A Bioinformatics approach to vaccine development.

Potential antigen target site for vaccine development among COVID variants

Nidhi Rathod,

Assistant Professor
Department of Biotechnology
N.B.Mehta Science college Bordi, Maharashtra

Abstract **- Epidemic are occurring with increasing frequency across the world. Traditional vaccines development methods were time consuming with some incidence of allergic reactions, occurrence of resistant strains and many other inadequacies. COVID-19 was one such transmissible disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). To fight this pandemic, which has cause a massive death toll around the globe, researchers are putting efforts into developing an effective vaccine against the pathogen. As genome sequencing projects for several coronavirus strains have been completed, a comprehensive study of the functions of the proteins and their 3D structures has gained increasing awareness. These data are a valuable asset to quicken the emerging field of immuno-informatics, which is aimed toward the identification of potential antigenic epitopes in viral proteins that can be targeted for the development of a vaccine design to evoke a high immune response. Bioinformatics platforms and various computational tools and databases are used for the identification of promising vaccine targets making the best use of resources for further experimental validation. This study is an effort to compare structural similarities and differences among different coronavirus strains and designing a multi epitopes based vaccine.**

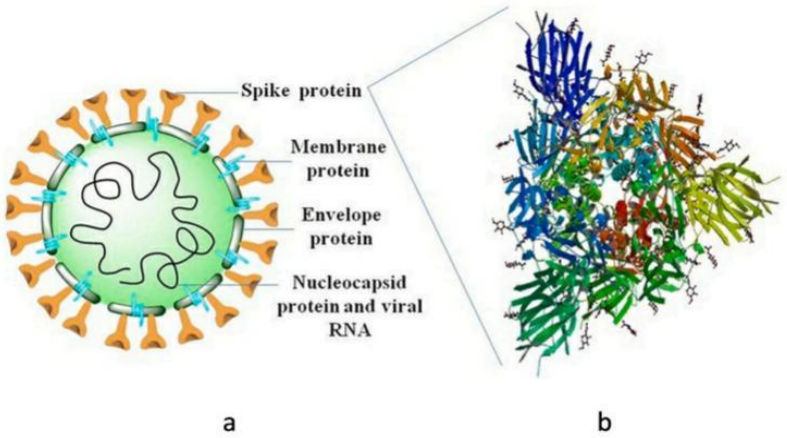
Keywords- **Epitopes, Bioinformatics, spike protein, COVID variants, multi epitope vaccine design.**

Introduction

Severe acute respiratory syndrome coronavirus type 2(SARS-CoV-2) is a causative agent of coronavirus disease 2019(COVID -19) outbreak in Wuhan China[4]. Coronaviruses are single-stranded, positive-sense RNA viruses belonging to the order Nidovirales, family Coronaviridae, and subfamily Coronavirinae [6].By early 2020, the disease was declared a global pandemic as a public health emergency of international concern The virus spreads from one individual to another by respiratory droplets in close contact between sick and asymptomatic people (within six feet)[4]. It is transmitted by aerosols and maybe contact with fomites is also a possibility, although this is not considered to be the most probable route. Pathogenesis of SARS cov2 depend on the viral spike protein binding to angiotensin-converting enzyme 2 (ACE2) receptors, to activate the viral protein cell entrance requires ACE2 receptor to be cleaved by a type 2 transmembrane serine protease[4]. Individual suffering from COVID-19 have a wide range of clinical symptoms, from moderate to severe, fast progressive, and acute disease[4]. The COVID-19 diagnosis is nonspecific, and the virus may exhibit itself in a variety of ways, ranging from asymptomatic to severe pneumonia and death[4]. The response plan of COVID-19 pandemic is based on the therapeutic alternatives and vaccine development.

Virion structure of coronavirus is made up of four major structural proteins namely, spike, envelope, membrane, nucleocapsid (Fig. 1). Entry of coronavirus into host cells is mediated by the transmembrane spike (S) glycoprotein that forms homotrimers projecting from the viral surface[16]. S protein contains two functional subunits responsible for binding to the host cell receptor (S1 subunit) and fusion of the viral and cellular membranes (S2 subunit). For many CoVs, S protein is spilt at the partition between the S1 and S2 subunits, which remain non-covalently bound in the conformation[16]. The S1 subunit interaction with the host receptor plays a major role in determining the host range virus.[6] Additionally, this brings about a conformational change in the S2 subunit of S protein thus exposing the hidden fusion peptide which aids in viral entry into the host cellular membrane.[6]. The conformational change leads to fusion the lipid bilayers and entry of viral nucleocapsid into the host cellular cytoplasm.[6]. Different coronaviruses use distinct domains within the S1 subunit to recognize a variety of attachment and entry receptors, depending on the viral species. SARS-CoV and several SARS-related coronaviruses interact directly with angiotensin-converting enzyme 2 (ACE2) to enter target cells[16]. As the coronavirus S glycoprotein is surface-exposed entity and mediates entry into host cells, it is the main target of neutralizing antibodies (Abs) upon infection and the focus of vaccine design.

Receptor recognition is the first step of viral infection and is a key determinant of host cell. Enhanced binding affinity between SARS-CoV S and hACE2 is known to correlate with increased virus transmissibility and disease severity in humans[16]. Indeed, SARS-CoV isolates from the three phases of the 2002–2003 epidemic were more efficiently transmitted among humans and are more pathogenic than the isolates associated with the 2003–2004 re-emergence, due to with their binding affinities for hACE2[16].
Fig.1.(a) structure of SARS-COV2 **(b)**structure SARS-COV2 spike glycoprotein obtained from electron microscopy (PDB ID:6VXX)**.**
The viral variants of SARS cov2 with mutations in their spike protein receptor-binding domain (RBD) that drastically improve binding affinity in the RBD-hACE2 complex while also causing fast transmission in human populations[4]. Increased viral



replication increases the probability of mutations in SARS cov2[4]. Most of the evolutionary changes in organisms occur as a consequence of mutations. The mutation is an alteration of the nucleotide sequence in the genome of an organism, these variations bring novelty that we observe in the course of evolution; also, natural selection can act upon these alterations [6]. Usually, small count of these variations is beneficial, some are inconsequential (neutral) and most of them are harmful or non-beneficial for the organisms. Whether an organism has a low mutation rate or high, the count of harmful mutations always outnumbers beneficial mutation[6]. RNA viruses usually have higher mutation rates compared to DNA viruses as well as a million times higher than their hosts; this is the reason for viral versatility[4,6]. Moreover, the ability to engage ACE2 from different animal species appears to reflect host susceptibility to SARS-CoV infection and facilitate the jump of the virus from animals to humans.
The development of SARS Cov2 was marked by the introduction of the terms “variant of interest” (VOI) and “variant of concern” (VOC) in late 2020[4]. The Delta variant (B.1.617.2) was discovered for the first time in late 2020 in India. In June 2021, the World Health Organization stated that the Delta strain is on its way to becoming the most prevalent strain in the world. Therefore, the Delta variant was changed from Variant of Interest to Variant of concern[4]. The Delta VOC mostly affects the individual who are unvaccinated or just partially vaccinated. The World Health Organization's Technical Advisory Group on Virus Evolution (TAG-VE) proposed that variant B.1.1.529, commonly known as Omicron, be identified as a VOC on November26, 2021[4]. The Omicron variant is the most varying strain seen in remarkable numbers so far during the pandemic, raising concerns that it may be linked to greater transmissibility, lower vaccine efficiency, and an increased risk of reinfection. Globally, the number of nations reporting SARS-CoV-2 Omicron VOC infections continues to rise, with a total of 352 confirmed cases reported by 27 countries as of December 1, 2021[4]. The existence of a high number of Omicron variant mutations is also a hallmark of the variants, indicating that viral evolution in immunocompromised persons may have played a significant role in their development[4]. Research is being conducted globally on this novel coronavirus and its variants to gain knowledge about their, evolutionary history, mutagenesis and transmission.
Reducing transmission through established proven disease control methods like avoiding introduction into animal populations are crucial aspects of the strategy to reduce the occurrence of mutations that have negative public implications[6]. The only option to end the pandemic is for effective vaccine development against circulating variations to be considerably and impartially delivered globally[6]. As nations are rushing to vaccinate their people within months, they risk SARS-CoV-2 evolving into a new variant that vaccines may not be able to effective against in other countries[4,6]. To resist some emerging SARS-CoV-2 strains, new vaccines may need to be developed regularly. With the origination of extremely infectious SARS-CoV-2 variants, more effective vaccine will be required to boost immunity[4]. The major function of the immune system, as in the united anatomic system and other systems, is to avoid disease in the human body. The immune system has its own mechanisms for maintaining a general physiological balance in life by co-operating with other systems of the body.[17]. Antibodies are produced as the result of immune induction, and antigens cannot trigger an immune response unless they bind with their corresponding antibodies or receptors.[17]. An antigen in is a entity that binds to a specific antibody or is any molecule or molecular fragment that can be bound by a major histocompatibility complex (MHC) and presented to a T cell receptor (TCR).[17] Antigenicity and immunogenicity, are generally used to describe each antigen. Immunogenicity is the ability to induce a humoral and/or cell-mediated immune response. Antigenicity is the ability to specifically combine with the final products of the immune response.[17]. Immunogenicity is however is a more profound measure of how well a vaccine works and measures the type of immune responses a vaccine produces and their extent over time. Accumulation of large scale immunological data gave rise to the field known as Immunoinformatics, which provides insights into the mechanisms of immune function.[2].
This study is an effort to compare structural similarities and diversifications among coronavirus strains, which can hint towards the susceptible antigen targets of SARS-CoV2 and its variants to come up with the potential therapeutic and precautionary treatment

to put an end to this public threat. Both patient/host response and virus-specific information are crucial in the clinical management of the disease including diagnosis and therapeutics. Recognition of the pathogen key protein by host cells to induce the immune system is of great importance as this is helpful in composing the vaccine. Identifying key pathogenic protein using bioinformatics tool is a viable strategy for designing of vaccine and therapeutic molecule. Further, identification of the acceptable and unacceptable regions is critical for identifying the potential drug targets with therapeutic efficacy. As, the genome and protein sequence information of SARS-CoV-2 is available, characteristics of the virus, as well as the epitopes presented in the pathogen, could be predicted by in silico analysis, which will greatly speed up the vaccine development [2]. By using different bioinformatics applications and immunoinformatics tools prediction of B cell and T cell epitopes on the spike glycoprotein of different coronavirus strains was done. Epitope information generated by this work may be helpful in development of a promising vaccine against SARS-COV2.

Literature Review

Recent pandemic pneumonia outbreak COVID-19, in Wuhan, China is spreading globally and has raised an urgent public health issue worldwide impacting millions of people with a continuous increase in both morbidity and mortality.[6]. The causative agent of this disease was identified and named as SARS-CoV2 because of its genetic relatedness to SARS-CoV species that was responsible for the 2003 coronavirus outbreak. The immense spread of the disease in a very small period demands urgent development of therapeutic and prophylactic interventions for the treatment of SARS-CoV2 infected patients.[6]

Vaccines stimulate the production of antibodies inside the human body before disease generation, in the same manner as antibodies are produced after the individuals are exposed to the disease pathogen. So, drug/vaccine repurposing is the current hot research area and researchers are aiming for the use of potential target antigen sequences of previously known coronavirus strains to come up with a suitable vaccine for novel SARS-CoV2 strain.[6]. Additionally, S glycoprotein is responsible for mediating viral entry into human cells and thus aids in determining host range as well as viral infectivity. Restricting its cleavage by manipulating the expression of S protein or ACE-2 receptor binding motif might be a great therapy. Moreover, S protein is the most potential target antigenic site suggesting that RBD-SARS CoV based vaccines have potential to prevent SARS-CoV2 infection. Therapeutics, Moderna Inovio Pharmaceuticals, Clover Biopharmaceuticals, Novavax, Johnson & Johnson, Codagenix, are a few among the various companies working for the development of SARS-CoV2 vaccines. [6]. Multi epitope vaccines using immunogenic epitopes specific to CD8+ and CD4+ cells and stimulating the immune system against these epitopes simultaneously and completely specifically are among the methods that have been considered in this regard. Conventional methods for producing vaccines are time-consuming and expensive.[25] By having a total map of virus epitopes and their Immunogenicity, it is vital to create an effective vaccine against COVID-19 virus disease. [25].

Although immense global research is being conducted on this novel coronavirus strain, based on the genetic comparison of various corona virus strains we have enlisted a few potential antigen target sites of SARS-CoV2 for the development of vaccines. More research and in-depth understanding of pathogenic mechanisms and genomic variability of SARS-CoV2 will further help in coming up with more targets for better therapy of COVID-19.[6].

Materials and Methods

1. Data retrieval
   Using Uniprot[8] database, the FASTA sequence of SARS-COV2 spike protein was obtained.(Accession no: PDOTC2). The structure of spike protein of delta variant(PDB ID:7TEY) and omicron variant (PDB ID:7T9J) of coronavirus was obtained from RCSB PDB [18]Database.
2. Analysis of physiochemical parameters
   Using ExPASy ProtParam[19], SARS-COV2, Delta, and Omicron variant sequences were analyzed. It calculates the molecular weight, theoretical pI, amino acid composition, anticipated half-life, instability index, aliphatic index, and grand average of Hydropathicity (GRAVY).
3. A Secondary structure prediction and comparison:
   GOR IV[9] was used to predict the secondary structure of the Delta, and Omicron variants. The Garnier–Osguthorpe–Robson (GOR) tool uses information theory and Bayesian statistics to analyze secondary protein structure. The goal of combining multiple sequence alignments using GOR is to gain knowledge for improved secondary structure differentiation.
4. B cell epitope prediction:
   ABCpred[10] servers was employed for B-cell epitope forecast. We used a threshold value of 0.85 to achieve a sensitivity between 95.5% and 99.5% for epitope prediction on ABCpred server. The length of linear B-cell epitopes normally varies from 5 to 30 residues[2]. IEDB[13] server was used to predict epitopes based upon protein structures. For linear epitopes Ellipro method was used and for Discontinuous Epitopes Discotope method was used. Vaxijen2.0 server[12] was utilized to analyze the antigenicity of chosen epitopes[2].
5. T cell epitopes were predicted using the free online service provided by IEDB[13] was used to forecast T-cell epitopes within membrane protein binding to MHC class and MHC class 2. Key features including antigenicity, toxicity, allergenicity, physiochemical properties were analysed via vaxijen 2.0[12], protein digest server, AllerTOP v2.0 server[14], and ToxinPred server[15].
6. Designing multi epitopes based vaccine
   A vaccine design was construct with selected T cell and B cell epitopes and it's physiochemical parameters, antigenicity and allergenicity was analyzed.

Result

1.1 Analysis of physiochemical parameters of the proteins

The spike glycoprotein of SARS-COV2 has 1273 amino acids, where as Delta and omicron have 1286 and 1285 respectively. There is and significant increase in the number of amino acids in both variants when compared to the wild type. According to the amino acid composition of Delta and omicron, there is a decrease in the composition of polar amino acids serine(S), glycine(G), threonine(T) in omicron which indicates its is more hydrophobic than delta variant. Both variants are more hydrophobic when compared to the wild type protein indicating hydrophobicity increase gradually among the variants. There increase in the composition of positively charged arginine (A),lysine(K) and negatively charged asparagine(D),glutamate(E) indicating increase in number charged residues that contribute to salt bridge formation and that the charged residues are exposed to a great extent among the variants, omicron has the more number of residues compared to delta variant.

The molecular weight of the protein and half life of the protein computed by ProtParam is shown**(Table.1.).** Theoretical pI or isoelectric point is value at which the net surface charge of a macromolecules is zero, that is protein is at neutral charge. The variants have high molecular weight and pI value than wild type. The pI value are 6.24<6.56<6.63 of SARS-COV2, delta variant and omicron respectively. That is omicron protein is more alkaline than delta variant and the wild type protein. The stability score of spike protein all SARS-COV2 types indicates structural stability, the value for omicron is 33.09, for delta variant is 31.61,for wild type is 33.01.

The aliphatic index is the measure of aliphatic amino acids(alanine, valine, isoleucine, leucine) filled on the side chains of a protein.[20].The higher the aliphatic index of a protein it is more thermostable. The aliphatic index of wild type SARS-COV2 is 84.67 and delta variant is 81.61 and omicron is 81.17. Thus, the variants are less thermostable than the wild type protein. The GRAVY value of the protein is the sum of hydropathy values of all amino acids, divided by the number of residues in the sequence.[20]. A protein with high grand average of hydropathicity value is polar and has low affinity towards water and is hydrophobic. Thus, wild type spike protein is more hydrophobic than Omicron and Delta variants.

Table.1. Physical parameters of spike protein of SARS-COV2, Delta and omicron predicted by ExPASy ProtParam server.

Physiochemical parameters	**SARS-COV2**	**Delta variant**	**Omicron**
No. Of amino acids	1273	1286	1285
Molecular weight	141178.47	142052.47	142452.30
Theoretical pI	6.24	6.52	6.63
Amino acids composition with high percentage			
Leu(L)	8.5%	8.2%	8.2%
Ser(S)	7.8%	8.2%	8.0%
Gly(G)	6.4%	7.5%	7.1%
Thr(T)	7.6%	7.4%	7.2%
Val(V)	7.6%	7.4%	7.3%
Asp+Glu(-ve)	110	109	113
Arg+Lys(+ve)	103	103	108
Half life	30 hrs	30 hrs	30hrs
Instability index	33.01(stable)	31.61(stable)	33.90(stable)
Aliphatic index	84.67	81.17	81.61
Hydropathicity (GRAVY)	-0.079	-0.183	-0.176

Note: -ve is negatively charged residues, +ve is positively charged residues.

1.2 Secondary structure prediction and comparison

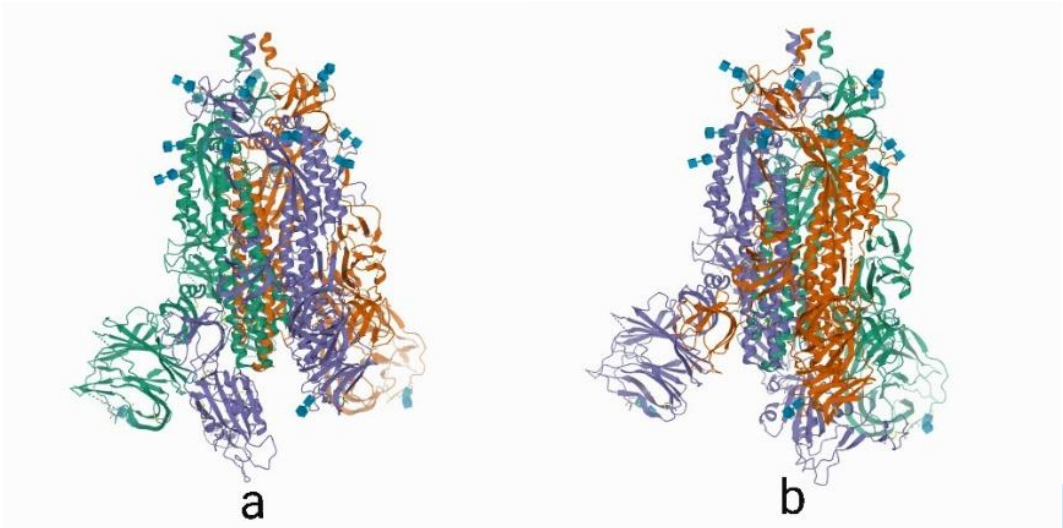


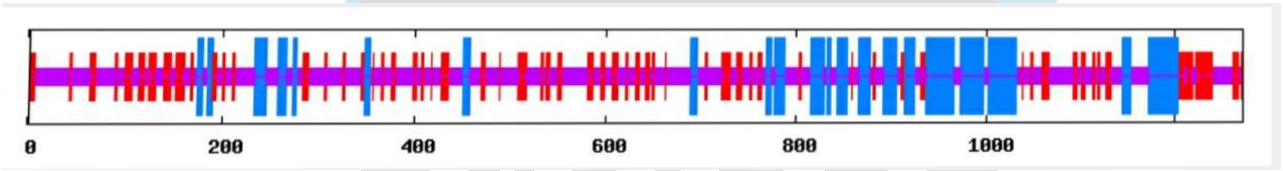
Fig.2.(a) cryo-EM structure of spike protein of delta variant (PDB IB:7TEY). **(b)** cryo-EM structure of spike protein of omicron variant (PDB ID:7T9J).

Table.2. secondary structure prediction and comparison of spike protein of SARS-COV2,Delta variant and Omicron by GOR IV

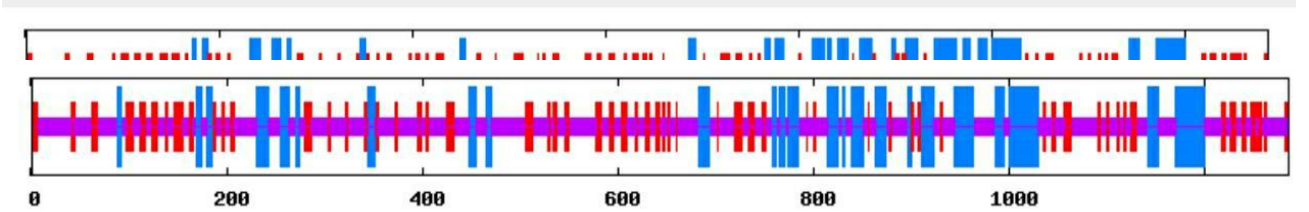
Sequence	**Strain**	**Alpha helix**	**Extended strands**	**Random Coil**
UNIPROT ID: P0DTC2	SARS-COV2	21.5%	22.07%	56.40%
PDB ID: 7TEY	Delta	19.13%	21.15%	59.72%
PDB ID:7T9J	Omicron	19.84%	19.69%	60.47%

Spike protein of Omicron has higher percent of alpha helix compared to delta variant.**(Table.2.)** whereas, delta variant has more number of extended strands compared to omicron. The parentage of alpha helix and extended strands is higher in wild type spike protein compared to the variants. There is a gradual increase in the random coiling which may be due to mutations.

(a) SARS-COV2



(b) Delta variant



(c) Omicron variant

Fig.3. Graphical representation of spike protein by GOR IV. Blue represents the alpha helix, orange represents the extended strands in the structure.

1.3 Linear continuous B cell prediction

A major task in vaccine design is to select and design proteins containing antibody-binding epitopes (B-cell epitopes) which are able to induce an efficient immune response. The selection can be aided by epitope prediction in relevant proteins or regions of proteins. Most existing methods for prediction of B-cell epitopes exclusively use protein sequences as input, and are best suited to predict epitopes composed of a continuous stretch of amino acids. In general, these methods are based on prediction of hydrophilicity, flexibility, b-turns, and surface accessibility using a number of amino acid propensity scales.[21]

Table.3. Linear B cell epitopes predicted using ABCpred online tool along with antigenicity score predicted by Vaxijen 2.0 online tool.

Sr. No.	Epitope Sequence	Position in delta variant PDB ID:7TEY	Position in Omicron PDB ID:7T9J	Vaxijen 2.0 score
1	EVRQIAPGQTGKIADY	404	403	1.3837
2	FPMQMAYRFNGIGVTQ	896	895	1.3601
3	HVTYVPAQEKNFTTAP	1062	1061	0.8933
4	HGVVFLHVTYVPAQEK	1056	1055	0.8847
5	DSLSSTPSALGKLQNV	934	933	0.8754
6	GCLIGAEHVNNSYECD	646	645	0.8480
7	GWTAGAAAYYVGYLQP	255	254	0.6210

In this study, sequence of spike proteins of delta variant and omicron were scanned for putative linear continuous B cell epitopes by using two bioinformatics tools with two different approaches. First, ABCpred [10] predicts B cell epitopes in antigen sequence using artificial neural networks[22]. Second, IEDB server Ellipro, a web-tool that implements a method for identifying continuous epitopes in the protein regions protruding from the protein's globular surface and together with a residue clustering algorithm, the MODELLER program and the JSmol viewer, allows the prediction and visualization of antibody epitopes in a given protein sequence or structure.[23]. Using ABCpred[10], a total number of non overlapping 26 and 27 epitopes were predicted in delta variant and omicron respectively with threshold set at 0.85. Antigenicity was calculated using Vaxijen 2.0 server and peptide present in both strains and with highest antigenicity score were selected **(Table.3.)**.

Table.4. Linear B cell epitopes prediction for spike protein of delta variant (PDB ID:7TEY) using IEDB server(Ellipro method) along with antigenicity score.

Chain	Start	End	Peptide	No. Of residues	Scores	Vaxijen 2.0 score
C	702	718	ENSVAYSNNSIAIPTNF	17	0.692	0.4481
C	553	564	TESNKKFLPFQQ	12	0.69	0.7903
C	577	585	RDPQTLEIL	9	0.664	0.7101
C	519	524	HAPATV	6	0.548	0.4273
C	528	537	KKSTNLVKNK	10	0.514	0.7187

Table.5. Linear B cell epitopes prediction for spike protein of omicron (PDB ID: 7T9J) using IEDB Server (Ellipro method) along with antigenicity score.

Chain	Start	End	Peptide	Length	Score	Vaxijen 2.0 score
C	14	29	QCVNLTTRTQLPPAYT	16	0.828	1.4165
C	702	718	ENSVAYSNNSIAIPTNF	17	0.698	0.4481
C	553	564	TESNKKFLPFQQ	12	0.678	0.7903
C	557	585	RDPQTLEIL	9	0.643	0.7101
C	206	218	KHTPIDLPQ	9	0.592	0.7129
C	519	524	HAPATV	6	0.536	0.4273
A	207	216	HTPIDL	6	0.529	1.0830

The structure of SARS-CoV-2 spike protein resolved with Cryo-electron microscopy (cryo-EM) which could greatly facilitate the process of vaccine development was obtained from PDB database. ElliPro accepts two types of input data: protein sequence or structure.[23].In this study, the second approach was implemented PBD file of structure of spike protein of delta variant(PDB ID:7TEY, chain A,B,C) and omicron (PDB ID:7T9J chain A,B,C) was submitted. ElliPro implements three algorithms performing the approximation of the protein shape as an ellipsoid, calculation of the residue protrusion index (PI) and clustering of neighboring residues based on their PI values.[23]. Peptide sequence with length between 5 to 30 peptides were selected and antigenicity was calculated using Vaxijen 2.0 server. Peptide sequences with highest antigenicity score were selected shown in **(Table.4 and Table.5.)**.

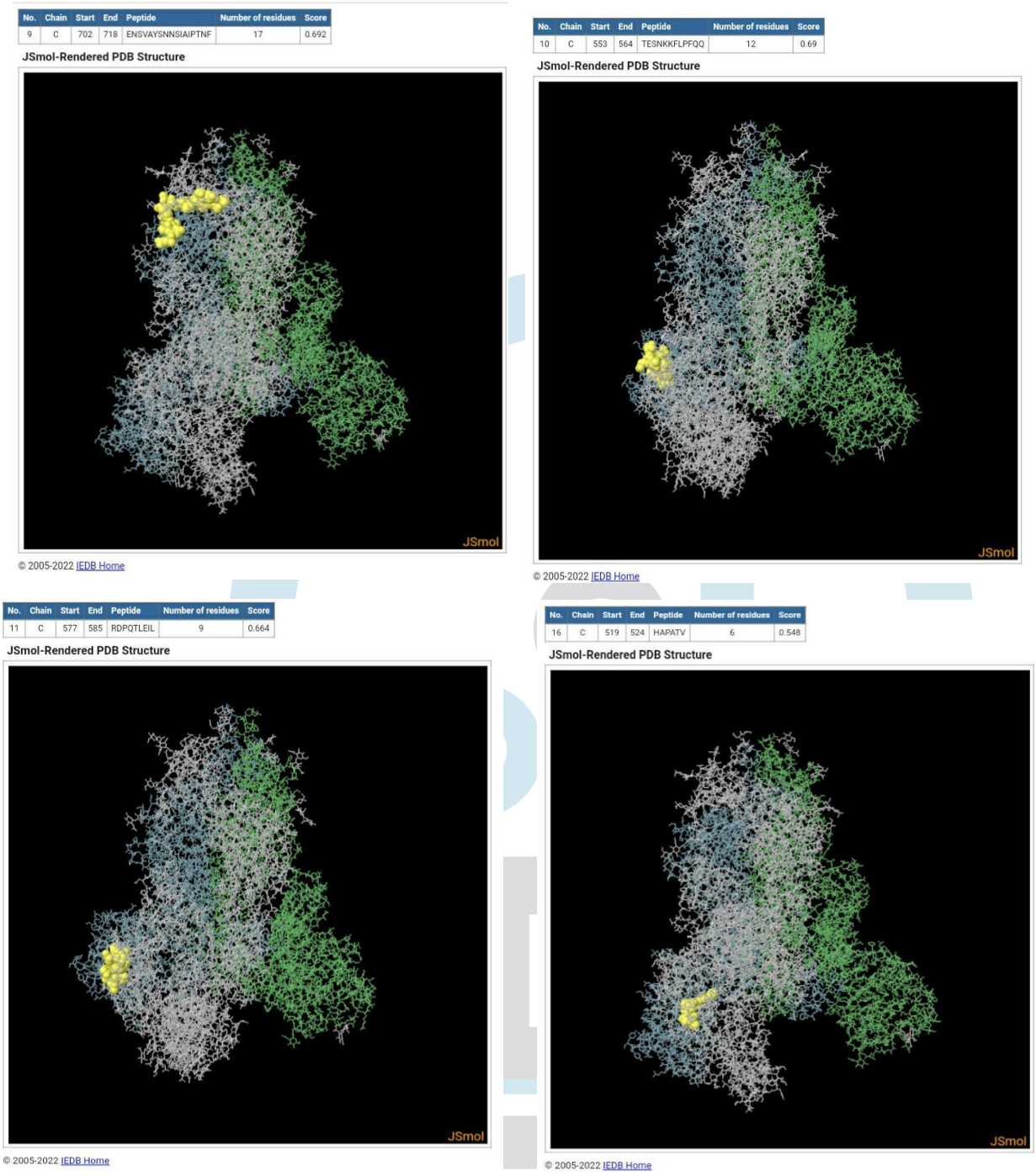
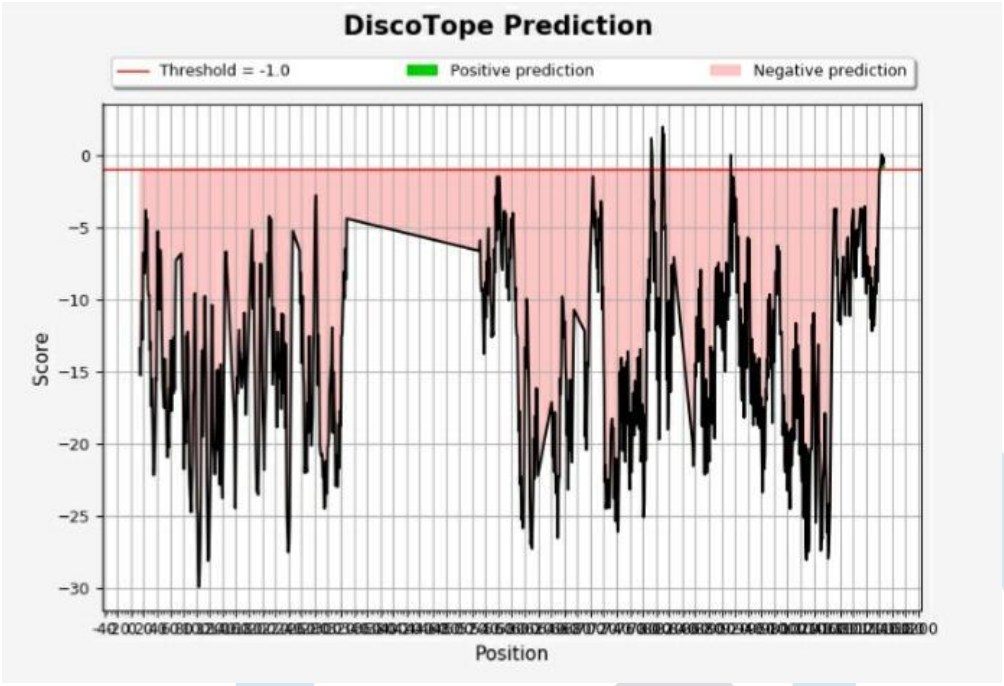
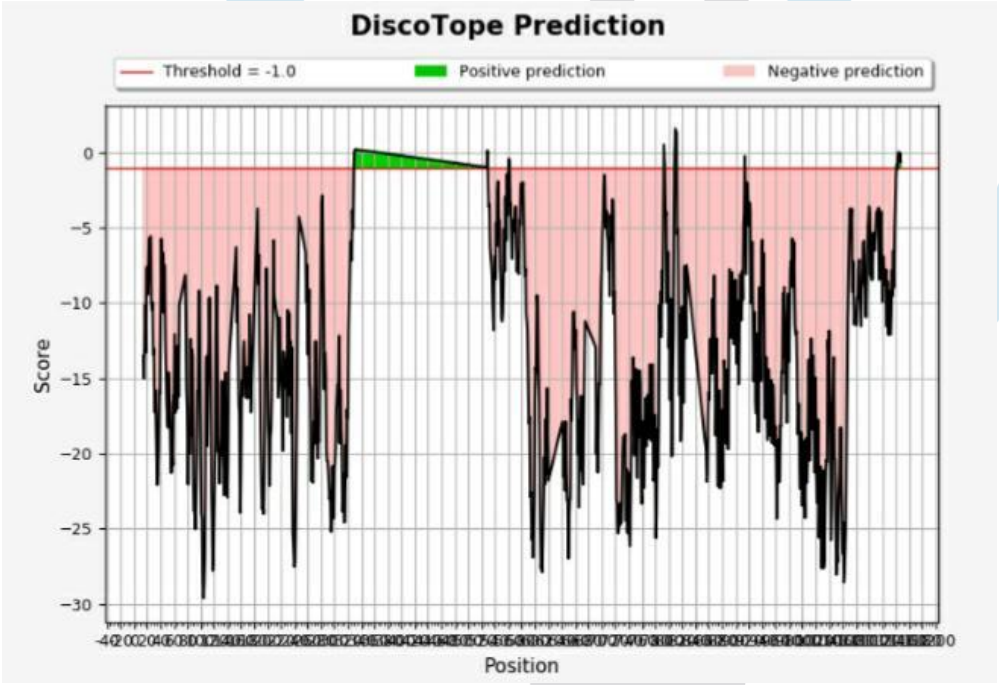


Fig.4. JSMOL render 3D view of spike protein where selected B cell epitopes are shown as yellow spheres.

1.4 Discontinuous B cell epitopes

(a)





(b)

Fig.5. Chart view of the discontinuous epitopes of B chain **(a)**spike protein of delta variant **(b)** spike protein of omicron.

Table.6. Discontinuous B cell epitopes prediction of spike protein (B chain)of Delta variant using Discotope method

Sr.no	Residue ID	Residues	Number of Contacts	Propensity score	Discotope score
1	793	PRO	0	1.294	1.145
2	794	ILE	4	0.448	-0.064
3	809	PRO	8	0.891	-0.131

4	810	SER	5	2.85	1.948
5	812	PRO	0	1.608	1.423
6	914	ASN	7	0.896	-0.012
7	1141	LEU	3	-0.615	-0.889
8	1141	GLN	8	0.201	-0.742
9	1143	PRO	6	0.496	-0.251
10	1144	GLU	4	0.548	-0.025
11	1145	LEU	4	-0.254	-0.685
12	1146	ASP	5	0.404	-0.218
13	1147	SER	5	0.13	-0.46

Discontinuous B cell epitopes were predicted using discotope method on IEDB server for delta variant as well as omicron variant. DiscoTope uses a combination of amino acid statistics, spatial information, and surface exposure. It is trained on a compiled data set of discontinuous epitopes from 76 X-ray structure of antibody/antigen protein complexes.[21]. A threshold value of − 1.0 was taken for the computation, which corresponds to a specificity of 85%and a sensitivity of 30%. The contact number, propensity score, and discotope score for each amino acid that passed the threshold were presented separately for delta variant and omicron in **Table.6 and Table .7** respectively. Total 13 residues were predicted for delta variant where as 18 residues were predicted for omicron by discotope method.

Table.7. Discontinuous B cell epitopes prediction of spike protein(B chain) of omicron using Discotope method.

Sr. No	Residue ID	Residues	Number of contacts	Propensity score	Discotope score
1	330	PRO	4	0.595	0.067
2	331	ASN	7	1.165	0.226
3	528	LYS	17	1.121	-0.963
4	529	LYS	2	0.425	0.146
5	561	PRO	0	-0.454	-0.402
6	562	PHE	0	-0.884	-0.783
7	793	PRO	0	0.605	0.535
8	794	ILE	7	0.082	-0.733
9	809	PRO	8	0.584	-0.403
10	810	SER	7	2.716	1.593
11	812	PRO	0	1.571	1.39
12	914	ASN	7	0.654	-0.226
13	1141	LEU	2	-0.662	-0.815
14	1143	PRO	6	0.474	-0.27
15	1144	GLU	4	0.555	0.031
16	1145	LEU	4	0.029	0.434
17	1146	ASP	5	0.639	-0.009
18	1147	SER	5	-0.033	-0.604

T cell epitopes prediction

Tcell epitopes were predicted using IEDB server The Immune Epitope Database and Analysis Resource (IEDB) is a freely available resource that contains an extensive collection of experimentally measured immune epitopes and a suite of tools for predicting and analyzing epitopes.[24]. Even a small virus can result in tens of thousands of peptide fragments as a result of processing by a cell's MHC class I pathway, experimentalist can rarely afford to measure each of them. Machine learning approaches can develop a function that predicts affinity binding for a given peptide sequence[24]. User can select a preferred method from a list including IEDB recommended, consensus, netMHCpan, ANN, scoring matrix method (SMM), SMMPMBEC, Combinatorial Library, PickPocket, and netMHCcons. The IEDB recommended method is the default setting and usually is consensus, a combination of three different methods (ANN, SMM, and Combinatorial Library).[24]

Table.8. MHC Class 1 T cell epitopes prediction of spike protein using IEDB server

Sr.no	Peptide	Percentile rank	MHC Restrictions	Vaxijen 2.0 score
1	GVVFLHVTY	0.33	HLA A*32:01,HLA A*1101,HLA B*46:06	1.4104

2	VGYLQPRTF	0.26	HLA A*02:01,HLA A*24:02,HLA B*51:01,HLA B*07:02,HLA*18:01,HLA B*15:01,HLA B*46:01	1.2441
3	VLKGVKLHY	0.11	HLA A*29:02,HLA B*15:01,HLA B*46:01	1.2378
4	QLTPTWRVY	0.26	HLA A*32:01,HLA A*30:02,HLA A*02:01,HLA B*07:02,HLA B*15:01,HLA B*46:01	1.2119
5	TLDSKTQSL	0.1	HLA A*02:01,HLA*08:01,HLA B*07:02,HLA B*13:01	1.0685
6	YSKHTPINL	0.26	HLA A*24:02,HLA B*46:01	1.0547
7	FAMQMAYRF	0.08	HLA B*51:01,HLAB*57:03,HLA B*35:01,HLA B*46:01	1.0278
8	ESNKKFLPF	0.72	HLA B*44:02,HLA B*46:01.	1.0278

In this study spike protein as well as membrane protein of both variants were scanned for the prediction of T cell epitopes both for MHC class 1 and class 2 binding affinity. The IEDB recommended method as default setting was applied.

Table.9. MHC Class 1 T cell epitopes prediction of membrane protein using IEDB server.

Sr.no.	Peptide	MHC Restrictions	Percentile rank	Vaxijen 2.0 score
1	AGDSGFAAY	HLA A*01:01,H B*15:01,HLA B*30:02,HLA B*46:01,HLA class 2	0.88	0.9095
2	SELVIGAVI	HLA B*44:02,HLA B*45:01,HLA B*40:01,HLA B*44:03,HLA B*40:02,HLA B*18:01,HLA class 2- HLA DRB1*01:01,HLA DRB1*09:01,HLADRB1*13:02,HLA DRB1*15:01,HLA DRB1*12:01,HDRB1*07:02,HLA DRB5*01:01,HLADRB3*02:02,HLA DRB3*01:01	0.69	0.6409
3	ATSRTLSYY	HLA A*01:01,HLA B*15:02,HLA B*11:01,HLA B*57:01,HLA B*01:01,HLA B*29:02,HLA A*03:01,HLA A*30:02,HLA A*68:01,HLA A*26:01,HLA A*31:01,HLA A*30:01,HLADQB1*05:03,HLA DRB1*16:01,HLA DQB1*05:01	0.3	0.6108
4	ITVATSRTL	HLADRB1*15:01,HLADRB1*07:01,HLA DRB1*03:01,HLADRB1*06:03,HLA DRB1*16:02,HLADRB1*14:01,HLA DRB1*14:06,HLADRB1*04:03,HLA DRB1*09:01,HLADRB1*11:01,HLA DRB1*13:02,HLADRB3*02:02,HLA DRB3*12:01,HLA DRB4*01:01	0.46	0.5800

MHC class 1 prediction of spike protein and membrane protein was performed separately using IEDB server. A total number of 276 peptide sequence for spike protein and a total number of 86 of membrane protein were predicted by IEDB server. These peptides were analyzed using tepitool provided by IEDB server for T cell epitopes prediction. Antigenicity of peptide sequence was calculated using Vaxijen 2.0 for spike protein 44 peptide sequence and for membrane protein 8 peptide sequences were found antigenic. Allergenicity was calculated using AllerTOP and toxicity was calculated using ToxinPred and peptide sequence that are non allergen, non toxic and with highest antigenicity score were selected and shown **(Table.8.)** for spike protein and **(Table.9)** for membrane protein. Immunogenicity of all the selected peptide sequence was predicted on IEDB server and shown **(Table.10.).** This tool on the IEDB server uses amino acid properties as well as their position within the peptide to predict the immunogenicity of a MHC class 1 complex.

Table.10. MHC Class 1 T cell epitopes Immunogenicity score using IEDB server

Sr.no.	Peptide(protein type)	Immunogenicity score
1	GVVFLHVTY (spike)	0.20837
2	VGYLQPRTF (spike)	-0.06924
3	VLKGVKLHY (spike)	-0.18916
4	QLTPTWRVY (spike)	0.31555

5	TLDSKTQSL (spike)	-0.52716
6	YSKHTPINL (spike)	0.09845
7	FAMQMAYRF(spike)	-0.28061
8	ESNKKFLPF (spike)	-0.33474
9	AGDSGFAAY (membrane)	0.03981
10	SELVIGAVI (membrane)	0.25658
11	ATSRTLSYY (membrane)	-0.11604
12	ITVATSRTL (membrane)	0.0012

MHC class 2 prediction of spike protein and membrane protein was performed separately using IEDB server. A total number of 311 sequence for spike protein and a total number of 82 of membrane protein were predicted by IEDB server. These peptides were analyzed using tepitool provided by IEDB server for T cell epitopes prediction. Non overlapping peptide sequence with threshold percentile rank below 10 were analyzed using this tool. Antigenicity of peptide sequence was calculated using Vaxijen 2.0 for spike protein 14 peptide sequence and for membrane protein 5 peptide sequences were found antigenic. Allergenicity was calculated using AllerTOP, The method is based on auto cross covariance (ACC) transformation of protein sequences into uniform equal-length vectors.[14]. Even the toxicity of the peptide sequence was calculated using ToxinPred. The peptide sequence that are non allergen, non toxic and with highest antigenicity score with their HLA alleles 6 peptide sequences were selected of spike protein and 3 peptide sequences were selected of membrane protein and are shown **(Table.11.)** for spike protein and **(Table.12.)**for membrane protein.

Table.11. MHC Class 2 T cell epitopes prediction of spike protein using IEDB server.

Sr.no.	Peptide	Percentile rank	MHC Restriction	Vaxijen 2.0 score
1	IPFAMQMAYRFNGIG	9.10	HLA DRB1*15:01,HLA DRB1*12:01,HLA DQB1*04:02,HLA DQB1*05:03,HLA DRB1*14:01	1.2828
2	NLLLQYGSFCTQLNR	0.75	HLA DRB1* 04:04,HLA DQB1*05:03,HLA-DRB1*15:01,HLA DRB1*07:01,HLA DRB1*04:01,HLA DRB1*08:02,HLA DRB1*03:01,HLA RB1*13:02,HLA DRB3*02:02,HLA DRB1*04:05,HLA DR1* 09:01,HLA DRB1*01:01,HLA DRB1*12:01,HLA DRB5*01:01,HLA DRB3*01:01	0.8668
3	VVLSFELLHAPATVC	6.10	HLA DRB1*14:01,HLA DQB1*05:03,HLA DRB1*12:01	0.8618
4	IGINITRFQTLLALH	3.60	HLA DRB1*07:01,HLA class 1	0.8391
5	RAAEIRASANLAATK	9.60	HLA DRB1*11:01,HLA DRB3*01:01,HLA DRB5*01:01,HLA DRB1*01:01HLA DRB1*01:04,HLA DRB1*07:01,HLA DRB1*15:01,HLA DRB1*03:01,HLA DRB1*04:04,HLA DRB1*04:01	0.7502
6	FKIYSKHTPINLVRD	9.20	HLA DRB1*13:01,HLA DRB1*07;01,HLA DRB1*01:01,HLA DRB1*03:01,HLA DRB1*04:01,HLA DRB1*15:01,HLA DRB1*11:04,*HLA DRB4*01:01,HLA DRB5*01:01,HLA DRB3*01:01	0.7292

Table.12. MHC class 2 T cell epitopes prediction of membrane protein using IEDB server.

Sr.no.	Peptide	Percentile rank	MHC Restrictions	Vaxijen 2.0 score
1	NLVIGFLFLTWICLL	6.70	HLA DRB1*01:02,HLA DQB1*05:01	1.2609
2	SELVIGAVILRGHLR	9.20	HLA DQB1*06:02,HLA DRB1*14:01,HLA DRB1*13:01,HLA DRB1*01:02,HLA DRB1*15:01,HLA DRB1*16:01,HLA DRB1*03:01,HLA	0.6768

			DRB1*08:02,HLA DRB1*11:01,HLA DQB1*06:03,HLA DRB3*02:02HLA DRB3*01:01	
3	LRIAGHHLGRCDIKD	2.60	HLA DRB1*04:04,HLA DRB3*01:01,HLA DPB1*04:01,HLA DRB1*01:01,HLA DRB1*04:01,HLA DRB1*15:01,HLA DRB1*01:01,HLA DRB1*07:01,HLA DRB1*03:01,HLA DRB1*01:01,HLA DRB1*11:04,HLA DRB1*11:01	0.4163

1.5 Designing multi epitope based vaccine

Identification of antigenic epitopes by the immune system is a key step in the immune response against the pathogen, identifying either epitopes that stimulate T cells or epitopes that are trapped by B cells. To induce both humoral and cell mediated immune response. In this study, T cell and B cell epitopes present in two coronavirus variants were selected based on antigenicity, allergenicity and toxicity. One T cell epitopes each of spike protein and membrane with affinity to MHC class 1 and one T cell epitopes of spike protein and one epitope of membrane protein with affinity towards MHC class 2 were selected. Four B cell epitopes which are present at same position in both variants were selected. Total of 8 epitopes were selected to construct a vaccine These epitopes were connected to design vaccine structure by AAY, GPGPG and KK linkers.

Human beta- defensin(hBD) increase the acquired immune response by chemically absorbing activity for monocytes, T cells and dendritic cells, and the activity of inducing cytokine production by monocytes and epithelial cells [25]. Accordingly, human beta-defensin 3 and 2 were added as adjuvants to the N and C ends of the designed structure by the EAAAK linker, respectively. EAAAK linker, due to its salt bridge related to glutamic acid and lysine, can prevent protein domains from converging by creating a stable helix structure [25].

The FASTA sequence for both hBD3(accession no. NP_001075020.1)and hBD2(accession no.AAC69554.1) were obtained from NCBI database[26]. The final vaccine construct consist of 252 amino acids. The antigenicity of the designed vaccine was and was analyzed using Vaxijen 2.0 server and was found to be 0.6183. The allergenicity and toxicity was checked using AllerTOP and ToxinPred, the design was found to be non allergenic and non toxic. The physiochemical parameters such molecular weight which is 27903.27, theoretical pI is 9.93, aliphatic index is 85.16, Hydropathicity which is (-0.060) of the designed vaccine were computed using ExPASy ProtParam server. The vaccine construct was found to be stable with instability index of 38.85 and the half life was 30 hours in mammalian reticulocytes, invitro. The final construct is shown dramatically in **Fig.6.**

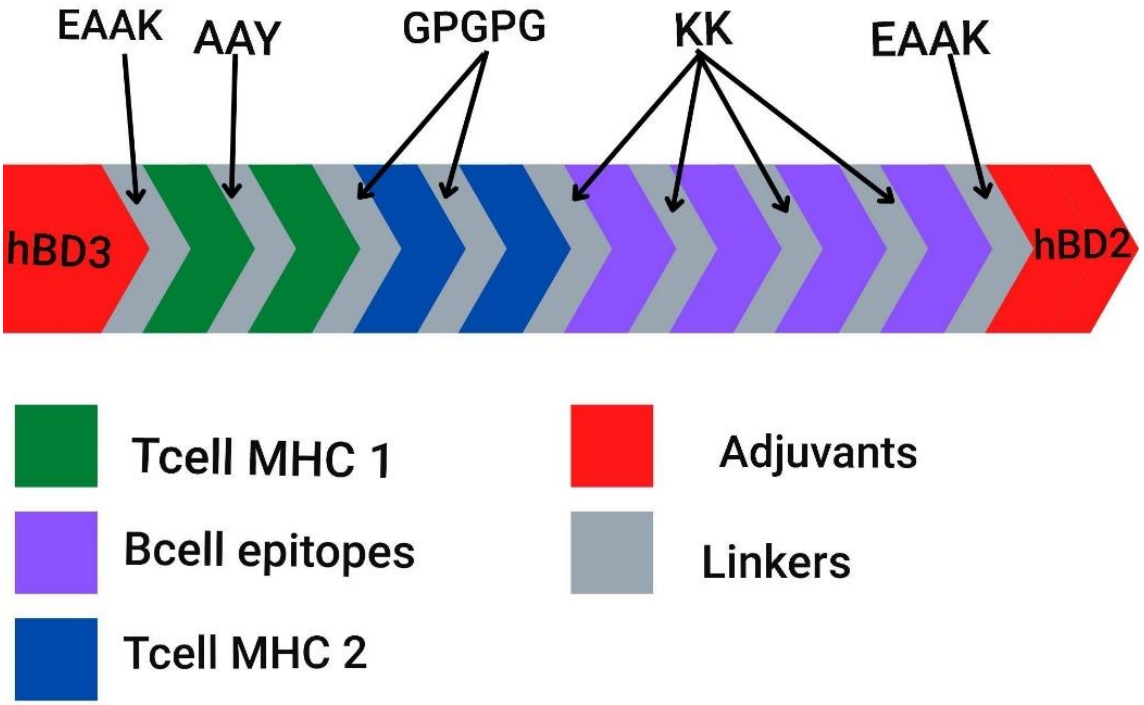


Fig.6. Diagrammatic representation of multi epitopes based vaccine design.

Discussion

Different from the less harmful human coronaviruses continuously circulating in the human population, effective and economically preventive approaches are in need urgently at the current situation of pandemic. Understanding the molecular and evolutionary origins of SARS-CoV-2, the underlying mechanisms of viral–host binding interaction, and the detection of potential antiviral peptides and epitope vaccine candidates as potential therapeutic choices against coronaviruses has improved our knowledge of the coronavirus pathogenesis[3]. The development of a vaccine is a multidisciplinary endeavour that combines molecular knowledge of host–pathogen interactions, selection of antigens to deliver an immune response, formulation aspects, and pre-clinical and clinical testing of the developed vaccine to assure an optimum therapeutic efficacy and safety to human[3]. Compared to traditional vaccine development, potent epitopes can be predicted via bioinformatics analysis, which makes the vaccine design straightforward and fast[2]. As the majority of spike protein is exposed outside the virion, it could be an ideal target to search for B-cell epitopes[2]. Besides B-cell epitope prediction on spike protein, we selected Membrane protein as a target protein for computation analysis as it is the most abundant protein present on the surface of the virus and is involved in virion assembly in the cells and it can generate a strong cellular immune response[6]. The future of immunological research is sharpened by the ability to make discoveries in biologics more effectively and efficiently using bioinformatics[1]. This will mean reduction and better targeting of wet laboratory experiments and will only be possible if wet laboratory experimentation is combined with bioinformatics techniques[1]. These advancements have been supplemented by the emergence of novel computational databases and tools specifically for coronavirus research, which have not only strengthened research strategies to prevent COVID-19 but also aimed to consolidate the massive volume of genomic data and relevant research observations on freely accessible centralized platforms to disseminate to the wider scientific community worldwide[3].

Conclusion

In this study, antigenic target site were predicted by comparing proteins of two variants of coronavirus and selected epitopes were used to design a multi epitopes based vaccine construct. Using different bioinformatics tools physiochemical parameters and structural similarities of spike protein of both variants were compared, they were almost similar. B cell epitopes on spike protein and T cell epitopes on spike as well as membrane protein were predicted and analyzed. Linear continuous B cell epitopes were predicted using ABCpred and Ellipro program. 5 peptide sequence in delta variant and 7 peptide sequence in omicron were identified with highest antigenicity score. Discontinuous B cell epitopes within spike protein were predicted using discotope program and 13 residues in delta variant and 18 residues in omicron were identified that crossed the set threshold. T cell were predicted using IEDB server in both spike as well as membrane proteins. Peptide sequence were predicted and analyzed for both MHC class 1 and MHC class 2 binding affinity. In spike protein 8 peptide sequences and in membrane protein 4 peptide sequence were identified with highest antigenicity score for MHC class 1 binding affinity where as 6 peptides for spike protein and 3 peptides for membrane protein were identified with highest antigenicity score for MHC class 2 binding affinity.

Based on there position, antigenicity, allergenicity, toxicity and peptide that were conserved in both variants were selected for designing the vaccine construct. . Finally, a different multi-epitope vaccine with 251 amino acid was designed. It consists of two adjuvants, with 4 B cell epitopes, 2 T cell epitopes with MHC class 1 affinity and 2 T cell epitopes with MHC class 2 affinity. It displays good antigenicity and satisfactory physiochemical characteristics, non-allergenicity and non-toxicity. It is expected that the epitopes predicted in this study would be an efficient vaccine formation against COVID-19. However, the conformation of the epitopes which were selected in this study as a vaccine candidate should be considered as laboratory studies.

Reference

1. Angus Nnamdi Oli, Wilson Okechukwu obialor and Gorden C Ibeanu Immunoinformatics and vaccine development: An overview Immunotargets and Therapy 2020 ,Dove press https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7049754/
2. Hong Zhi Chen, Ling Li Tang, Xin Ling Yu. Bioinformatics analysis of epitope based vaccine design against the novel SARS COV2. Infectious diseases of poverty 2020 https://idpjournal.biomedcentral.com/articles/10.1186/s40249-020-00713-3
3. Rahul Chatterjee et.al. Next generation bioinformatics approaches and resources for coronavirus vaccine discovery and development: A perspective review Vaccines 2021 https://pubmed.ncbi.nlm.nih.gov/34451937/
4. Suresh Kumar et.al. Omicron and Delta variant of SARS-COV2: A comparative computational study of spike protein by dept. of Diagnostic and Allied Health science. Faculty of health and life sciences. J Med Viral 2021 https://pubmed.ncbi.nlm.nih.gov/34914115/#:~:text=We%20used%20computational%20studies%20to,domain%20(RBD)%2C%20indicating%20a
5. Delphine Planas et.al. Reduced sensitivity of SARS-COV2 variant delta to antibody neutralization by Article in Nature vol 596, 12 August 2021. https://www.nature.com/articles/s41586-021-03777-9
6. Navpreet Kaur et.al. Genetic comparison among various coronavirus strains for the identification of potential vaccine targets of SARS-COV2 by dept. Of Biophysics Punjab university, Chandigarh, India. Infection, Genetics and evolution 2020. https://pubmed.ncbi.nlm.nih.gov/32745811/
7. WHO| Novel Coronavirus (2019-nCoV) Situation Report-128. Report 2020. Available from: https://www.whoint/docs/default-source/coronaviruse/situation-reports/20200527-covid-19-sitrep-128pdf.
8. Uniprot database https://www.uniprot.org/
9. GOR IV https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html
10. ABCpred server http://crdd.osdd.net/raghava/abcpred/

11. BepiPred https://services.healthtech.dtu.dk/service.php?BepiPred-2.0
12. Vaxijen 2.0 server http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html
13. IEDB https://www.iedb.org/
14. AlletopV2.0 server https://www.ddg-pharmfac.net/AllerTOP/method.html
15. ToxinPred http://crdd.osdd.net/raghava/toxinpred/
16. Alexandra C. Walls et.al. April 16 2020. Structure, Function and Antigenicity of SARS-COV2 spike glycoprotein by Department of Biochemistry, University of Washington, Seattle, WA 98195, USA https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7102599/
17. Jianguo Zhang and Ailin Tao, 2015, Antigenicity, Immunogenicity and Allergenicity, Guangdong Provincial Key Laboratory of Allergy and Clinical Immunology, Allergy Bioinformatics, Translational Bioinformatics https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7123983/#:~:text=Allergenicity%20refers%20to%20the%20ability,function%20disorder%20or%20tissue%20damage.
18. RCSB PDB https://www.rcsb.org/
19. ExPASyrotParam https://web.expasy.org/protparam/#:~:text=ProtParam%20(References%20%2F%20Documentation)%20is,a%20user%20entered%20protein%20sequence.
20. Gasteiger E. et.al. Protein identification and analysis tool on ExPASy server. The protein protocols handbook ,Humana press 2005. https://web.expasy.org/protparam/protparam-doc.html
21. PERNILLE HASTE ANDERSEN, MORTEN NIELSEN, AND OLE LUND Prediction of residues in discontinuous B cell epitopes using 3D structure
    Center for Biological Sequence Analysis, BioCentrum, Technical University of Denmark.2006. https://pubmed.ncbi.nlm.nih.gov/17001032/
22. Sudipto saha et.al Proteins 2006 Nov. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network
    Institute of Microbial Technology, Chandigarh, India. PMID: 16894596 https://pubmed.ncbi.nlm.nih.gov/16894596/
23. Julia Ponomarenko et.al. ElliPro: a new structure-based tool for the prediction of antibody epitopes University of California, USA. Dec 2008 PMID: 19055730 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2607291/
24. Ward Fleri et.al. The immune epitope Database and Analysis Resource in epitope Discovery and Synthetic vaccine Design. Division of Vaccine Discovery, La Jolla Institute for Allergy and Immunology, La Jolla, CA, USA 2017 https://www.frontiersin.org/articles/10.3389/fimmu.2017.00278/full
25. Arian Karimi Rouzbahani et.al. Design of a multi-epitope-based peptide vaccine against the S and N proteins of SARS-COV-2 using immunoinformatics approach Egyptian Journal of Medical Human Genetics. 2022 https://jmhg.springeropen.com/articles/10.1186/s43042-022-00224-w
26. NCBI database https://www.ncbi.nlm.nih.gov/
27. PSIPRED server http://bioinf.cs.ucl.ac.uk/psipred/