Vaccine Research
ISSN: 2423-4923
eISSN: 2383-2819

In Silico Design of Multiepitope Vaccine Candidate Against SARS CoV and CoV2

Krupanidhi Sreerama*

Department of Biotechnology, Vignan's Foundation for Science, Technology and Research, Vadlamudi, 522213 Andhra Pradesh, India

ARTICLE INFO

Research Article

VacRes, 2020
Vol. 7, No.2, 61- 73
Received: Dec 10, 2020
Accepted: May 09, 2021
Pasteur Institute of Iran

***Corresponding Author:**
Department of Biotechnology, Vignan's Foundation for Science, Technology and Research, Vadlamudi, 522213 Andhra Pradesh

Email: krupanidhi.srirama@gmail.com
Tel/Fax: +91-9440984668

KEYWORDS: COVID-19; Vaccination,Vaccine ethics, Vaccine hesitancy

ABSTRACT

Introduction: The rampant widespread of COVID-19 crossing continents in the year 2020 is primarily because of its unique spike protein architecture. Hence, the spike protein sequence similarities among the representatives of coronaviruses were evaluated so as to arrive at possible conservancy in its epitopes. **Methods:** Multiple sequence alignment and molecular phylogenetic analysis were done using MEGA software version X. Clustal omega open software was adapted to develop Percent Identity Matrix of spike proteins. Online IEDB tools were used to explore linear epitopes within the RBD region of spike protein of SARS CoV2 isolate of Wuhan-Hu-1 and their conservancy across the species of the chosen CoVs. **Results:** The constructed phylogenetic tree showed a primary cluster between SARS CoV2s of Wuhan and Bangladesh strains. The branch length of this primary cluster reflected their recency in emergence. Further, this primary cluster developed as an offshoot of yet another primary cluster between SARS CoV and bat SARS CoV. All betacoronaviruses grouped as one tertiary cluster, wherein MERS CoV formed as an independent offshoot and its branch length reflected that it is phylogenetically older. Both SARS CoV2s are the closest relatives to SARS CoV and Bat SARS CoV of China, and hence the similar pattern was confirmed through MEGA analysis. Ten linear epitopes were identified within the RBD region of the spike protein for the population of the State of Andhra Pradesh among Indian Asians based on their HLA haplotype diversity. Further, Conservancy Analysis of spike protein suggested that SARS CoV2 and SARS CoV shared 53% predicted epitopes. The physicochemical features of the envisaged polytope indicated the presence of 12.96% charged residues with instability index showing stable nature and more hydrophilicity as revealed through GRAVY values indicating that residues of polytope possibly interact well in aqueous environment. The secondary structure of the envisaged polytope showed predominantly coils, moderate number beta pleated sheets and α-helices with 41.2% residues in favoured region and 44.7% in allowed regions of Ramachandran Plot. **Conclusion:** The derived predicted scores of T-cell and B-cell immunogenicity, MHC class I binding, non-toxic and non-allergic due to the identified multi-epitope confirmed to be antigenic and elicit both T-cell and B-cell immune responses. Population coverage tool (IEDB) showed an adequate fraction of individuals predicted to respond to a given set of identified epitopes with known MHC restrictions in Indian Asian population.

Citation:
Sreerama K. *In silico* design of multiepitope vaccine candidate against SARS CoV and CoV2. vacres. 2020; 7 (2) :61-73
URL: http://vacres.pasteur.ac.ir/article-1-250-en.html

INTRODUCTION

The recent pandemic of COVID-19 is primarily because of the unique protein architecture of Severe Acute Respiratory Syndrome Coronavirus (SARS CoV2) which rampantly made it spread across nations [1-4]. As of 26th January, 2021, a total of 546,267 new cases of coronavirus were reported (COVID-19). There were possibly 100 million sufferers globally due to COVID-19, of which ~ 2.1 million deaths were recorded. This data possibly reflects that the healthcare measures other than prophylactic vaccines do not possibly protect the human population. (https://www.statista.com/statistics/1103046/new-coronavirus-covid19-cases-number-worldwide-by-day/).

[DOI: 10.52547/vacres.7.2.61]

There are eleven proteins transcribed by SARS CoV2 +sense RNA genome [4, 5].Of which, spike protein plays an important role in spreading infection as it anchors angiotensin converting enzyme 2 (ACE2) receptors of alveolar epithelial cells and /or luminal surface of intestinal epithelial cells and facilitates membrane fusion followed by endocytosis [6-8]. Spike proteins possess two sub units namely S1 and S2. S1 domain of the spike protein anchors host ACE2 receptor through Receptor binding domain (RBD), whereas S2 mediates membrane fusion [3,9]. It is reported that the amino acid sequence of RBD of spike protein of SARS CoV2 is ~74% similar to SARS CoV [3,9]. Comparisons of protein sequences allowed tracing possible evolutionary affinities, structural similarities, functional domains, immune elicitation, cross-protection, virus reservoirs and clues to discriminate the potential hosts [10-15]. Whereas the nucleotide-based observations of the phylogeny are having limited scope compared to the protein residue-based sequence analyses primarily in relation to the attributes such as the evaluation of antigenicity and immune elicitation. The structural proteins such as spike proteins and envelope proteins are used to evaluate the antigenicity as they anchor to receptors rather than non-structural proteins. The envelope proteins of a few representatives CoVs of pet and domesticated animals with which human beings normally dwell were found to be 100 % identical between CoVs of pangolin and bat with varying similarities ranging from 20%-95% with the rest of the nine specimens reported by Tilocca et al [15,16] and suggested that the past contact elicit passive immunity and cross-protection against SARS CoV2. Further, as the host human beings are habituated to different geo-climatic and socio-cultural environments with varied life styles, there ought to have been the potential for the aforementioned viruses to develop into various clades, a few of them are authenticated as the most virulent [17]. The compiled data from GISAID shown 400,508full genomes of SARS CoV2, of which the major clades are found to be S, L, V, G, GR,GH, GV etc., (gisaid.org/spike) and they are reported occurring across continents and they are primarily classified based on spike protein sequence variation. Of the full genomes delineated by GISAID, it is reported that the variation is primarily due to mutations in the spike protein of SARS CoV2 [17]. Interestingly, the South Indian population is heterogeneous and its haplotypes of HLA A, B and DRB1 loci in five linguistic groups are reported [18]. This data on HLA haplotype diversity of South Indian population is used in the present in-silico analysis considering the epitopes of spike protein of SARS CoV2 to display on HLA so as to predict T-cell and B-cell immune responses.

The multi-epitope vaccine proposed in this article is aimed to predict the immune elicitation against antigenic targets of spike protein. It is reported that the conjugated polytope vaccine as a dendrimer constructed against HIV-1 elicited higher Th1 response [19]. In yet another instance, a polytope prediction from four conserved envelope protein of dengue virus serotypes was recognized by neutralizing IgGs [20]. Therefore, in this context, it is pertinent to build the molecular affinity based on spike protein sequences among the members of the family Coronaviridae and explore the possible multi-epitope-based vaccine candidate for the population of Andhra Pradesh among Indian Asians.

MATERIALS AND METHODS

Protein Sequences Retrieval

The spike protein sequences of seven CoVs enlisted in Table 1 were retrieved in FASTA format from GeneBank database. These full length sequences were all saved in a text document. The number of amino acids, length of conserved region, receptor binding domain and corresponding host receptor were all noted.

Table 1. Spike proteins of the representative coronaviruses extracted from GenBank database along with the length of conserved region, RBD and corresponding receptors in human host are shown.

NCBI Protein Id	Source	No. of Amino acids	Length of conserved S2 region	RBD located in S1 region	Receptor in human host
ACU31032.1	Bat SARS CoV, China	1241	634-1238	321-554	ACE2
ABA02260.1	SARS CoV, China	1255	648-1252	317-569	ACE2
AMO03401.1	MERS CoV, Egypt	1353	761-1352	383-502	CD26 (DPP4)
YP_009724390.1	SARS CoV2 isolate Wuhan-Hu-1	1273	662-1270	319-541	ACE2
QLF97795.1	SARS-CoV-2, Bangladesh	1273	615-1175	420-515	ACE2
AFV53148.1	Human NL63 CoV	1356	724-1355	-	-
NP_073551.1	Human CoV 229E	1173	543-1172	-	-

[DOI: 10.52547/vacres.7.2.61]

Multiple Sequence Alignment

The Molecular Evolutionary Genetics Analysis (MEGA) software version X [21, 22] was employed for aligning the retrieved spike protein sequences of CoVs. The spike protein sequences saved in FASTA format in the text document were retrieved using the option in MEGA viz., "retrieve sequence from a file". Protein sequences were appeared in MX Alignment explorer. The terminals of the same were trimmed and subjected by clicking "Align using the MUSCLE algorithm" with default setting which yielded the multiple sequence alignment (MSA) with conserved residues and gaps. The obtained MSA was exported in MEGA format and kept in a folder for the subsequent use in phylogeny.

Phylogenetic Analysis

The test of phylogeny was done using Maximum Likelihood statistical method with 1050 bootstrap replications in GTR (General Time Reversible) model, Tamura-Nei Model of substitution and an option of complete deletion of gaps [21, 22]. The resultant Tree was saved as PNG file (Fig. 1A). The Percent Identity Matrix of spike protein sequences of the seven Coronaviridae strains was evaluated by using online Clustal 2.1 (http://www.ebi.ac.uk/Tools/msa/clustalo/) to evaluate their pair-wise similarity/identity.

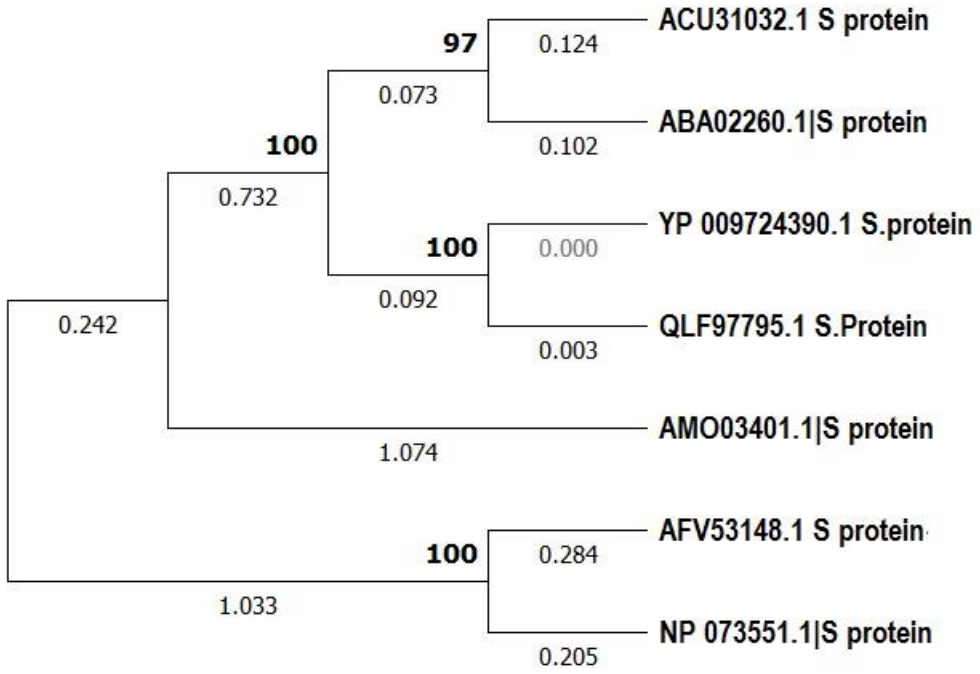

Fig. 1A. Phylogenetic tree of representative coronaviruses constructed with spike protein amino acid sequences using Maximum Likelihood Method and JTT matrix-based model in MEGA X [21].This analysis involved 7 amino acid sequences. All positions containing gaps and missing data were eliminated (complete deletion option). There were a total of 1023 positions in the final dataset. NCBI accession numbers of representative spike proteins along with the source as shown in the figure are given in Table 1: ACU31032.1 BAT SARS CoV; ABA02260.1 SARS CoV; YP_009724390.1 SARS CoV2 (Wuhan); QLF97795.1 SARS CoV2 (Bangladesh); AMO03401.1 MERS CoV; AFV53148.1 NL63 CoV; NP_073551.1 229E CoV.

Epitope Prediction

The Immune Epitope Database (IEDB) online software tool (www.iedb.org) was employed to predict the epitopes on YP_OO9724390.1 (spike protein of SARS CoV2 isolate Wuhan-Hu-1). In addition, HLA I haplotypes of the population of Andhra Pradesh, India were submitted to IEDB software for the analysis of MHC Class I processing epitope predictions. There were 68266 hits appeared as an output with 8 to 14 sequence peptides as possible epitopes with the rank varied between 0.01 to 100. This huge data was transferred to the Excel sheet and classified this data by filtering using the maximum rank namely 100 and 8 and 9 lengths of peptide sequences. This filtering resulted in 276 hits. Further, the search was done manually for the consensus 'start' and 'end' epitope sequences in receptor binding domain (RBD) of spike proteins (Table 1) restricted to five MHC Class I haplotypes of the population of Andhra Pradesh, India and they were tabulated (Table 2).

Epitope Conservancy Analysis (ECA) and Clumped cluster (clique) Analysis were done to show interconnectivity among predicted peptides by adopting IEDB tools viz., (http://tools.iedb.org/conservancy/) and (http://tools.iedb.org/cluster/), respectively (Figure 1B and Table 3).

[DOI: 10.52547/vacres.7.2.61]

Table 2. Epitopes identified within the RBD site of spike protein (YP_009724390.1) of SARS CoV2(shown in Table 1) using IEDB online software tool for the population of Andhra Pradesh, India based on their class I HLA haplotypes.
HLA haplotypes of the population of Andhra Pradesh, India: A*01-B*57-DRB1*07; A*33-B*44-DRB1*07; A*02-B*40-DRB1*15; A*24-B*07-DRB1*15; A*24-B*40-DRB1*15 [18].

S.No.	RBD		Length of peptide	Peptide	Core	Icore	Rank
	Epitope Start Sequence	Epitope End Sequence					
1	343	354	12	LNDLCFTNVYAD	LNFTNVYAD	LNDLCFTNVYAD	100
2	384	395	12	DFTGCVIAWNSN	DFTGAWNSN	DFTGCVIAWNSN	100
3	385	395	11	FTGCVIAWNSN	FTVIAWNSN	FTGCVIAWNSN	100
4	392	402	11	WNSNNLDSKVG	WNSNNLDSG	WNSNNLDSKVG	100
5	435	444	10	PCNGVEGFNC	PCNGVGFNC	PCNGVEGFNC	100
6	435	447	13	PCNGVEGFNCYFP	PVEGFNCYF	PCNGVEGFNCYF	100
7	477	488	12	PATVCGPKKSTN	PATPKKSTN	PATVCGPKKSTN	100
8	483	496	14	PKKSTNLVKNKCVN	PKKSTNLVV	PKKSTNLVKNKCV	100
9	484	496	13	KKSTNLVKNKCVN	KKSTNLVKN	KKSTNLVKNKCVN	100
10	517	527	11	PFQQFGRDIAD	PQFGRDIAD	PFQQFGRDIAD	100

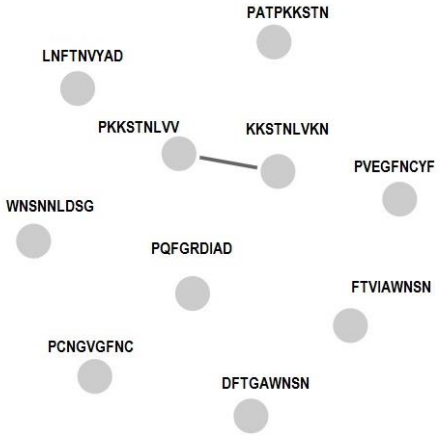

Fig. 1B. Clumped cluster (clique) showing the interconnectivity among peptides developed by adopting Epitope Cluster Analysis online tool viz,http://tools.iedb.org/cluster/ [29]. Peptides are presented as circles and line connecting two circles means the two peptides shared identity above given threshold.

Secondary Structure Analysis and Homology Modeling Validation

Each of the ten identified epitopes (Table 3) was connected through the spacer residues viz., GS [23]. The total amino acid residues in the generated polytope were 108. This polytope sequence was submitted to online service tool to obtain secondary structure prediction through PEP2D web server http://crdd.osdd.net/raghava/pep2d/result/pep2dress.php, which predicted solvent accessibility for the residues of polytope in cellular environment derived through https://zhanglab.ccmb.med.umich.edu/I-TASSER/output/S599287/. Dihedral angles viz., psi against phi of the residues in polytope were determined through Ramachandran plot in PROCHECK (https://servicesn.mbi.ucla.edu/PROCHECK/) and homology modeling of sequentially arranged multi-epitopes with the spacer (GS) in the polytope was determined using I-TASSER online services (https://zhanglab.ccmb.med.umich.edu/I-TASSER/output/S599287/). Ligand binding site residues were determined using COFACTOR and COACH approach in I-TASSER.

[DOI: 10.52547/vacres.7.2.61]

Table 3. Predicted epitopes of spike protein of SARS CoV2 based on conservancy analysis using (http://tools.iedb.org/conservancy/)

Epitope No.	Epitope sequence	Epitope length	Identity in Maximum %						
			SARS CoV2 Wuhan	SARS CoV2 Bangl.	Bat CoV	SARS CoV	MERS CoV	229E CoV	NL63 CoV
1	LNFTNVYAD	9	88.89	88.89	77.78	77.78	44.44	44.44	44.44
2	DFTGAWNSN	9	55.56	55.56	44.44	44.44	44.44	44.44	44.44
3	FTVIAWNSN	9	77.78	77.78	55.56	44.44	44.44	44.44	44.44
4	WNSNNLDSG	9	88.89	88.89	44.44	44.44	33.33	44.44	44.44
5	PCNGVGFNC	9	55.56	55.56	33.33	44.44	55.56	44.44	44.44
6	PVEGFNCYF	9	88.89	88.89	33.33	44.44	44.44	44.44	33.33
7	PATPKKSTN	9	66.67	66.67	44.44	44.44	44.44	44.44	33.33
8	PKKSTNLVV	9	88.89	88.89	66.67	55.56	44.44	44.44	44.44
9	KKSTNLVKN	9	100.00	100	77.78	66.67	44.44	44.44	44.44
10	PQFGRDIAD	9	88.89	88.89	66.67	66.67	44.44	44.44	44.44
	Average		80.00	80.00	54.44	53.33	44.44	44.44	42.22

Physicochemical features

Physicochemical properties namely MW, pI, NC residues, PC residues, II, aliphatic index and GRAVY of polytope given in Table 3 were determined through ProtParam online tool (https://web.expasy.org/cgi-bin/protparam/protparam) and tabulated in Table 4.

Table 4. Physicochemical properties of polytope are shown using ProtParam online tool (https://web.expasy.org/cgi-bin/protparam/protparam)

Number of amino acids in the polytope	MW	pI	Total number of negatively charged Residues (Asp+Glu)	Total number of positively charged Residues (Arg+Lys)	II	Aliphatic index	GRAVY
108	11222.16	8.54	6	8	21.92 (Stable)	45.09	-0.567

MW= Molecular weight, pI = Isoelectric point, II = Instability index, GRAVY= Grand Average Hydropathicity

Prediction of T and B Cell Epitopes and Antigenicity

The HLA haplotypes of the population of the State of Andhra Pradesh among Indian Asians were retrieved from Dedhia [18]. MHC Class I immunogenicity prediction using tools.iedb.org/immunogenicity, MHC Class I binding predictions with IEDB analysis resource NetMHCpan (ver. 4.1) tool [24] and B-cell epitope predictions using Bepipred Linear Epitope Prediction 2.0 tool were carried out for the identified multi-epitope of SARSCoV2. The overall prediction for the protective antigenicity of the designed polytope using VaxiJen V2.0 tool is 0.5008 at a threshold of 0.4 ("Probable ANTIGEN"); (http://www.ddg-pharmfac.net/vaxijen/scripts/VaxiJen_scripts/VaxiJen3.pl)

Population Coverage

Multi-epitope, each comprising of nine amino acid residues, were submitted to http://tools.iedb.org/population/ along with HLA I alleles of Indian Asian population localizing to the State of Andhra Pradesh. The fraction of individuals predicted to respond to a given set of epitopes with known MHC restrictions in Indian Asian population were evaluated (tools.iedb.org/population/example_genotype_locus/).

RESULTS

Phylogenetic Analysis

The viral pandemic in the year 2020 spread across nations indicate that there is an utmost need to develop herd immunity to protect human population by evaluating the origin and molecular affinities among CoVs [26]. In the present study, seven representatives of the family Coronaviridae are chosen to build the affinity of SARS CoV2 isolate from Wuhan, China. The constructed phylogenetic tree is rooted on the out-group primary cluster namely alpha CoVs (AFV53148.1 and NP_073551.1). Furthermore, the SARS CoV2 isolate of Wuhan-hu-1 clustered with the isolate of Bangladesh is derived as a primary cluster showing the recency of their emergence (Fig.1).The branch lengths of other primary cluster formed between bat SARS CoV and SARS CoV showed that they are ancient to SARS CoV2 indicating possible transmission from the animal source namely bat and hence it is nonetheless zoonotic in its origin. Further, branch lengths (Fig.1) of Wuhan and Bangladesh isolates compared to others indicated that CoV2s eruption is recent. Possibly, in consonance with the difference in sequences and RBD located in spike protein region I (Table 1), alpha CoV representatives namely N63 and 229E constituted another independent primary cluster away from the tertiary cluster (Fig.1). Interestingly, MERS CoV

[DOI: 10.52547/vacres.7.2.61]

formed as an offshoot of the tertiary cluster of beta CoVs and also its branch length indicated that MERS CoV is phylogenetically older among betacoronaviruses. YP_009724390.1 and QLF97795.1of SARS CoV2s isolate from Wuhan and Bangladesh are showing relatively high percent identity with SARS and Bat SARS CoVs as revealed in Percent Identity Matrix with the values namely 77.30 and 77.14 respectively compared to other strains chosen in the present study (Table 5) which confirmed that SARS CoV2 and SARS are the least in their divergence from bat CoV.

Table 5. Percent Identity Matrix of the spike protein sequences of the chosen members of Coronaviridae strains created by Clustal 2.1 (https://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=clustalo-I20210516-091135-0411-93578101-p1m&analysis=summary)

S.No.	**Protein Id from NCBI**	1	2	3	4	5	6	7
1	AFV53148.1	100.00						
2	NP_073551.1|S	64.15	100.00					
3	AMO03401.1|S	25.97	28.27	100.00				
4	YP_009724390.1	27.79	28.63	31.85	100.00			
5	QLF97795.1	27.70	28.63	31.77	99.69	100.00		
6	ACU31032.1	26.70	28.02	31.46	76.56	76.39	100.00	
7	ABA02260.1|S	26.12	28.00	31.59	77.30	77.14	80.16	100.00

Epitope Prediction and Polytope Construction

Ten epitopes, each with nine amino acid residues were identified within the receptor binding domain of the spike protein (YP_009724390.1) of SARS CoV2 restricting to the HLA haplotypes of the population of the State of Andhra Pradesh within Indian Asians. The predicted multi-epitope conservancy analysis (Tables 3 & 5) revealed that alpha and beta coronaviruses are more diverged. Therefore, the vaccine to be designed with epitope conservancy data (Tables 3 & 5) would also serve to combat the future eruption of SARS CoV. Between two epitopes within the identified multi-epitope, glycine and serine residue spacer was used as a link and generated a polytope to evaluate the secondary structure, homology model and ligand binding site residues to assess the antigenicity of epitopes.

Physicochemical Features of Poytope

The polytope comprising of 10 epitopes of RBD region of spike protein of SARS CoV2 as given in Table 3 were submitted to ProtParam tool- ExPASy. The resulted data is presented in Table 4 with the features namely molecular weight, pI, negatively and positively charged residues, instability index, aliphatic index and GRAVY. These features indicate that polytope is stable with 12.96% charged residues and solvent accessibility with more hydrophilicity.

Secondary Structure Analysis and Homology Modelling Validation

Secondary structure prediction and homology modelling (Fig. 2) of sequentially arranged ten cognate epitopes in the polytope (Table 3) with two spacers link between them namely G (glycine) and S (serine) were determined using I-TASSER online service (https://zhanglab.ccmb.med.umich.edu/I-TASSER/output/S599287/) tool. Importantly, the predicted secondary structure yielded the presence of random coils, β-strands and alpha helices (Fig. 2A).

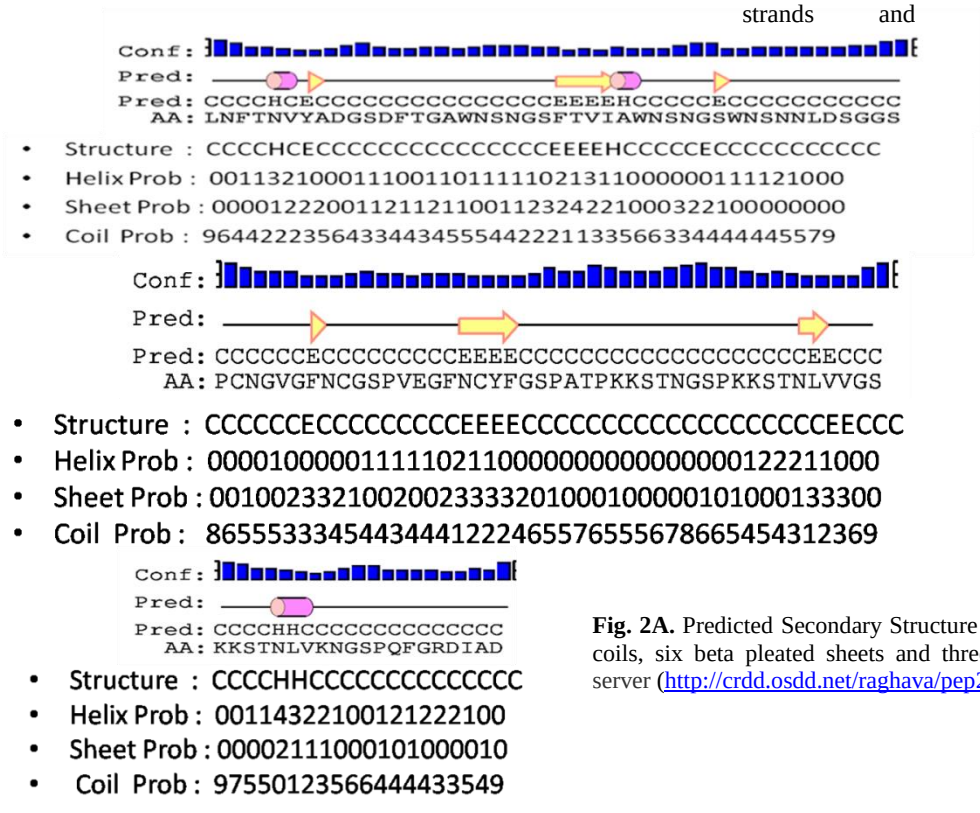

Fig. 2A. Predicted Secondary Structure of the polytope showing predominantly coils, six beta pleated sheets and three α-helix regions through PEP2D web server (http://crdd.osdd.net/raghava/pep2d/result/pep2dress.php).

[DOI: 10.52547/vacres.7.2.61]

All residues in the generated polytope are exposed so as to interact with aqueous environment prevailing in the host system (Fig. 2B).

The phi angles against psi angels of residues of polytope were shown in Ramachandran Plot with 41.2% residues in the most favoured and 44.7% residues in the additional allowed regions of the plot (Fig. 2C).

```
                    20        40        60        80        100
                    |         |         |         |         |
Sequence    LNFTNVYADGSDFTGAWNSNGSFTVIAWNSNGSWNSNNLDSGGSPCNGVGFNCGSPVEGFNCYFGSPATPKKSTNGSPKKSTNLVVGSKKSTNLVKNGSPQFGRDIAD
Prediction  662442334344132314562311000024634144642554343033232413332532302233433345546444444231222354434225534352445348

            Values range from 0 (buried residue) to 9 (highly exposed residue)
```

Fig. 2B. Predicted solvent accessibility for the residues of polytope in cellular environment derived through (https://zhanglab.ccmb.med.umich.edu/I-TASSER/output/S599287/) .

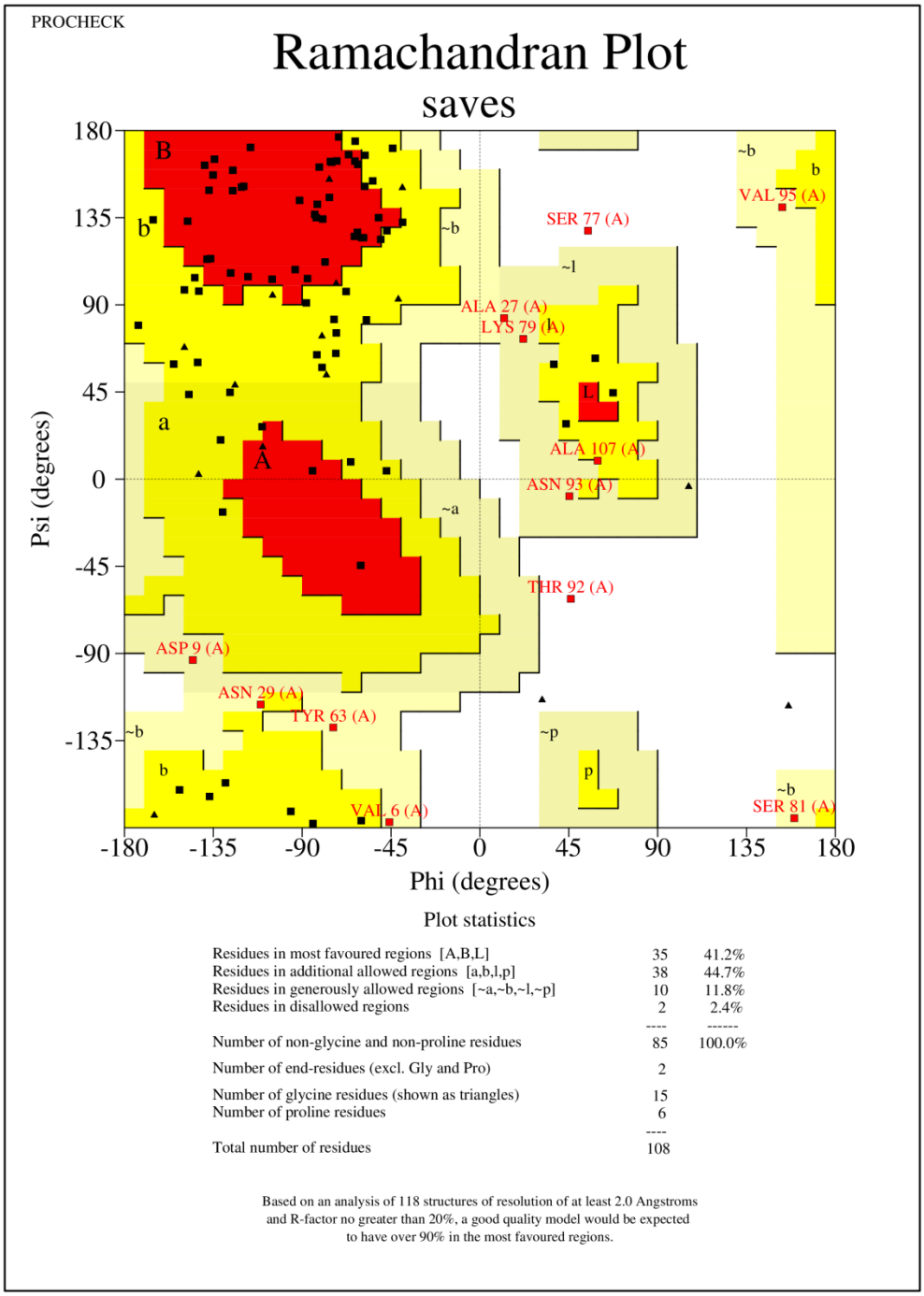

Fig. 2C. Ramachandran plot developed for the polytope using PROCHECK. 41.2% residues of polytope are in the most favoured region and 44.7% are in allowed regions.

[DOI: 10.52547/vacres.7.2.61]

Total amino acids in the polytope are 108 including spacers. The homology model for the polytope was derived from I-TASSER uses the SPICKER program (Fig. 2D). After analysis, the models with the highest confidence score (C-score) were selected for refinement analysis. The solvent accessibility of the predicted model is high because all residues are well exposed (Fig. 2B). Ligand binding site residues based on COFACTOR and COACH approach on the I-TASSER structure prediction of the generated polytope are:4,6,7,17,20,25,26,27,28,29,0,31,3,34,40,41,46,47,48,50,60, 61,65,66,67,69,70,85 which suggests that this polytope does invariably bind to MHC Class I so as to enable to present through Protein-Major Histocompatibility Complex (pMHC) complex to T-cell for elicitation of immune response.

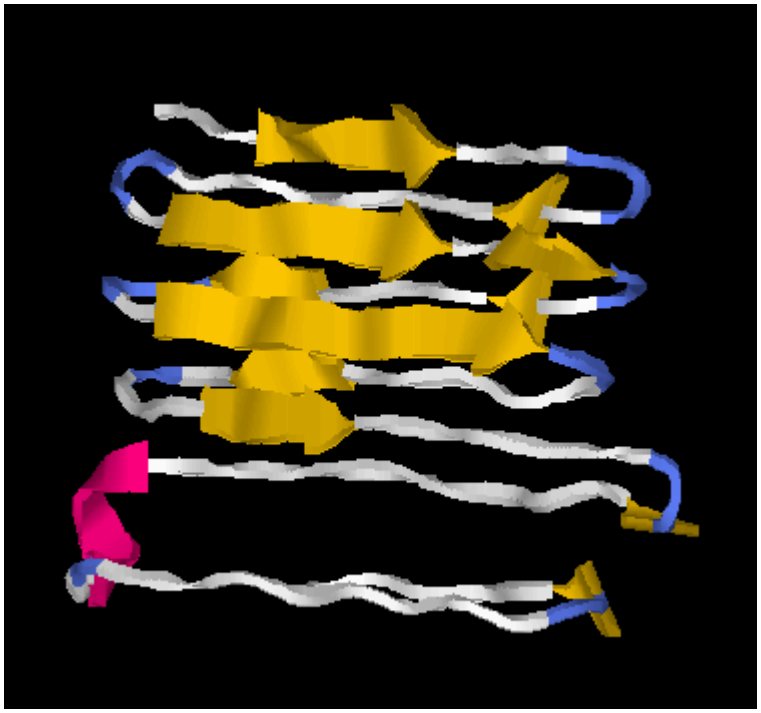

Fig. 2D. Homology modelling of sequentially arranged multi-epitope with the spacer (G,S) in the polytope was determined using online I-TASSER.

Prediction of T and B cell Epitopes and Antigenicity

The multi-epitope of spike protein of SARS CoV2 derived through IEDB tool was submitted to IEDB in conjunction with HLA alleles of Indian Asian population to evaluate immunogenicity. The low score of linear epitopes is an indication of high immunogenic peptides (Table 6). The epitopes are given in the order of relatively low to high immunogenicity in Table 6.

Table 6. MHC Class I Immunogenicity prediction due to multi-epitopes of spike protein of SARS CoV2 given in the Table 4 using IEDB online software tool. As a default, 1 and 2 C- terminal residues are masked. (tools.iedb.org/immunogenicity)

Epitope No.	Peptide	Length	Score
3	FTVIAWNSN	9	0.29181
10	PQFGRDIAD	9	0.27856
5	PCNGVGFNC	9	0.19912
2	DFTGAWNSN	9	0.19119
1	LNFTNVYAD	9	0.12936
6	PVEGFNCYF	9	0.12685
9	KKSTNLVKN	9	-0.12254
4	WNSNNLDSG	9	-0.15489
8	PKKSTNLVV	9	-0.19
7	PATPKKSTN	9	-0.5285

[DOI: 10.52547/vacres.7.2.61]

The MHC Class I binding potential shown in Table 7 also revealed the low score indicating high level affinity and immunogenicity of pMHC complex. HLA haplotypes of the population of Andhra Pradesh, India: A*01-B*57-DRB1*07; A*33-B*44-DRB1*07; A*02-B*40-DRB1*15; A*24-B*07-DRB1*15; A*24-B*40-DRB1*15 [18].

The alleles of the population of Andhra Pradesh represented in Indian Asians considered in the present analysis: HLA-A*01:01, HLA-A*02:01, HLA-A*24:02, HLA-A*31:01, HLA-B*07:02, HLA-B*15:01, HLA-B*27:01, HLA-B*40:01, HLA-DRB1*07:01, HLA-DRB1*15:01 [18] (tools.iedb.org/population/example genotype_locus/).

Table 7. MHC binding predictions of multi-epitopes of spike protein of SARS CoV2 using the IEDB analysis resource NetMHCpan (ver. 4.1) tool [24] with frequent occurrence of MHC Class I and II alleles in Indian Asian population are given. The toxicity, antigenicity and allergenicity parameters due to the epitopes are shown.

Toxicity	Antigenicty Score	Allergenicity	% PC among Indian Asians		Length	Prediction score for MHC I binding	Peptide
			MHC I	MHC II			
Non-Toxin	0.5637	Non-Allergen	46.52	41.78	9	0.29181	PATPKKSTN
Non-Toxin	0.3985 (NA)	Non-Allergen	46.52	41.78	9	0.27856	DFTGAWNSN
Non-Toxin	1.0582	Non-Allergen	46.52	41.78	9	0.19119	WNSNNLDSG
Non-Toxin	1.1898	Non-Allergen	46.52	41.78	9	0.12936	LNFTNVYAD
Non-Toxin	0.8976	Non-Allergen	46.52	41.78	9	-0.12254	FTVIAWNSN
Non-Toxin	0.4876	Non-Allergen	46.52	41.78	9	-0.15489	KKSTNLVKN
Non-Toxin	-0.5389 (NA)	Non-Allergen	46.52	41.78	9	-0.5285	PQFGRDIAD

NA: Non-Antigen, PC: Population Coverage.
The URLs for predicting the parameters indicated in the table are mentioned hereunder:
https://webs.iiitd.edu.in/raghava/toxinpred/multi_submit.php
http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html
http://www.pharmfac.net/allertop/
http://tools.iedb.org/population/result/
http://tools.iedb.org/immunogenicity/

The identified multi-epitopes are of high hydrophilic nature as shown through GRAVY score in addition to highly immunogenic revealing that they are supposedly good antigenic and therefore have high propensity to be recognized by TCR complex. Further, the antigenic nature of multi-epitope is confirmed through the Bepipred Linear Epitope Prediction 2.0 tool which gave the score more prone towards B-cell epitopes (Fig.3). The following table 8 is an extract from ElliPro (Epitope 3D Structures for filemz5lpe6l.pdb) on the discontinuous epitopes of the designed polytope.

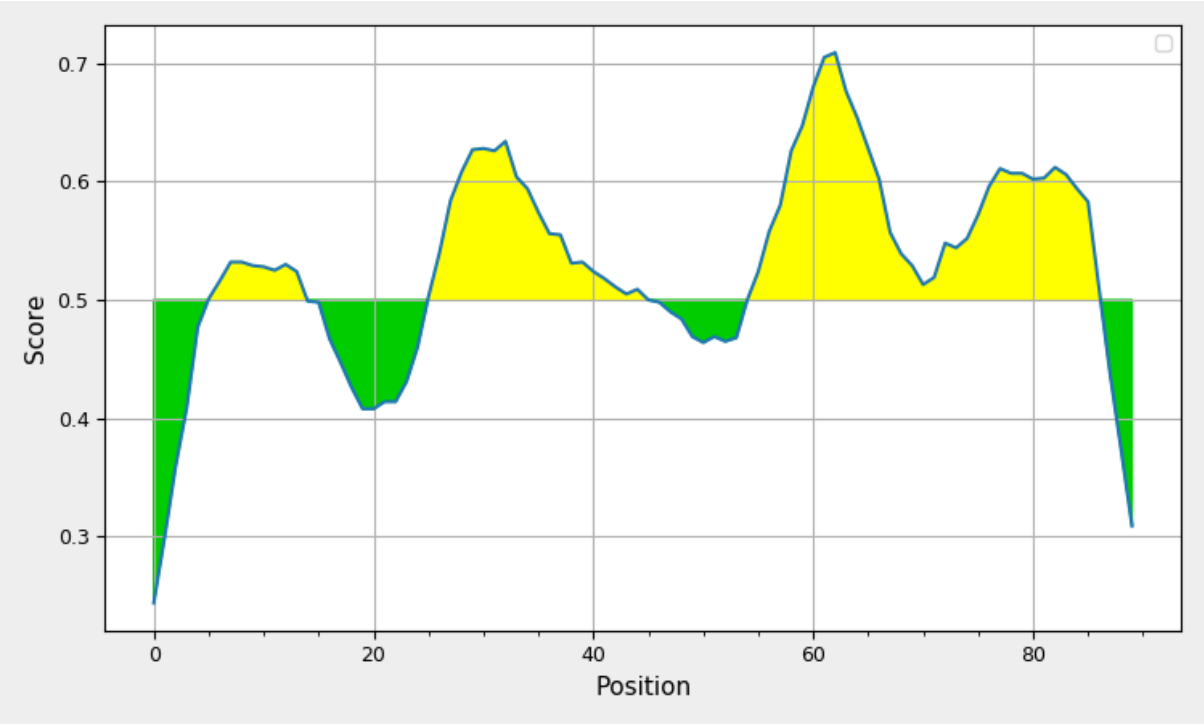

Fig. 3. B cell epitope prediction for multi-epitopes given in the Table 3 at a threshold value of 0.500 with a maximum score 0.709. This graphic form is derived through Bepipred Linear Epitope Prediction 2.0. The yellow colour indicates higher epitope prediction scores.

[DOI: 10.52547/vacres.7.2.61]

Table 8. Predicted discontinuous B-cell epitopes of the designed polytope obtained from online IEDB Ellipro tool. http://tools.iedb.org/ellipro/result/predict/

No.	Residues	Number of residues	Score
1	A:L1, A:N2, A:F3, A:T4, A:N5, A:V6, A:Y7, A:A8, A:D9, A:G10, A:S11, A:D12, A:F13, A:T14, A:G15, A:A16, A:W17, A:N18, A:S19, A:N20, A:A27, A:W28, A:N29, A:S30, A:N31, A:G32, A:W34, A:N35, A:S36, A:N37, A:N38, A:L39, A:D40, A:S41, A:G42, A:G43, A:S44, A:F51, A:N52, A:C53, A:G54, A:S55, A:P56, A:E58, A:Y63, A:F64, A:G65, A:S66, A:P67, A:A68, A:T69, A:S73, A:T74, A:N75, A:G76, A:S77, A:P78, A:K79, A:K80, A:S81, A:G87, A:S88, A:K89, A:K90, A:S91, A:T92, A:N93, A:L94, A:V95, A:K96, A:S99, A:P100, A:Q101, A:F102, A:G103, A:R104, A:D105, A:I106, A:A107, A:D108	80	0.59

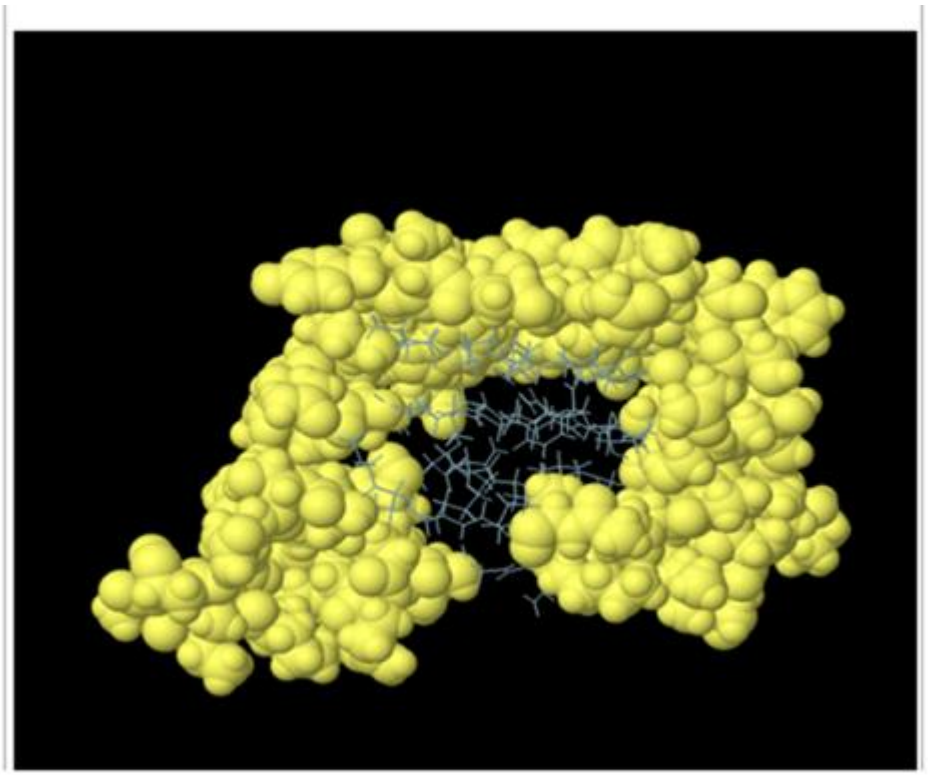

Fig. 4.The pose of predicted discontinuous epitopes of polytope (shown in table 8) binding to the antibody (B-cell) retrieved from tools.iedb.org/ellipro/result/jsmol/?epitope=1&no=0.

Population Coverage

The IEDB-based tool was used to predict population coverage of T-cell epitope-based multi-epitope vaccine with restricted HLA Class I binding of Indian Asians. Accordingly, multi-epitopes are shown in Table 9 to elicit immune response in compliance with prevailing HLA restriction in the destined population.

The identified epitopes were found to hit all HLA haplotypes under consideration. The previous observations on the physicochemical properties of polytope (Table 4), propensity for interaction in an aqueous environment (Fig.2B), T and B cell epitope prediction (Tables 6 & 7) along with HLA hits suggest that the variability of coverage and complexity are minimized with adequate prediction in the destined ethnic group namely the population of the State of Andhra Pradesh among Indian Asians. The designed polytope is reported to be "probable non-allergen" through AllerTop online software tool, (http://www.pharmfac.net/allertop/feedback.py). The following table 10 shows the toxicity evaluation of the polytope using online software tool (https://webs.iiitd.edu.in/raghava/toxinpred/monofreq_sp.php?ran=46813).

[DOI: 10.52547/vacres.7.2.61]

Table 9. Population coverage tool (IEDB) showing the fraction of individuals predicted to respond to a givenset of epitopes with known MHC restrictions in Indian Asian population. (tools.iedb.org/population/example_genotype_locus/)

Epitope	Coverage	HLA allele (genotypic frequency (%))							Total HLA hits
	Class I	HLA-A*01:01 (8.36)	HLA-A*02:01 (3.81)	HLA-A*24:02 (11.25)	HLA-A*33:01 (0.59)	HLA-B*07:02 (2.29)	HLA-B*40:01 (1.24)	HLA-B*44:02 (1.26)	
Epitope #1: LNFTNVYAD	37.78%	+	+	+	+	+	+	+	7
Epitope #2: DFTGAWNSN	37.78%	+	+	+	+	+	+	+	7
Epitope #3: FTVIAWNSN	37.78%	+	+	+	+	+	+	+	7
Epitope #4: WNSNNLDSG	37.78%	+	+	+	+	+	+	+	7
Epitope #5: PCNGVGFNC	37.78%	+	+	+	+	+	+	+	7
Epitope #6: PVEGFNCYF	37.78%	+	+	+	+	+	+	+	7
Epitope #7: PATPKKSTN	37.78%	+	+	+	+	+	+	+	7
Epitope #8: PKKSTNLVV	37.78%	+	+	+	+	+	+	+	7
Epitope #9: KKSTNLVKN	37.78%	+	+	+	+	+	+	+	7
Epitope #10: PQFGRDIAD	37.78%	+	+	+	+	+	+	+	7
Epitope set	**37.78%**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**70**

+: restricted

The identified epitopes were found to hit all HLA haplotypes under consideration. The previous observations on the physicochemical properties of polytope (Table 4), propensity for interaction in an aqueous environment (Fig.2B), T and B cell epitope prediction (Tables 6 & 7) along with HLA hits suggest that the variability of coverage and complexity are minimized with adequate prediction in the destined ethnic group namely the population of the State of Andhra Pradesh among Indian Asians. The designed polytope is reported to be "probable non-allergen" through AllerTop online software tool, (http://www.pharmfac.net/allertop/feedback.py).

The following table 10 shows the toxicity evaluation of the polytope using online software tool (https://webs.iiitd.edu.in/raghava/toxinpred/monofreq_sp.php?ran=46813).

Table 10. Prediction of toxicity score of fragmented polytope (<50) using ToxinPred tool.

Fragment	Peptide Sequence	SVM Score	Prediction	Hydrophobicity	Hydropathicity	Hydrophilicity	Charge	Molecular Weight
1	LNFTNVYADGSDFTGAWNSNGSFTVIAWNSNGSWNSNNLDSGGS	-1.14	Non-Toxin	-0.07	-0.51	-0.41	-3.00	4647.44
2	PCNGVGFNCGSPVEGFNCYFGSPATPKKSTNGSPKKSTNLVV	-0.42	Non-Toxin	-0.11	-0.38	-0.11	3.00	4292.45
3	GSKKSTNLVKNGSPQFGRDIAD	-0.55	Non-Toxin	-0.29	-1.02	0.50	2.00	2319.88

SVM: Support Vector Machine

DISCUSSION

The virus causing COVID-19 has become human host-specific because of its affinity to anchor ACE2 receptors of alveolar/intestinal epithelial cells (Table 1) [27]. As a result, normal function and lifestyle of human population are affected drastically [1-4]. In a few countries, the COVID-19 confirmed cases are in the linear decrease and at the same time, the resistance is developed among the immune competent youth and they happened to be the carriers of the virus. Hence, herd immunity practices have to be taken up to restore the normal function and life style of human population. In this context, it is

[DOI: 10.52547/vacres.7.2.61]

imperative to understand the molecular affinity of SARS CoV2 with its counterparts in the family, Coronaviridae. The strains of alpha CoV group are found to erupt sporadically causing mild illness without serious issues [26]. Whereas, beta CoVs cause acute respiratory distress syndrome along with other symptoms such as organ failure [1, 9, 12]. These viruses possess single-stranded positive-sense RNA as a genetic material having nearly 30,000 nucleotides with unique spike protein containing RBD (Table 1). The phylogenetic tree constructed for the retrieved spike protein sequences (Table 1) of the representative members of the family Coronaviridae is shown in fig. 1. The tree is rooted on the out-group alpha CoVs, which formed a primary cluster [21, 22].

SARS CoV2 isolate of Wuhan-hu-1 constituted yet another primary cluster with Bangladesh isolate with short branch lengths suggesting its recent eruption. The SARS outbreak appeared in the year 2003 and the present pandemic of SARS CoV2 started in the year 2019. Incidentally, in this span of 16 years, the newly emerged strain of coronavirus became highly infectious to human population. Thus, it has become a challenge for scientists to show an evidence based prevention of COVID-19 infection. MERS coronavirus strain whose outbreak was in Saudi Arabia [27, 28] is yet another offshoot in the tertiary cluster of beta coronaviruses in the phylogenetic tree (Fig.1) that created pandemic in the year 2012 has now become calm. Thus, not only the genome but also the amino acid sequence of spike protein of SARS CoV2 reflects that it is showing high molecular affinity with SARS than MERS coronavirus and authenticates a possible origin through zoonotic infection from bats in Wuhan, China [13, 17, 27]. Whereas, the other two alpha strains of coronaviruses namely NL63 and 229E formed as an out group primary cluster and possibly rooted the rest of the four beta coronaviruses under study primarily because their S1 region of the spike proteins varies and the bootstrap value of 100 revealed that betacoronaviruses evolved from their alpha counterparts. Tilocca et al [15] analyzed epitopes of spike proteins of CoV2 in the sequence ranges of 424-437, 447-458, 754-764, 789-799 and 1139-1152 and reported 80%, 75%, 83%, 57% and 70% identity respectively. Among them, the first two sequence ranges fall in the RBD of SARS CoV2 and the % identity reported is comparable with our results shown in Tables 1 and 3. Furthermore, the least percent identity values are shown for the alpha coronavirus strains namely NL63 and 229E and they are 27.79 and 28.63 respectively with SARS CoV2 indicating that alpha and Betacoronaviruses are inherently different clades (Table 1, Fig. 1). In the present investigation it is shown that the percent identity of spike proteins of SARS CoV2 with SARS is 77.30, the value of which is akin to the observations made by Ou [3]. Therefore, due to the nearest molecular affinity between SARS and SARS CoV2 as evidenced through the appearance of sister taxa in tertiary cluster in the phylogenetic tree (Fig.1), it is suggested in the present observation that there must also have been the similarity among the epitopes of spike proteins. Furthermore, the epitope conservancy analysis (Tables 3 and Fig.1) revealed that SARS and SARS CoV2 shared relatively higher epitope conservancy which suggests a common prospective vaccine candidate against these two viruses. The novel clustering online tool http://tools.iedb.org/cluster2/ [29] includes “clique method” that was adopted in the present investigation to observe the defined level of identity with each of the predicted epitopes of spike protein of SARS CoV2.

The physicochemical features of polytope (Table 4) revealed that its residues interact in aqueous environment and the construct is shown as stable with 12.96% charged residues. The MHC I binding prediction of epitopes as shown in Table 7 reveal that the conceived multi-epitope are antigenic with the scores (Tables 6 and 7) that authenticate its immunogenicity. The generated polytope also showed the secondary structure (Fig. 2A) with exposed residues and ligand binding residue sites suggest that pMHC complex elicits T-cell immune response [25]. Interestingly, the observation made in line with One-Health concept by Tilocca et al [16] displayed the conserved homology in 3-D structures of yet another structural envelop protein of human enteric CoV and CoVs of bovine and dog and reported that they share a common antigenic potential that possibly elicit cross-protection among the interacting synanthropic animal species with which unintentionally a few human communities interacted in their routine lifestyle. With known MHC restrictions of the population of the State of Andhra Pradesh (18), the most promising multi-epitope candidate with the desired physicochemical features is designed for the Indian Asian population including the State of Andhra Pradesh, which showed the adequate coverage.

In conclusion, the spike protein sequence of SARS CoV2 reflected that it is showing molecular affinity with the chosen Betacoronaviruses as they all constituted a single tertiary cluster. Further, the physicochemical features of the polytope authenticate that the designed poytope interacts well in aqueous environment. MHC I immunogenicity prediction, B-cell epitope, ToxinPred and AllrTop prediction revealed the multi-epitope are antigenic, non-toxic, non-allergic and hence the suggested polytope would be the possible vaccine candidate to attempt for experimentation.

ACKNOWLEDGEMENT

The author acknowledges VFSTR (Deemed to be university) administration and DST-FIST (LSI- 576/2013) for providing the networking facility to carry out this work.

CONFLICT OF INTEREST

The author declares he has no conflict of interests.

REFERENCES

1. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, et al. Author Correction: A new coronavirus associated with human respiratory disease in China. Nature. 2020;580(7803):E7.

2. Krupanidhi S, Abraham Peele K, Venkateswarulu TC, Ayyagari VS, Nazneen Bobby M, John Babu D, et al. Screening of phytochemical compounds of Tinospora cordifolia for their inhibitory activity on SARS-CoV-2: an in silico study. J Biomol Struct Dyn. 2020;1–5.

3. Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat Commun. 2020;11(1):1620.

4. Prajapat M, Sarma P, Shekhar N, Prakash A, Avti P, Bhattacharyya A, et al. Update on the target structures of SARS-CoV-2: A systematic review. Indian J Pharmacol. 2020;52(2):142–9.

5. Zhang H, Penninger JM, Li Y, Zhong N, Slutsky AS. Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. Intensive Care Med. 2020;46(4):586–90.

6. Peele KA, Potla Durthi C, Srihansa T, Krupanidhi S, Ayyagari VS, Babu DJ, et al. Molecular docking and dynamic simulations for antiviral compounds against SARS-CoV-2: A computational study. Inform Med Unlocked. 2020;19(100345):100345.

7. Ayyagari VS, T C V, K AP, Srirama K. Design of a multi-epitope-based

[DOI: 10.52547/vacres.7.2.61]

vaccine targeting M-protein of SARS-CoV2: an immunoinformatics approach. J Biomol Struct Dyn. 2020;1–15.

8. Song W, Gui M, Wang X, Xiang Y. Cryo-EM structure of the SARS coronavirus Spike glycoprotein in complex with its host cell receptor ACE2, PLoSPathogens. Vol. 13;14(8):e1007236. 2018.

9. Xia S, Liu M, Wang C, Xu W, Lan Q, Feng S, et al. Inhibition of SARS-CoV-2 infection (previously 2019-nCoV) by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion [Internet]. bioRxiv. 2020. Available from: http://dx.doi.org/10.1101/2020.03.09.983247

10. Mandal A, Viswanathan C. Natural killer cells: In health and disease. Hematol Oncol Stem Cell Ther. 2015;8(2):47–55.

11. Gralinski LE, Sheahan TP, Morrison TE, Menachery VD, Jensen K, Leist. Complement Activation Contributes to Severe Acute Respiratory Syndrome coronavirus Pathogenesis, Host-Microbe Biol. Vol. 9. 2018. p. 01753–18.

12. Saqr M, Wasson B. COVID-19: Lost opportunities and lessons for the future. Int J Health Sci (Qassim). 2020;14(3):4–6.

13. Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, et al. Bats are natural reservoirs of SARS-like coronaviruses. Science. 2005;310(5748):676–9.

14. Tilocca B, Soggiu A, Sanguinetti M, Musella V, Britti D, Bonizzi L, et al. Comparative computational analysis of SARS-CoV-2 nucleocapsid protein epitopes in taxonomically related coronaviruses. Microbes Infect. 2020;22(4–5):188–94.

15. Tilocca B, Soggiu A, Sanguinetti M, Babini G, De Maio F, Britti D, et al. Immunoinformatic analysis of the SARS-CoV-2 envelope protein as a strategy to assess cross-protection against COVID-19. Microbes Infect. 2020;22(4–5):182–7.

16. Tilocca B, Soggiu A, Musella V, Britti D, Sanguinetti M, Urbani A, et al. Molecular basis of COVID-19 relationships in different species: a one health perspective. Microbes Infect. 2020;22(4–5):218–20.

17. Hamed SM, Elkhatib WF, Khairallah AS, Noreddin AM. Global dynamics of SARS-CoV-2 clades and their relation to COVID-19 epidemiology [Internet]. Research Square. 2020. Available from: http://dx.doi.org/10.21203/rs.3.rs-89876/v1

18. Dedhia L, Gadekar S, Mehta P, Parekh S. HLA haplotype diversity in the South Indian population and its relevance. Indian j transplant. 2015;9(4):138–43.

19. Abdoli A, Radmehr N, Bolhassani A, Eidi A, Mehrbod P, Motevalli F, et al. Conjugated anionic PEG-citrate G2 dendrimer with multi-epitopic HIV-1 vaccine candidate enhance the cellular immune responses in mice. Artif Cells Nanomed Biotechnol. 2017;45(8):1762–8.

20. Himmah K, Fitriyah, Ardyati T, Afiyanti M, Rifa'i M, Widodo. Designing a polytope for use in a broad-spectrum dengue virus vaccine. J Taibah Univ Med Sci. 2018;13(2):156–61.

21. S. K, G. S, M. L, C. K, K T. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution. 2018;35:1547–1549.

22. Mello B. Estimating time-trees with MEGA and the TimeTree resource. Molecular biology and evolution. 2018;35(9):2334-42.

23. Himmah K, Brawijaya University, Fitriyah F, Ardyati T, Deocaris C, Widodo N, et al. Polytope prediction for dengue vaccine candidate based on conserved envelope glycoprotein of four serotypes of dengue virus and its antigenicity. J pure appl chem res. 2016;5(2):101–7.

24. Reynisson B, Alvarez B, Paul S, Peters B, NetMHCpan-4 NM. 1 and NetMHCIIpan-4. 0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic acids research. 2020;Jul 2;48(W1):W449-54.

25. Shi Y, Wang Y, Shao C, Huang J, Gan J, Huang X, et al. COVID-19 infection: the perspectives on immune responses. Cell Death Differ. 2020;27(5):1451–4.

26. Qiu Y, Zhao Y-B, Wang Q, Li J-Y, ZHou Z-J, Liao C-H, et al. Predicting the angiotensin converting enzyme 2 (ACE2) utilizing capability as the receptor of SARS-CoV-2 [Internet]. Preprints. 2020. Available from: http://dx.doi.org/10.20944/preprints202003.0091.v1

27. Lu G, Hu Y, Wang Q, Qi J, Gao F, Li Y, et al. Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. Nature. 2013;500(7461):227–31.

28. To KKW, Hung IFN, Chan JFW, Yuen K-Y. From SARS coronavirus to novel animal and human coronaviruses. J Thorac Dis. 2013;5 Suppl 2:S103-8.

29. Dhanda SK, Vaughan K, Schulten V, Grifoni A, Weiskopf D, Sidney J, et al. Development of a novel clustering tool for linear peptide sequences. Immunology. 2018;155(3):331–45.

[DOI: 10.52547/vacres.7.2.61]