions are black.

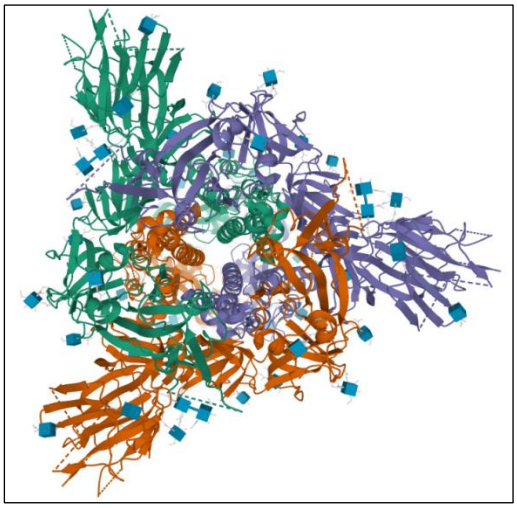

Figure S2: The 3D structure of the SARS-CoV-2 S protein.

Table 1: Potential linear B-cell epitopes on the surface of SARS-CoV-2 spike protein predicted using IEDB and BCPRED servers.

Position	Sequence	Antigenicity scores
177-189	MDLEGKQGNFKNL	1.2592
265-273	YYVGYLQPR	1.4692
331-340	NITNLCPFGE	1.2370
369-393	YNSASFSTFKCYGVSPTKLNDLCFT	1.4031
386-403	KLNDLCFTNVYADSFVIR	1.0694
404-426	GDEVRQIAPGQTGKIADYNYKLP	1.1017
506-519	YQPYRVVVLSFELLH	0.9711
580-599	QTLEILDITPCSFGGVSVIT	1.3356
584-601	ILDITPCSFGGVSVITPG	1.1031
718-730	FTISVTTEILPVS	1.0546
1035-1043	GQSKRVDFC	1.7790
1194-1202	NESLIDLQE	0.9767
1261-1269	SEPVLKGVK	0.9704
1262-1273	EPVLKGVKLHYT	1.4118

Kolaskar and Tongaonkar antigenicity measurements for the B-cell epitopes on S protein with threshold value at 1.041 estimated the antigenic tendency values to be 1.041 (average), 0.866 (minimum) and 1.261 (maximum). The results of Kolaskar and Tongaonkar analysis is shown in Figure 1a. To find the surface availability of the B-cell epitopes and its hydrophilicity, Parker-hydrophilicity with threshold value 1.238 and Emini surface accessibility prediction with threshold value 1.000 were set. Their results in visual representations are shown in Figure 1b and 1c respectively and the calculated values were 1.238 (average), - 7.629 (minimum), 7.743 (maximum); and 1.000 (average), 0.042 (minimum), 6.051 (maximum) respectively.

Chou and Fasman beta turn analysis algorithm threshold adjusted at 0.997, computed average: 0.997, minimum: 0.541 and maximum: 1.484 values. Chou and Fasman's results in graphical representation are shown in Figure 1d. The results indicate that regions from 251 to 257 amino acids are more disposed to persuade B turns in peptide structure. Karplus and schulz flexibility analysis revealed that the area from amino acids 854 to 860 sequence positions are highly versatile as shown in Figure 1e.

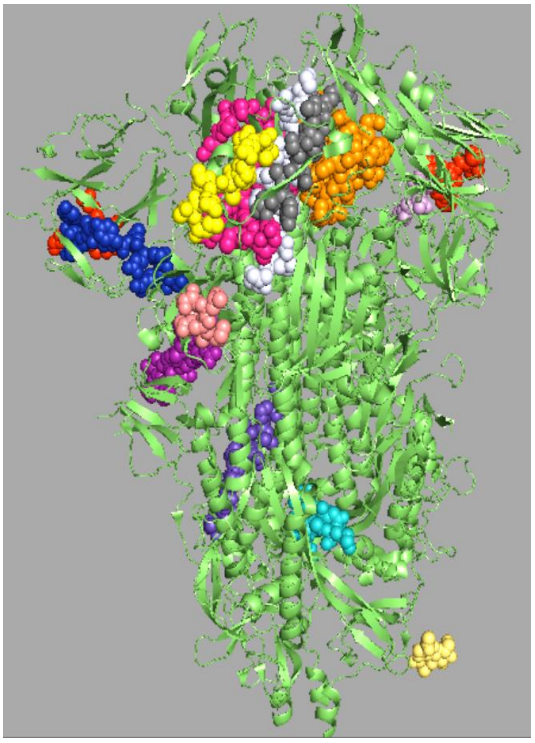

Figure 2: Representation of sites of linear B cells predicted epitopes (colourful spheres) on the crystal structure of SARS-CoV-2 envelope S protein (green cartoon).

The positions of selected linear B cell epitopes located on the surface of 3-D structure of SARS-CoV-2 S protein was predicted by using Pepsurf and generated using Pymol is shown in Figure 2.

Recognition of discontinuous B-cell epitopes

Discotope 2.0 server was used to calculate the surface availability of the discontinuous B-cell epitopes in terms of residue contact number and discotope scores. A total of 35 discontinuous epitopes were predicted at different exposed surface areas (Table 2). The positions of 8 predicted epitopes with high propensity and discotope scores on the surface of 3D structure of S protein shown in Figure 3.

Table 2: Discontinuous epitopes on SARS-CoV-2 envelope S protein as predicted by discotope 2.0 server.

Residue ID	Residue Names	Contact numbers	Propensity scores	Discotope scores
440	ASN	5	-1.222	-1.657
443	SER	19	0.432	-1.803
444	LYS	8	1.79	0.664
447	GLY	14	2.779	0.849
448	ASN	27	1.336	-1.922
449	TYR	3	0.877	0.431
454	ARG	14	-0.401	-1.965
489	TYR	0	-0.418	-0.37
490	PHE	9	-0.819	-1.76
491	PRO	11	0.123	-1.156
492	LEU	11	1.006	-0.375
493	GLN	11	1.042	-0.343
494	SER	14	0.753	-0.944
496	GLY	2	2.019	1.557
498	GLN	4	3.076	2.262
499	PRO	4	2.854	2.066
500	THR	1	4.501	3.868
501	ASN	24	3.882	0.676
503	VAL	2	0.217	-0.038
505	TYR	10	0.727	-0.507
558	LYS	0	-2.012	-1.781
704	SER	3	-1.384	-1.57
793	PRO	0	-1.559	-1.38
809	PRO	5	-1.511	-1.912
810	SER	4	1.115	0.526
812	PRO	3	0.186	-0.18
914	ASN	7	-0.803	-1.516
1140	PRO	8	-0.752	-1.586
1141	LEU	3	-0.71	-0.974
1142	GLN	7	0.439	-0.416
1143	PRO	6	0.445	-0.296
1144	GLU	4	0.586	0.059
1145	LEU	5	-0.203	-0.754
1146	ASP	6	1.013	0.206
1147	SER	5	-0.017	-0.59

Recognition of T-cell epitopes

Propred-I (47 MHC class-I alleles) and Propred (51 MHC class-II alleles) were utilized for prediction of T-cell epitopes on SARS-CoV-2 S protein. Additionally, antigenicity scores of the peptides were estimated using Vaxijen 2.0. Just 10 potential peptides were chosen for next processing on the basis of their antigenicity-score (Table 3).

A peptide which has capacity to attach with larger number of alleles is observed as most important peptide due to its potential to bring a powerful defence response. Among MHC class-I predicted epitopes, the peptide 'KIADYNYKL' indicated higher antigenicity score 1.6639 attaching with number of alleles including HLA-A2, HLA-A*0201, HLA-A*0205, HLA-A*1101, HLA-A24, HLA-A3, HLA-A*3101, HLA-B*2705, HLA-B*3501, HLA-B*3902, HLA-B*0702.

Propred, predicted T-cell peptides, which can interact with MHC class-II alleles. Subsequent screening was done with Vaxijen 2.0 and just 6 high scoring epitopes were chosen for further analyses (Table 4). The peptide 'VVFLHVTYV' was considered more antigenic for its higher antigenicity score 1.5122 and it demonstrated virtual attachment with many alleles DRB1_0101, DRB1_0102, DRB1_0404, DRB1_0408, DRB1_0410, DRB1_0423.

Eminent features profiling of selected T cells epitopes

Peptides digested by fewer enzymes are highly stable and more favourable vaccine candidates. Peptides digesting enzymes were predicted through Protein digest server. Allergen FP 1.0 was used for allergenicity prediction of epitopes.

ToxinPred was utilized for toxicity prediction of chosen epitopes. Mutations, hydropathicity, hydrophilicity, hydrophobicity, and charge were also checked with their results in Table 5.

Conservation analyses of selected epitopes

Sequences of SARS-CoV-2 S proteins from 10 different countries subjected to multiple-sequence-alignment showed that all the chosen epitopes were conserved in all sequences utilized for analysis as shown in Additional file 1: Figure S3.

A phylogenetic tree was also created to indicate the evolutionary relationship of SARS-CoV-2 spike proteins from 10 distinct countries as shown in Figure 4.

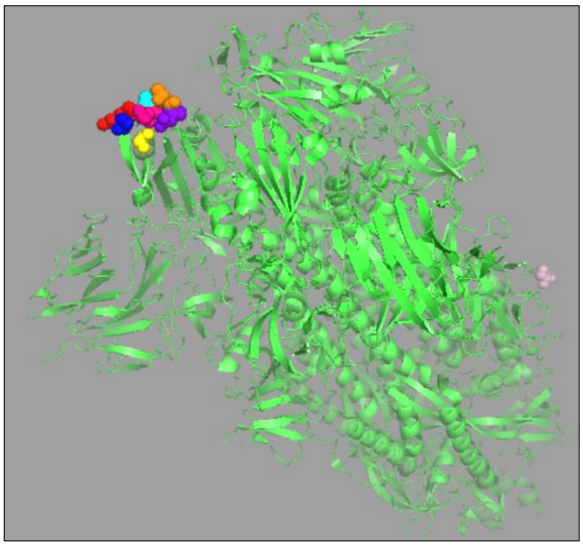

Figure 3: Representation of sites of B cells discontinues epitopes with higher propensity and discotope scores predicted through DISCOTOPE 2.0 Server on the crystal structure of SARS-CoV-2 envelope S protein as cartoon representation (green).

Table 3: MHC class-I allele binding peptides in SARS-CoV-2 S protein predicted via Propred-I with their antigenicity scores.

At Position	Peptides	MHC class-I alleles	Vaxijen score
417	KIADYNYKL	HLA-A2, HLA-A*0201, HLA-A*0205, HLA-A*1101, HLA-A24, HLA-A3, HLA-A*3101, HLA-B*2705, HLA-B*3501, HLA-B*3902, HLA-B*0702	1.6639
379	CYGVSPTKL	HLA-A24, HLA-A*3302,HLA-Cw*0401,HLA-Cw*0702,MHC-Kd	1.4263
510	VVVLSFELL	HLA-A*0205,HLA-A*1101,HLA-A24,HLA-A*3101,HLA-A*3302, HLA-A68.1,HLA-B*3701,HLA-B*51,HLA-B7,HLA-Cw*0301,MHC-Db,MHC-Db revised,HC-Kb, MHC-Kd	1.0909
181	GKQGNFKNL	HLA-A20 Cattle,HLA-B*3902, HLA-Cw*0301,MHC-Db,MHC-Db revised,MHC-Dd,MHC-Kb,	1.0607
109	TLDSKTQSL	HLA-A1,HLA-A2,HLA-A*0201, HLA-A3,HLA-A20 Cattle, HLA-A2.1,HLA-B*2705,HLA-B*3801, HLA-B*3901,HLA-B8,MHC-Dd,	1.0685
969	NFGAISSVL	HLA-A24,HLA-B*3801,HLA-B*5201,HLA-Cw*0401,HLA-Cw*0702,MHC-Kd	0.9894
218	QGFSALEPL	HLA-B14,HLA-B*3901,HLA-B40, HLA-B*5101,HLA-B*5102,HLA-B*5103,HLA-B*5201,HLA-B*5401,HLA-B61,HLA-Cw*0401, MHC-Dd,MHC-Kb	0.8462
798	GGFNFSQIL	HLA-A2.1,HLA-B14,HLA-B*3902, HLA-B40,HLA-B*5101,HLA-B*5102,HLA-B*5103,HLA-B*5201,HLA-B61,HLA-Cw*0602, MHC-Dd,MHC-Kb	0.7967
1016	AEIRASANL	HLA-B*3701,HLA-B40,HLA-B*4403,HLA-B60,HLA-B61,HLA-Cw*0301,MHC-Kk,MHC-Ld	0.7082
168	FEYVSQPFL	HLA-A*0201,HLA-A*0205,HLA-A2.1,HLA-B*2702,HLA-B*2705, HLA-B*3701,HLA-B40,HLA-B*4403,HLA-B*5301,HLA-B*51, HLA-B60,HLA-B61,HLA-Cw*0301, MHC-Kb,MHC-Kk,MHC-Ld	0.6324

Table 4: MHC class-II allele binding peptides of SARS-CoV-2 S protein predicted via propred with their antigenicity scores.

At position	Peptides	MHC class-II alleles	Vaxijen score
1059	VVFLHVTYV	DRB1_0101, DRB1_0102, DRB1_0404, DRB1_0408, DRB1_0410, DRB1_0423	1.5122
230	IGINITRFQ	DRB1_0402, DRB1_1102, DRB1_1114, DRB1_1121, DRB1_1304, DRB1_1322, DRB1_1323	1.3386
485	FNCYFPLQS	DRB1_0801, DRB1_0802, DRB1_0813, DRB1_0817, DRB1_1502	1.0649
200	FKIYSKHTP	DRB1_0801, DRB1_0802, DRB1_0804, DRB1_0813, DRB1_1120, DRB1_1302	0.9886
1	FVFLVLLPL	DRB1_0101, DRB1_0102, DRB1_0701, DRB1_0703, DRB1_0817, DRB1_1101, DRB1_1104, DRB1_1106, DRB1_1128, DRB1_1305, DRB1_1307, DRB1_1311, DRB1_1321, DRB1_1502, DRB5_0101	0.8601
4	LVLLPLVSS	DRB1_0804, DRB1_1102, DRB1_1104, DRB1_1106, DRB1_1114, DRB1_1121, DRB1_1301, DRB1_1307, DRB1_1311, DRB1_1322, DRB1_1323,DRB1_1327, DRB1_1328	0.6523

The epitope-conservancy study through IEDB epitope conservancy analysis tool shows that all of the selected B-cell and T-cell (MHC class-I and II) epitopes have 100% identity and conserved in all isolates of distinct countries.

Docking study of selected epitopes with HLA allele

3D structures of three (NFGAISSVL, QGFSALEPL, GGFNFSQIL) MHC class-I attaching peptides were predicted via PEPFOLD 3.5 and the best model in each

was used for the study (Additional file 1: Figure S4). The allele HLA-B*5201 that was predicted to be immunogenic to all of the three selected epitopes was docked against NFGAISSVL, QGFSALEPL, GGFNFSQIL MHC class-I binding peptides. A very negative docking score corresponds to a strong binding. All the three peptides exhibited strong binding with the HLA-B*5201 allele with negative docking scores (Table 6). The amino acid residues involved in the hydrogen bonding is listed in Table 6. Molecular docking of HLA-B*5201 allele with MHC class-I binding peptides is shown in Figure 5a,5c,5e and the hydrogen bonding with amino acid residues is shown in Figure 5b,5d,5f.

Table 5: Digestion enzymes, mutation, toxicity, allergenicity, hydro and physiochemical profiling of selected MHC class I and class II binding peptides in SARS-CoV-2 S protein.

MHC class-I binding peptides									
Allergenicity	**Peptide Sequence**	**Mutation**	**Toxicity**	**Digesting enzymes**	**Hydrophobicity**	**Hydrop -hilicity**	**Charge**	**pI**	**Mol. wt.**
A	KIADYN YKLL	NM	Non-Toxin	Asp-N endopeptidase, Chymotrypsin Trypsin	-0.22	0.06	1.00	8.76	1127.43
NA	CYGVSP TKLL	NM	Non-Toxin	Chymotrypsin Trypsin	-0.04	-0.41	1.00	8.54	967.27
A	VVVLSF ELL	NM	Non-Toxin	Chymotrypsin Staphylococcal peptidase I	0.33	-1.01	-1.00	4.00	1018.40
NA	GKQGN FKNL	NM	Non-Toxin	Chymotrypsin Trypsin	-0.30	0.26	2.00	10.02	1005.28
A	TLDSKT QSL	NM	Non-Toxin	Asp-N endopeptidase, Trypsin	-0.26	0.27	0.00	6.19	992.22
NA	NFGAIS SVL	NM	Non-Toxin	Chymotrypsin	0.18	-0.81	0.00	5.88	907.16
NA	QGFSAL EPL	NM	Non-Toxin	Chymotrypsin Staphylococcal peptidase I	0.05	-0.34	-1.00	4.00	961.21
NA	GGFNFS QIL	NM	Non-Toxin	Chymotrypsin	0.13	-0.88	0.00	5.88	982.24
A	AEIRAS ANL	NM	Non-Toxin	Clostripain Staphylococcal peptidase I Trypsin	-0.14	0.16	0.00	6.36	944.17
NA	FEYVSQ PFL	NM	Non-Toxin	Chymotrypsin Staphylococcal peptidase I	0.07	-0.79	-1.00	4.00	1129.40
MHC class-II binding peptides									
NA	VVFLH VTYV	NM	Non-Toxin	Chymotrypsin	0.30	-1.50	0.50	7.09	1076.44
NA	IGINITR FQ	NM	Non-Toxin	Chymotrypsin Trypsin Clostripain	-0.03	-0.54	1.00	10.11	1061.39
A	FNCYFP LQS	NM	Non-Toxin	Chymotrypsin	0.02	-1.04	0.00	5.84	1118.39
NA	FKIYSK HTP	NM	Non-Toxin	Chymotrypsin Trypsin	-0.19	-0.13	2.50	9.72	1120.44
NA	FVFLVL LPL	NM	Non-Toxin	Chymotrypsin	0.48	-1.69	0.00	5.88	1060.53
A	LVLLPL VSS	NM	Non-Toxin	*Proline-endopeptidase*	0.29	-1.07	0.00	5.88	940.33

DISCUSSION

Epitope-based vaccination strategy poses decreased biohazard risk, could be rationally engineered and optimized structurally eliciting strong immunity.[10,11] Generation of conserved immunodominant epitopes, lack of infectious potential and its stability makes epitope-based vaccines effective against a wide range of infectious agents, including parasitic, bacterial, fungal, and viral infections.[12] The basic premise of epitope-based approach to vaccine development is that, in certain cases, the responses induced by the natural immunogen are not optimal, and can be improved upon by isolation or optimization of specific components of the response.[13] Regardless of the datum that antibody memory response can be effortlessly avoided by antigenic drift over the period of time, T-cell immunity normally elicits enduring immunity.[14] In the clinical trials of various cancers, peptide vaccine has entered phases I and II with satisfactory and promising clinical outcomes.[15,16]

Spike proteins are considered to be potential target for vaccine design because of their ability to induce a faster and longer-term mucosal immune response than that of

the other proteins.[17] For this reason, spike protein has gained much popularity among the researchers in vaccine design.[18,19]

Figure S3: Multiple sequence alignment showing conservation of the S protein of SARS-CoV2 isolated from 10 different countries (Australia, Finland, India, Sweden, China-Wuhan, China-Yunnan, USA, Colombia, Italy, Japan and Vietnam).

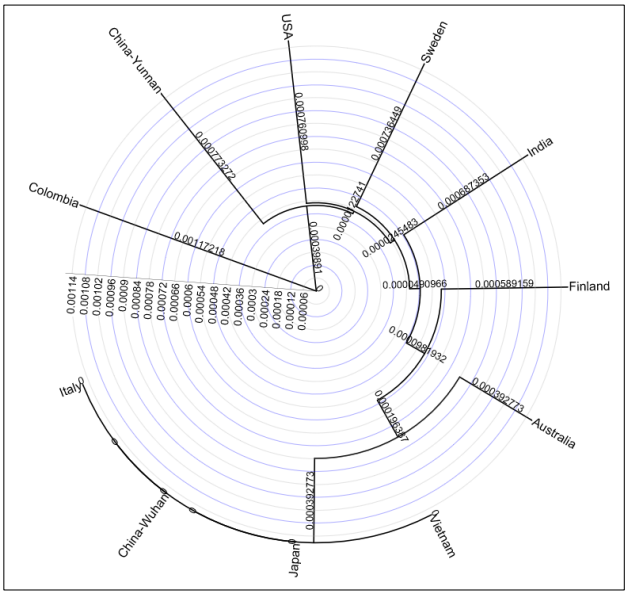

Figure 4: Phylogenetic tree illustrating evolutionary relationships among SARS-CoV-2 S proteins from 10 different countries (Australia, Finland, India, Sweden, China-Wuhan, China-Yunnan, USA, Colombia, Italy, Japan and Vietnam).

The identification of MHC binding peptides for consideration as potential T-cell epitopes has application in peptide vaccine design and immunotherapy.[20] During T cell activation, peptides derived from protein antigens are presented by MHC molecules. This process is accomplished by the intracellular fragmentation of protein antigens, followed by binding of derived peptides to HLA molecules and subsequent presentation on antigen presenting cell (APC) surface, for recognition by T cell receptors.[21] In the present study, well conserved ten T-cell epitopes that could bind to MHC class I alleles and 6 epitope peptides that bind to MHC class II alleles were predicted. The epitopes were also subjected to allergenicity analyses and most of them were reported to be non-allergens.

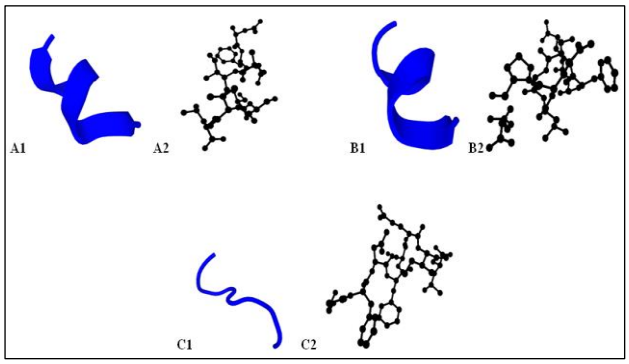

Figure S4: 3D (A1, B1,C1) and stick with balls structures (A2,B2,C2) representation of selected MHC class-I alleles binding peptides.

Table 6: Molecular docking results of HLA-B*5201 allele with MHC class-I binding peptides of SARS-CoV-2 S protein.

MHC class-I binding peptides	Interacting residues	Docking scores
NFGAISSVL	Ser-57, Tyr-26, Asp-31, Arg-240, Gln-33, Tyr-63, Phe-56, Tyr-28, Arg-49,Glu-54, Lys-41, Tyr-78, Thr-71	-167.580
QGFSALEPL	Phe-56, Tyr-63, Tyr-28, Gly-238, Asp-31, Asp-30, Arg-240, Ser-20, Ser-230, Arg-49, Glu-54, Lys-41, Thr-71, Thr-78	-160.193
GGFNFSQIL	Glu-233, Tyr-26, Thr-234, Arg-49, Gln-33, Asp-31, Arg-240	-164.498

Previous researches report that B-cell epitopes, as the least immune unit, are strong enough to elicit a potent humoral immune response without harmful side effects to the human body, therefore, the identification of B-cell epitopes is important in understanding the

immunotherapeutic, and immunodetection, characteristics of a vaccine antigen.[22]

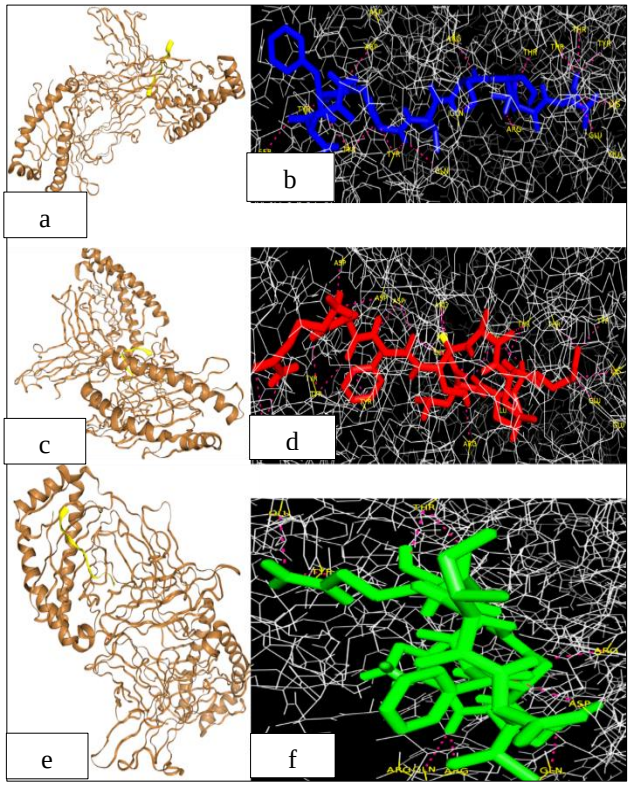

Figure 5: (a,c,e) Cartoon representation of docking between human HLA-B*5201 protein (shown in brown) and MHC class-I alleles binding peptides (shown in yellow). (b,d,f) Binding interaction of the epitope residues and HLA-B*5201 through the hydrogen bonds. The lines (shown in grey) represent HLA-B*5201, epitopes as red, blue and green sticks, hydrogen bonds as hot pink dotted lines and residue names are labelled in yellow.

Unlike the T-cell epitopes, the majority of the functional B-cell epitopes are discontinuous non-linear epitopes having 3D-conformational structures.[23] In this study, a total of fourteen linear B-cell epitopes were detected among which GQSKRVDFC exhibited the highest antigenicity score. A similar study has already reported various other epitope based peptides on the spike protein of the same virus.[24]

To completely describe and characterize epitope variability, measures of identity and conservancy are typically utilized. Identity is the extent to which two amino acid sequences are invariant while conservancy is the fraction of protein sequences that contain the epitope considered at or above a specified level of identity.[25] In the present study, the epitope-conservancy study through IEDB epitope conservancy analysis tool showed that all of selected B-cell and T-cell (MHC class-I and II) epitopes had 100% identity and conserved in all the isolates of 10 distinct countries analysed.

Molecular modelling utilizes detailed knowledge of the crystal structure of MHC molecules and of protein–peptide interactions.[26,27] Molecular modelling provides a detailed insight into specific 3-D structures and interactions. Docking work can be extended to the prediction of peptide binding affinities using free energy scoring functions.[28] A very negative docking score corresponds to a strong binding. All the three peptides docked in our study exhibited strong binding with the HLA-B*5201 allele with negative docking scores. The docking analysis concluded that our selected epitopes could bind competently to human HLA.

CONCLUSION

In the current study, a computational approach was adopted to identify and screen immunogenic surface-exposed peptides in SARS-CoV2 S protein that can serve as vaccine candidates. The results in this study have provided 14 linear B-cell epitopes, 10 MHC class I binding T-cell peptides and 6 MHC class II binding T-cell peptides that are all conserved potential epitopes that might serve as potential peptide vaccine candidates for SARS-CoV-2 vaccine with high specificity which is a mandatory requirement in the current global coronavirus threat scenario.

ACKNOWLEDGEMENTS

Authors would like to thank staff members of Biotechnology Department, Malankara Catholic College for their encouragement throughout this work. Authors are also thankful to Rev. Fr. Jose Bright, Correspondent/Secretary and Dr. J. Thampi Thanka Kumaran, Principal, Malankara Catholic College, Mariagiri for their constant encouragement and support.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES

1. Khot W, Nadkar M. The 2019 Novel Coronavirus Outbreak - A Global Threat. J Assoc Physicians India. 2020;68:67-71.
2. Zheng J. SARS-CoV-2: an Emerging Coronavirus that Causes a Global Threat. Int J Biol Sci. 2020;16(10):1678-85.
3. Sah R, Rodriguez-Morales A, Jha R, Chu D, Gu H, Peiris M, et al. Microbiol Resour Announc. 2020;9(11):e00169-20.
4. He Y, Zhou Y, Siddiqui P, Jiang S. Inactivated SARS-CoV vaccine elicits high titers of spike protein-specific antibodies that block receptor binding and virus entry. Biochem Biophys Res Commun. 2004;325(2):445-52.
5. Cann A. Genomes. In: Principles of Molecular Virology. 5th ed. UK: Academic Press; 2012.

6. Qamar T, Saleem M, Ashfaq S. Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: an immune-informatics study. J Transl Med. 2019;17:362.
7. Dimitrov I, Flower D, Doytchinova I. AllerTOP-a server for in silico prediction of allergens. BMC bioinformatics. Bio Med Central. 2013;14(6):S4.
8. Buchan D, Minneci F, Nugent T, Bryson K, Jones D. Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res. 2013;41:W349-57.
9. Yang H, Kim D. Peptide and Protein Vaccines. Curr Opin Immunol. 2003;4:461-70.
10. Tung T, Zacharias A, Arun K, Raúl G, Stig T, Melanie S, et al. The COVID-19 vaccine development landscape. Nat Rev Drug Discov. 2020.
11. Hajissa K, Zakaria R, Suppian R, Mohamed Z. Epitope-based vaccine as a universal vaccination strategy against Toxoplasma gondii infection: A mini-review. JAVAR. 2019;6(2):174-82.
12. Patronov A, Doytchinova I. T-cell epitope vaccine design by immunoinformatics. Open Biol. 2013;3(1):1201-39.
13. Sette A, Fikes J. Epitope-based vaccines: an update on epitope identification, vaccine design and delivery. Curr Opin Immunol. 2003;15(4):461-70.
14. Galvani AP, Ndeffo-Mbah ML, Wenzel N, Childs JE. Ebola vaccination: if not now, when?. Ann Int Med. 2014;161(10):749-50.
15. Naz R, Dabir P. Peptide vaccines against cancer, infectious diseases, and conception. Front Biosci J Virtual Lib. 2006;12:1833-44.
16. Ma C, Li Y, Wang L. Intranasal vaccination with recombinant receptor-binding domain of MERS-CoV spike protein induces much stronger local mucosal immune responses than subcutaneous immunization: Implication for designing novel mucosal MERS vaccines. Vaccine. 2014;32(18):2100-8.
17. Yang Z, Kong W, Huang Y. A DNA vaccine induces SARS coronavirus neutralization and protective immunity in mice. Nature. 2004; 428(6982):561-4.
18. Agnihothram S, Gopal R, Yount B. Platform strategies for rapid response against emerging coronaviruses: MERS-CoV serologic and antigenic relationships in vaccine design. J Infect Dis. 2013;209(7):995-1006.
19. Zhao B, Sakharkar K, Lim C, Kangueane P, Sakharkar M. MHC-Peptide binding prediction for epitope based vaccine design. IJIB Design. 2007;1(2):127.
20. Pamer E, Cresswell P. Mechanisms of MHC class I--restricted antigen processing Annu Rev Immunol. 1998;16:323-58.
21. Varshney A, Wang X, Cook E, Dutta K, Scharff M, Goger M. Generation, characterization, and epitope mapping of neutralizing and protective monoclonal antibodies against staphylococcal enterotoxin B-induced lethal shock. J Biol Chem. 2011;286(11):9737-47.
22. Ansari H, Raghava G. Identification of conformational B-cell epitopes in an antigen from its primary sequence. Immunome Res. 2010;6:1-9.
23. Bhattacharya M, Sharma A, Patra P, Ghosh P, Sharma G, Bidhan C, et al. Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics Approach J Med Virol. 2020;2020.
24. Bui H, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC Bioinform. 2007;8:361.
25. Rognan D, Scapozza L, Folkers G, Daser. A Molecular dynamics simulation of MHC-peptide complexes as a tool for predicting potential T cell epitopes. Biochem. 1994;33:11476-85.
26. Altuvia Y, Sette A, Sidney J, Southwood S, Margalit H. A structure-based algorithm to predict potential binding peptides to MHC molecules with hydrophobic binding pockets, Hum. Immunol. 1997;58:1-11.
27. Brusic V, Bajic V, Petrovsky N. Computational methods for prediction of T-cell epitopes-a framework for modelling, testing, and applications. Methods. 2004;34:436-43.
28. Schueler-Furman O, Altuvia Y, Sette A, Margalit H. Structure-based prediction of binding peptides to MHC class I molecules: application to a broad range of MHC alleles. Protein Sci. 2000;9:1838-46.

Cite this article as: Bency BJ, Helen PAM. Novel epitope based peptides for vaccine against SARS-CoV-2 virus: immunoinformatics with docking approach. Int J Res Med Sci 2020;8:2385-94.