AI-Based Computational Vaccine Design against SARS-CoV-2

Pratik Sambhaji Sawai[1], Deekshitha P[2], Vaishnavi pawar,[3] Bhoomika Varshney[4], Hridya Ramesh[5], Adinath Palve[6], VipinHiremath[7]
[1,3,5]*Chhatrapati Sambhajinagar university Navi Mumbai*
[2]*Visvadyalayam Tirupati, Andhra Pradesh*
[4,6,7]*Progenome Life Science*

***Abstract*—The cases of highly pathogenic viruses such as zika virus, Ebola virus, Marburg virus, and SARS-CoV-2 are increasing constantly and posing serious threats to global public health and highlighting the critical need for quick, safe, and efficient preventive measures against these diseases [1]. Vaccination remains one of the most effective strategies for preventing viral infections. However, conventional vaccine development techniques are costly, time-consuming, and limited by stringent biosafety regulations [2]. Due to the integration of artificial intelligence with immune-informatics, it becomes a promising and effective method for the vaccine design.**

Therefore, we are focusing the multi-epitope vaccine design against SARS-CoV-2. For multi-epitope vaccine design, the membrane protein of SARS-CoV-2 was selected as the target due to its high antigenic potential and non-allergenic nature [3] Physicochemical properties of the protein were analyzed using ExPASy ProtParam, while secondary and tertiary structures were predicted using PSIPRED and trRosetta, respectively. Cytotoxic T-lymphocyte (MHC class I) and helper T-lymphocyte (MHC class II) epitopes were predicted using the Immune Epitope Database (IEDB). The selected epitopes were assembled into a multi-epitope vaccine construct using suitable linkers and an adjuvant to enhance immunogenicity [4].

Subsequently, the designed vaccine construct underwent structural assessment, molecular docking analysis using ClusPro, and immune simulation using C-ImmSim. The results demonstrated the potential of the vaccine construct to elicit a strong immune response. Overall, this study provides a computationally validated framework for the rational design of multi-epitope vaccines against high-risk viral pathogens and establishes a foundation for future experimental validation.

***Index Terms*—Vaccine, Multi-Epitopes, Artificial Intelligence, Immuno-informatics, Antigenicity, Allergenicity**

I. INTRODUCTION

Covid-19 is one of the life-threatening infectious diseases, caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) [5]. The first cluster of pneumonia cases associated with SARS-CoV-2 was reported in December 2019 in Wuhan, China. As of now, WHO (world health organization, (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports), reported 410 million cases with 5.8 million reported deaths

This virus has spread globally very fast and generated a lockdown condition throughout the world because the spread rate of this virus was very fast; approximately 5.8 million deaths are reported due to this virus. This virus not only impacts health but also impacts the financial, social, and political landscape throughout the world [6,7,8].

Coronavirus is responsible for covid-19 diseases, in 2019 this is known as novel coronavirus (2019-nCoV) or the severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) [9]. Rapid mutations in the viral genome have led to the emergence of multiple variants with altered transmissibility and immune-evasion capabilities. The list of variants is mention in table 1.1

Sno.	Name of Variants	First Origin	Impact	Reference
1	Alpha	US	Highly Transmitted virus	[10]
2	Beta	South Africa	Increase the immune escape	[11]
3	Gamma	Brazil	Highly transmitted and reinfection potential	[12]
4	Delta	India	Very fast transmit throughout the world	[14]
5	Omicron	South Africa/Botswana	Have no of variants such as BA Lineages • BA.1 • BA.2 • BA.3 • BA.4 • BA.5 Recombinant and Descendant Lineages • XBB • XBB.1.5 • XBB.1.16 (Arcturus) • EG.5 (Eris) • HV.1 • JN.1 and strong immune evasion	[15]

The corona virus is an enveloped virus with positive sense single- standard RNA (+ssRNA) genome of approximately 29.9 kbp belongs to the Beta-coronavirus genus with different feature such as Replicase Genes (ORF1a , ORF1b) , structural gene ([S, E,M,N (Spike protein, Envelope protein membrane protein and nucleocapsid protein respectively)] and Accessory Genes (ORF3a,ORF6,ORF7a,ORF7b,ORF8,ORF9b,ORF10)[16] Among all these, membrane protein plays a very crucial role in viral assembly and immune recognition[17]. Due to the rapid mutation of the SARS-CoV-2 genome, several new variants have emerged over time. These variants may affect the effectiveness of existing therapeutic and preventive measures. Therefore, there is a need to develop broad-spectrum and adaptable vaccines that can provide protection against multiple variants of the virus [18].

Although several conventional vaccine development approaches, such as live-attenuated and inactivated (heat-killed) vaccines, have been successfully employed against various infectious diseases, these methods possess certain limitations. Live-attenuated vaccines may not be suitable for immune-compromised individuals and pregnant women due to the potential risk of reversion to a virulent form. In contrast, inactivated vaccines often require multiple booster doses to maintain long-term immunity [19]. Furthermore, both approaches are generally time-consuming, expensive, and require extensive laboratory experimentation, biosafety measures, and clinical validation. Additionally, the continuous emergence of new SARS-CoV-2 variants may reduce the effectiveness of existing vaccines [20]. Therefore, the development of rapid, cost-effective, and adaptable vaccine design strategies has become increasingly important.

Due to the emergence of artificial intelligence (AI) and Bioinformatics in immunology have revolutionized this field and develop and computational approach such as epitope based, multi-epitopes, reverse vaccinology and structured based designing of vaccine. In immunoiformatics, we use different web server, software and immunological database to predict antigenic protein, B-cell epitopes, T-cell epitopes, allergenicity, allergenicity and population coverage, population conservancy, thereby reducing the dependence on extensive laboratory experimentation [21].

Furthermore, by using artificial intelligence and learning -based algorithms, we can easily analyse large biological datasets, to predict the 3d structure of vaccine, docking with ligand, simulation study and expression analysis with high accuracy, these computational approaches significantly reduce the time, cost, and resources required for vaccine designing and become valuable tools for combating rapidly evolving pathogens such as SARS-CoV-2[22].

For designing a vaccine, there are different computation approaches are available such reverse

vaccinology, epitope-Based Vaccine Design, multi-epitopes- based vaccine design structure-based vaccine designing approaches, AI/ML based vaccine design etc. In multi-epitope vaccines designing we combine multiple cytotoxic T-lymphocyte (CTL), helper T-lymphocyte (HTL), and B-cell epitopes, by using linker and adjuvants to stimulate both cellular and humoral immune responses. These adjuvants and linkers are used to enhance the stability, immunogenicity, and efficacy of the vaccine. In addition, multi-epitope vaccines offer improved accuracy and specificity, safety, and broad-spectrum protection against multiple viral strains and its variants [23].

The aims of my study are to design epitope-based vaccine by using an artificial intelligence and immune-informatics-based against SARS-CoV-2. We target the membrane protein of SARS-CoV-2, its high conservation among SARS-CoV-2 variants, significant antigenic potential, and important role in viral assembly and host immune recognition. Various computational web server and AI/ML based tools and software were employed to predict epitopes, construct the vaccine candidate, evaluate its physicochemical and immunological properties, perform molecular docking analysis, and immune simulation [24]. All findings of my study may contribute to the development of effective and broad-spectrum vaccine candidates against SARS-CoV-2 and its variants [25,26. 27].

II. MATERIAL AND METHODS

The methodology contains in-silico genome sequence, membrane protein sequence retrieval analysis, epitope prediction, and immunological screening to evaluate antigenicity, safety, and population relevance, structural modelling and protein-protein docking and immune simulation.

We retrieved SARS-CoV-2 genome (accession no MT520466) in fasta format from the National Center for Biotechnology Information (NCBI) database in FASTA format. Subsequently, the Membrane (M) protein (Accession No. QKG90185.1) was retrieved and selected as the target antigen for vaccine design. We analysed different parameter of selected protein such as physicochemical property antigenicity, allergenicity, secondary structure target prediction, tertiary structural prediction, disulphide bond prediction by using expasy protparam, vaxijen 2.0, allertop 2.0, psipred, alpha fold dianna 1.1 web server respectively.

The physicochemical analysis included the determination of molecular weight, theoretical pI, instability index, aliphatic index, and GRAVY score, disulphide bond prediction by using DiANNA server to evaluate structural stability, secondary structure prediction was performed by using PSIPRED and tertiary structure prediction by using the trRosetta server.

Furthermore, B-cell, MHC Class I (CTL), and MHC Class II (HTL) epitopes were predicted using the Immune Epitope Database (IEDB). Then predicted epitopes were screened and filtered based on binding affinity (IC50 value), percentile rank, antigenicity, allergenicity, toxicity, epitope conservancy, and population coverage. We were selected highly immunogenic, non-allergenic, non-toxic, and conserved epitopes for the construction of the multi-epitope vaccine candidate.

After the prediction and screening of B-cell, MHC Class I (CTL), and MHC Class II (HTL) epitopes based on antigenicity, allergenicity, toxicity, epitope conservancy, and population coverage, the selected epitopes were assembled to construct a multi-epitope vaccine candidate. To enhance the immunogenicity of the vaccine, β-defensin was incorporated as an adjuvant at the N-terminal region of the construct. Different linkers were used to connect the selected epitopes and maintain their individual immunological functions. The EAAAK linker was used to connect the adjuvant with the vaccine construct, whereas AAY, GPGPG, and KK linkers were used to join CTL, HTL, and B-cell epitopes, respectively. These linkers facilitate efficient antigen processing and presentation while preventing the formation of junctional epitopes. The final vaccine construct was designed to induce both humoral and cellular immune responses.

Furthermore, the designed vaccine construct was evaluated for its physicochemical properties, antigenicity, allergenicity, toxicity, and structural characteristics. Secondary and tertiary structure analyses were performed to assess the stability and

structural integrity of the vaccine candidate. Subsequently, Toll-Like Receptor 4 (TLR4; PDB ID: 2A0Z) was selected as the immune receptor for molecular docking studies. The docked complex was further analysed through molecular dynamics simulation using the iMODS server to evaluate its stability and flexibility. Finally, immune simulation analysis was performed to predict the immune response generated by the designed vaccine construct.

III. RESULT AND DISCUSSION

The complete genome sequences of SARS-CoV-2 (MT520466) were successfully retrieved from the NCBI database. Then extraction of all viral proteins was conducted. Based on immunogenicity profiling, primary targets were identified: the Membrane (M) protein of SARS-CoV-2. These proteins were prioritized as the foundation for the vaccine constructs due to their critical roles in viral structure and pathogenesis. Also, the physicochemical analysis of the selected proteins was done which shows favourable stability, solubility, and hydrophobicity profiles, supporting their suitability as vaccine targets. Furthermore, disulfide bond prediction using the DiANNA server revealed structurally significant cysteine residues involved in disulfide bridge formation, indicating proper folding and enhanced structural stability of both the Membrane protein. Secondary structure prediction of the Membrane protein and the secondary secreted glycoprotein was performed using PSIPRED. Furthermore, tertiary structure modelling of both proteins was carried out using trRosetta.

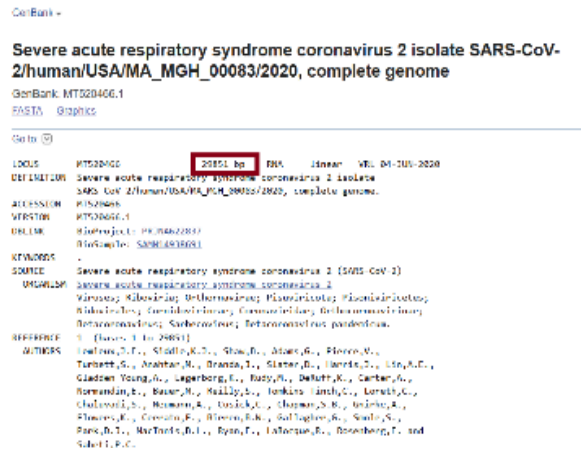


Fig no: 01 Retrieval of Viral Genome Sequence of SARS-CoV-2 (MT520466)

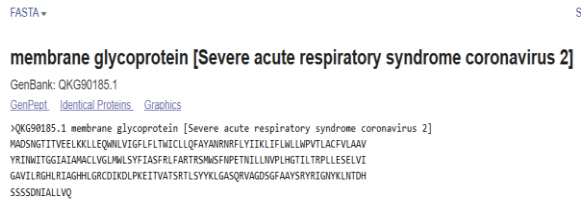


Fig no:02 Retrieval of Fasta Sequence of Membrane (M) protein (QKG90185.1)

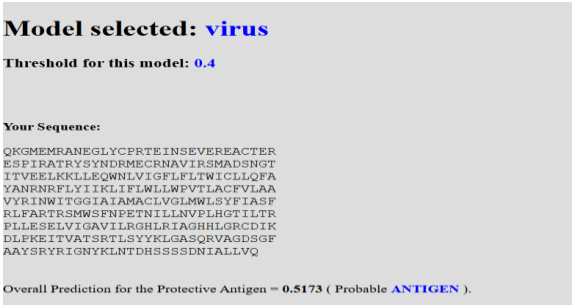


Fig no: 03 The antigenicity of all viral proteins in SARS-CoV-2 was predicted using VaxiJen v2.0

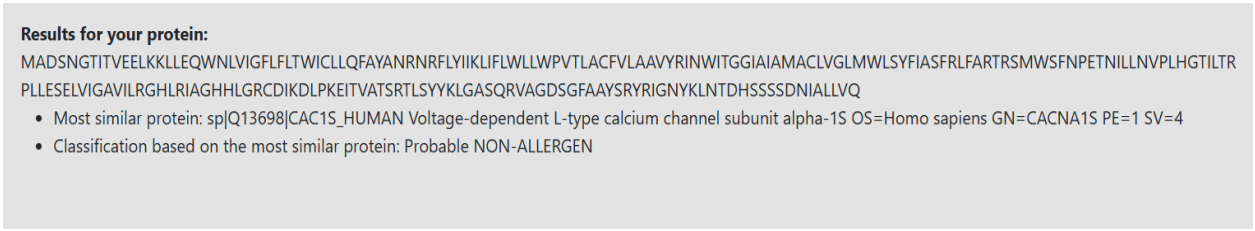


Fig no: 04 The Allergenicity of all viral proteins in SARS-CoV-2 were predicted using AllerTop v2.0

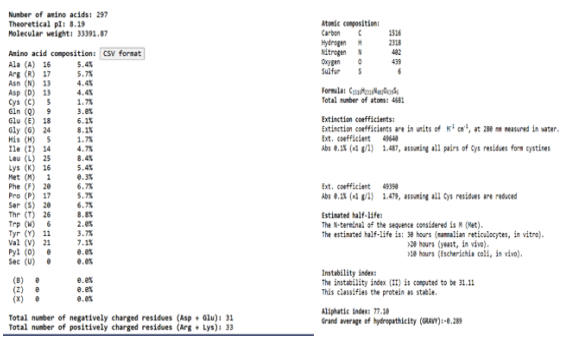
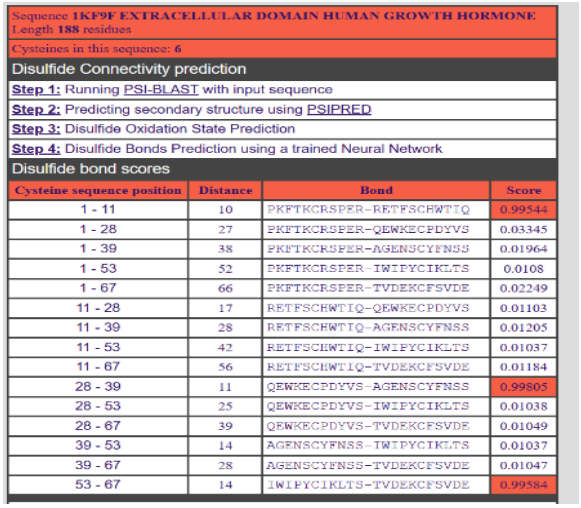


Fig no: 05 Physiochemical properties of Membrane (M) protein (QKG90185.1)



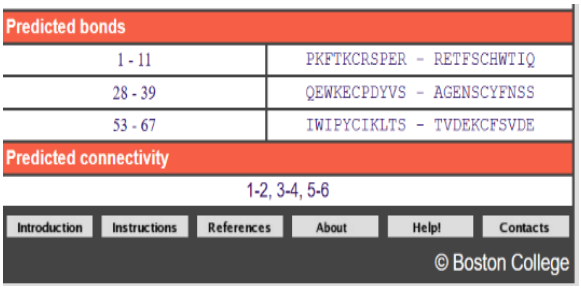


Fig no: 06 Computational prediction of disulfide bond connectivity for the SARS-CoV-2 Membrane (M) protein (QKG90185.1).

The schematic illustrates the potential linkage of six cysteine residues within the N-terminal and transmembrane regions Neural network-based confidence scores for Cys-Cy's pairings in the SARS-CoV-2 M protein. High-probability scores (P > 0.99) suggest a structured disulfide network (1-11, 28-39, and 53-67) that may stabilize the protein's dimer interface or membrane orientation.

Fig no: 07: PSIPRED-based secondary structure prediction of the SARS-CoV-2 membrane protein showing the distribution of α-helices, β-strands, and coil regions along the amino-acid sequence.

Fig no: 08 Graphical output of PSIPRED displaying confidence scores and three-state (helix, strand, coil) secondary structure assignment for the SARS-CoV-2 membrane protein.

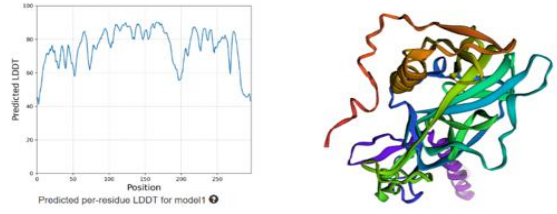
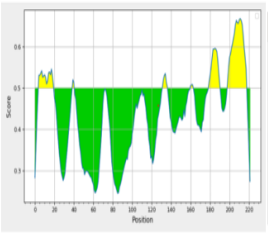


Fig no: 09 Predicted three-dimensional (tertiary) structure of the SARS-CoV-2 membrane protein generated using the trRosetta modelling approach.Per-residue pLDDT confidence plot for the trRosetta-predicted SARS-CoV-2 membrane protein model.

IV. PREDICTION OF LINEAR B-CELL EPITOPES

The primary sequences of the SARS-CoV-2 Membrane (M) protein were analysed to identify potential B-cell epitopes. The proteins were screened for surface-exposed regions which have the ability to trigger a humoral immune response. This was done using the BepiPred-2.0 algorithm.



No.	Start	End	Peptide	Length
1	5	20	NGTITVEELKKLLEQW	16
2	40	41	AN	2
3	132	137	PLLESE	6
4	161	163	IKD	3
5	180	191	KLGASQRVAGDS	12
6	199	218	YRIGNYKLNTDHSSSSDNIA	20

Fig no:10 For the SARS-CoV-2 membrane protein, epitope prediction focused on extracellular regions. High-scoring epitopes were mainly found in the N-terminal tail and extracellular loops between transmembrane helices, and these surface-exposed regions were selected to ensure effective immune recognition. B-cell epitope mapping of the SARS-CoV-2 Membrane (M) protein. Highlights represent immunodominant extracellular loops identified for their high surface accessibility.

T-Cell Epitope Prediction and Immunological Safety Screening

The T-cell epitopes of MHC-I and MHC-II molecules were predicted by the help of IEDB tools. This was done to make sure safety and efficacy of the vaccine.

Biological Safety Assessment

The predicted T-cell epitopes were screened for antigenicity, allergenicity, and toxicity to ensure safety and immunogenicity. VaxiJen, AllerTop, ToxinPred confirmed the predicted epitopes are potential antigen, non-allergenic and non-toxic.

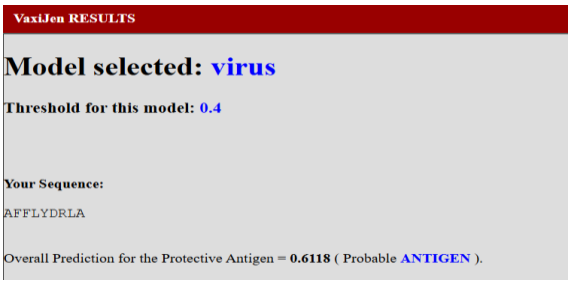


Fig no:11: Evaluation of antigenicity using VaxiJen

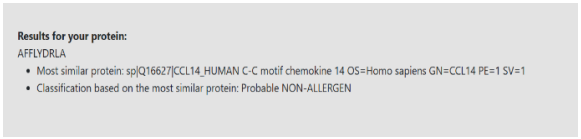
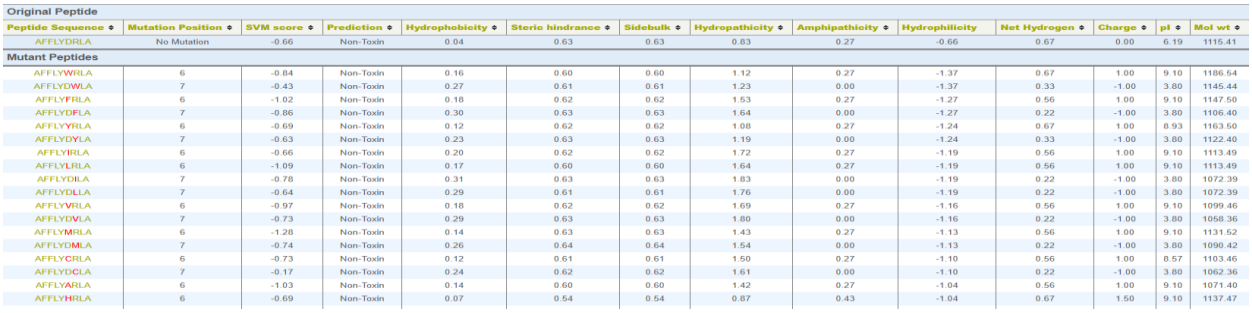


Fig no: 12 Evaluation of allergenicity using AllerTop

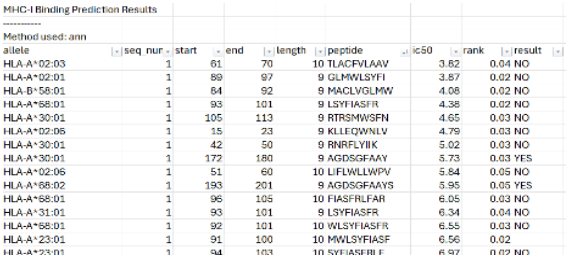


Peptide Sequence	Mutation Position	SVM score	Prediction	Hydrophobicity	Steric hindrance	Sidebulk	Hydropathicity	Amphipathicity	Hydrophilicity	Net Hydrogen	Charge	pI	Mol wt
Original Peptide													
AFFLYDRLA	No Mutation	-0.66	Non-Toxin	0.04	0.63	0.63	0.83	0.27	-0.66	0.67	0.00	6.19	1115.41
Mutant Peptides													
AFFLYWRLA	6	-0.84	Non-Toxin	0.16	0.60	0.60	1.12	0.27	-1.37	0.67	1.00	9.10	1186.54
AFFLYDWLA	7	-0.43	Non-Toxin	0.27	0.61	0.61	1.23	0.00	-1.37	0.33	-1.00	3.80	1145.44
AFFLYFRLA	6	-1.02	Non-Toxin	0.18	0.62	0.62	1.53	0.27	-1.27	0.56	1.00	9.10	1147.50
AFFLYDFLA	7	-0.86	Non-Toxin	0.30	0.63	0.63	1.64	0.00	-1.27	0.22	-1.00	3.80	1106.40
AFFLYYRLA	6	-0.69	Non-Toxin	0.12	0.62	0.62	1.08	0.27	-1.24	0.67	1.00	8.93	1163.50
AFFLYDYLA	7	-0.63	Non-Toxin	0.23	0.63	0.63	1.19	0.00	-1.24	0.33	-1.00	3.80	1122.40
AFFLYIRLA	6	-0.66	Non-Toxin	0.20	0.62	0.62	1.72	0.27	-1.19	0.56	1.00	9.10	1113.49
AFFLYLRLA	6	-1.09	Non-Toxin	0.17	0.60	0.60	1.64	0.27	-1.19	0.56	1.00	9.10	1113.49
AFFLYDILA	7	-0.78	Non-Toxin	0.31	0.63	0.63	1.83	0.00	-1.19	0.22	-1.00	3.80	1072.39
AFFLYDLLA	7	-0.64	Non-Toxin	0.29	0.61	0.61	1.76	0.00	-1.19	0.22	-1.00	3.80	1072.39
AFFLYVRLA	6	-0.97	Non-Toxin	0.18	0.62	0.62	1.69	0.27	-1.16	0.56	1.00	9.10	1099.46
AFFLYDVLA	7	-0.73	Non-Toxin	0.29	0.63	0.63	1.80	0.00	-1.16	0.22	-1.00	3.80	1058.36
AFFLYMRLA	6	-1.28	Non-Toxin	0.14	0.63	0.63	1.43	0.27	-1.13	0.56	1.00	9.10	1131.52
AFFLYDMLA	7	-0.74	Non-Toxin	0.26	0.64	0.64	1.54	0.00	-1.13	0.22	-1.00	3.80	1090.42
AFFLYCRLA	6	-0.73	Non-Toxin	0.12	0.61	0.61	1.50	0.27	-1.10	0.56	1.00	8.57	1103.46
AFFLYDCLA	7	-0.17	Non-Toxin	0.24	0.62	0.62	1.61	0.00	-1.10	0.22	-1.00	3.80	1062.36
AFFLYARLA	6	-1.03	Non-Toxin	0.14	0.60	0.60	1.42	0.27	-1.04	0.56	1.00	9.10	1071.40
AFFLYHRLA	6	-0.69	Non-Toxin	0.07	0.54	0.54	0.87	0.43	-1.04	0.67	1.50	9.10	1137.47

Fig no: 13 Evaluation of toxicity using ToxinPred

MHC Class I and Class II Epitope Evaluation

T-cell epitope mapping was carried out for both viral targets. For SARS-CoV-2, MHC Class I and Class II binding analyses identified multiple epitopes across diverse HLA alleles, ensuring broad cellular immune coverage.



MHC-I Binding Prediction Results

Method used: ann

allele	seq_num	start	end	length	peptide	ic50	rank	result
HLA-A*02:03	1	61	70	10	TLACFVLAAV	3.82	0.04	NO
HLA-A*02:01	1	89	97	9	GLMWLSYFI	3.87	0.02	NO
HLA-B*58:01	1	84	92	9	MACLVGLMW	4.08	0.02	NO
HLA-A*68:01	1	93	101	9	LSYFIASFR	4.38	0.02	NO
HLA-A*30:01	1	105	113	9	RTRSMWSFN	4.65	0.03	NO
HLA-A*02:06	1	15	23	9	KLLEQWNLV	4.79	0.03	NO
HLA-A*30:01	1	42	50	9	RNRFLYIIK	5.02	0.03	NO
HLA-A*30:01	1	172	180	9	AGDSGFAAY	5.73	0.03	YES
HLA-A*02:06	1	51	60	10	LIFLWLLWPV	5.84	0.05	NO
HLA-A*68:02	1	193	201	9	AGDSGFAAYS	5.95	0.05	YES
HLA-A*68:01	1	96	105	10	FIASFRLFAR	6.05	0.03	NO
HLA-A*31:01	1	93	101	9	LSYFIASFR	6.34	0.04	NO
HLA-A*68:01	1	92	101	10	WLSYFIASFR	6.55	0.03	NO
HLA-A*23:01	1	91	100	10	MWLSYFIASF	6.56	0.02	
HLA-A*23:01	1	94	103	10	SYFIASFRLF	6.97	0.02	NO

Fig no: 14: Evaluation of MHC class-1 Epitopes of SARS-CoV-2 Membrane protein based on percentile rank.

Final Selection of High-Affinity T-cell Epitopes

T-cell epitopes from MHC Class I and MHC Class II peptides of SARS-CoV-2 were selected because they showed strong binding ability, with IC50 values below 50 nM and percentile rank below 2.0, indicating high potential to trigger effective immune responses

MHC-I Binding Prediction Results

Method used: ann

allele	seq_nur	start	end	length	peptide	ic50	rank	result
HLA-A*30:01	1	172	180	9	TSRTLSYYK	5.73	0.03	YES
HLA-A*68:02	1	193	201	9	FAAYSRYRI	5.95	0.05	YES
HLA-A*02:06	1	65	73	9	FVLAAVYRI	7.66	0.07	YES
HLA-A*11:01	1	171	180	10	ATSRTLSYYK	7.92	0.02	YES
HLA-A*30:01	1	171	180	10	ATSRTLSYYK	8.47	0.05	YES
HLA-A*68:01	1	5	14	10	NGTITVEELK	9.26	0.06	YES
HLA-A*68:01	1	137	146	10	ELVIGAVILR	9.88	0.06	YES
HLA-A*23:01	1	44	53	10	RFLYIIKLIF	11.12	0.03	YES
HLA-A*68:01	1	138	146	9	LVIGAVILR	11.41	0.08	YES
HLA-A*23:01	1	46	55	10	LYIIKLIFLW	12.03	0.03	YES
HLA-A*33:01	1	178	186	9	YYKLGASQR	13.37	0.03	YES
HLA-A*02:03	1	71	80	10	YRINWITGGI	15.86	0.26	YES
HLA-A*02:01	1	65	73	9	FVLAAVYRI	16.52	0.16	YES
HLA-A*01:01	1	171	179	9	ATSRTLSYY	16.73	0.04	YES
HLA-A*02:01	1	88	97	10	VGLMWLSYFI	17.08	0.16	YES
HLA-B*40:01	1	136	145	10	SELVIGAVIL	17.83	0.05	YES
HLA-A*68:01	1	6	14	9	GTITVEELK	18.08	0.14	YES

Fig no: 15 T-cell epitopes of MHC-1 peptides of SARS-CoV-2 below IC50 value and below 2.0 Rank

MHC-II Binding Prediction Results

Method used: netmhciipan_ba

allele	seq_nur	start	end	length	core_peptide	peptide	ic50	rank	result
HLA-DQA1	1	136	150	15	VIGAVILRG	SELVIGAVILRGHLR	7.16	0.1	YES
HLA-DQA1	1	137	151	15	VIGAVILRG	ELVIGAVILRGHLRI	12.51	0.15	YES
HLA-DRB1	1	100	114	15	FARTRSMWS	FRLFARTRSMWSFNP	12.59	0.16	YES
HLA-DRB1	1	99	113	15	FRLFARTRS	SFRLFARTRSMWSFN	13.59	0.18	YES
HLA-DQA1	1	71	85	15	ITGGIAIAM	YRINWITGGIAIAMA	13.7	0.22	YES
HLA-DRB1	1	99	113	15	FARTRSMWS	SFRLFARTRSMWSFN	14.59	0.23	YES
HLA-DQA1	1	138	152	15	VIGAVILRG	LVIGAVILRGHLRIA	14.87	0.48	YES
HLA-DRB1	1	100	114	15	FARTRSMWS	FRLFARTRSMWSFNP	15.49	0.53	YES
HLA-DQA1	1	70	84	15	WITGGIAIA	VYRINWITGGIAIAM	16.34	0.54	YES
HLA-DRB1	1	99	113	15	FARTRSMWS	SFRLFARTRSMWSFN	16.88	0.6	YES
HLA-DRB1	1	99	113	15	FARTRSMWS	SFRLFARTRSMWSFN	17.44	0.65	YES
HLA-DRB1	1	100	114	15	FARTRSMWS	FRLFARTRSMWSFNP	19.76	0.69	YES
HLA-DRB1	1	99	113	15	FARTRSMWS	SFRLFARTRSMWSFN	19.92	0.69	YES
HLA-DRB3	1	68	82	15	YRINWITGG	AAVYRINWITGGIAI	20.82	0.74	YES
HLA-DRB5	1	138	152	15	AVILRGHLR	LVIGAVILRGHLRIA	20.86	0.8	YES
HLA-DQA1	1	133	147	15	VIGAVILRG	LLESELVIGAVILRG	20.91	0.87	YES
HLA-DRB1	1	100	114	15	FARTRSMWS	FRLFARTRSMWSFNP	21.24	0.94	YES
HLA-DRB5	1	139	153	15	AVILRGHLR	VIGAVILRGHLRIAG	21.84	0.96	YES
HLA-DRB3	1	67	81	15	YRINWITGG	LAAVYRINWITGGIA	22.53	0.99	YES
HLA-DQA1	1	71	85	15	ITGGIAIAM	YRINWITGGIAIAMA	23.54	1.2	YES

Fig no: 16 T-cell epitopes of MHC-2 peptides of SARS-CoV-2 below IC50 value and below 2.0 Rank

Epitope Conservancy

To make sure broad protection against viral variants, selected T-cell epitopes were analysed for sequence conservancy. Only those epitopes with ≥90% conservation across viral strains were chosen for the final construct.

Epitope #	Epitope name	Epitope sequence	Epitope le	Percent of protein	Minimum i	Maximum i
1	ws-separated-0	AEGVVAFLILPQAKK	15	100.00% (1/1)	100.00%	100.00%
2	ws-separated-1	AFHKEGAFFLYDRLA	15	100.00% (1/1)	100.00%	100.00%
3	ws-separated-2	AFLILPQAKKDFFSS	15	100.00% (1/1)	100.00%	100.00%
4	ws-separated-3	ATGFGTNETEYLFEV	15	100.00% (1/1)	100.00%	100.00%
5	ws-separated-4	DFAFHKEGAFFLYDR	15	100.00% (1/1)	100.00%	100.00%
6	ws-separated-5	DRLASTVIYRGTTFA	15	100.00% (1/1)	100.00%	100.00%
7	ws-separated-6	EGVVAFLILPQAKKD	15	100.00% (1/1)	100.00%	100.00%
8	ws-separated-7	ETEYLFEVDNLTYVQ	15	100.00% (1/1)	100.00%	100.00%
9	ws-separated-8	EYLFEVDNLTYVQLE	15	100.00% (1/1)	100.00%	100.00%
10	ws-separated-9	FAFHKEGAFFLYDRL	15	100.00% (1/1)	100.00%	100.00%
11	ws-separated-10	FFLYDRLASTVIYRG	15	100.00% (1/1)	100.00%	100.00%
12	ws-separated-11	FLWVIILFQRTFSIP	15	100.00% (1/1)	100.00%	100.00%
13	ws-separated-12	FQRTFSIPLGVIHNS	15	100.00% (1/1)	100.00%	100.00%
14	ws-separated-13	FSIPLGVIHNSTLQV	15	100.00% (1/1)	100.00%	100.00%
15	ws-separated-14	GVVAFLILPQAKKDF	15	100.00% (1/1)	100.00%	100.00%
16	ws-separated-15	ILFQRTFSIPLGVIH	15	100.00% (1/1)	100.00%	100.00%
17	ws-separated-16	IPLGVIHNSTLQVSD	15	100.00% (1/1)	100.00%	100.00%
18	ws-separated-17	LESRFTPQFLLQLNE	15	100.00% (1/1)	100.00%	100.00%
19	ws-separated-18	LFEVDNLTYVQLESR	15	100.00% (1/1)	100.00%	100.00%
20	ws-separated-19	LTYVQLESRFTPQFL	15	100.00% (1/1)	100.00%	100.00%
21	ws-separated-20	LWVIILFQRTFSIPL	15	100.00% (1/1)	100.00%	100.00%

Fig no:17 MHC Class II (HTL): The 15-mer Class II epitopes were 100% conserved, indicating these regions are highly stable and unlikely to mutate. Their ability to bind multiple HLA-DR, DP, and DQ alleles supports broad helper T-cell activation and global immune coverage.

SARS-CoV-2 Membrane (M) Protein Conservancy Results

#	Epitope na	Epitope sequenc	Epitope ler	Percent of prote	Minimum i	Maximum i
1	ws-separa	AGDSGFAAYS	10	100.00% (1/1)	100.00%	100.00%
2	ws-separa	ASFRLFARTR	10	100.00% (1/1)	100.00%	100.00%
3	ws-separa	ATSRTLSYYK	10	100.00% (1/1)	100.00%	100.00%
4	ws-separa	ELVIGAVILR	10	100.00% (1/1)	100.00%	100.00%
5	ws-separa	FAAYSRYRI	9	100.00% (1/1)	100.00%	100.00%
6	ws-separa	FVLAAVYRI	9	100.00% (1/1)	100.00%	100.00%
7	ws-separa	GTITVEELKK	10	100.00% (1/1)	100.00%	100.00%
8	ws-separa	LVIGAVILR	9	100.00% (1/1)	100.00%	100.00%
9	ws-separa	LVIGFLFLTW	10	100.00% (1/1)	100.00%	100.00%
10	ws-separa	LYIIKLIFLW	10	100.00% (1/1)	100.00%	100.00%
11	ws-separa	NGTITVEELK	10	100.00% (1/1)	100.00%	100.00%
12	ws-separa	RFLYIIKLIF	10	100.00% (1/1)	100.00%	100.00%
13	ws-separa	SELVIGAVI	9	100.00% (1/1)	100.00%	100.00%
14	ws-separa	SFRLFARTR	9	100.00% (1/1)	100.00%	100.00%
15	ws-separa	SGFAAYSRYR	10	100.00% (1/1)	100.00%	100.00%
16	ws-separa	TSRTLSYYK	9	100.00% (1/1)	100.00%	100.00%
17	ws-separa	YRINWITGGI	10	100.00% (1/1)	100.00%	100.00%
18	ws-separa	YYKLGASQR	9	100.00% (1/1)	100.00%	100.00%

Fig no:18 MHC Class I (CTL): Conserved, high-affinity peptides from the SARS-CoV-2 membrane protein were selected to maintain effectiveness against emerging variants.

#	Epitope na	Epitope sequence	Epitope le	Percent of prote	Minimum i	Maximum i	V
1	ws-separa	AAVYRINWITGGIAI	15	100.00% (1/1)	100.00%	100.00%	
2	ws-separa	DLPKEITVATSRTLS	15	100.00% (1/1)	100.00%	100.00%	
3	ws-separa	ELVIGAVILRGHLRI	15	100.00% (1/1)	100.00%	100.00%	
4	ws-separa	FRLFARTRSMWSFNP	15	100.00% (1/1)	100.00%	100.00%	
5	ws-separa	LAAVYRINWITGGIA	15	100.00% (1/1)	100.00%	100.00%	
6	ws-separa	LVIGAVILRGHLRIA	15	100.00% (1/1)	100.00%	100.00%	
7	ws-separa	SELVIGAVILRGHLR	15	100.00% (1/1)	100.00%	100.00%	
8	ws-separa	SFRLFARTRSMWSFN	15	100.00% (1/1)	100.00%	100.00%	
9	ws-separa	VIGAVILRGHLRIAG	15	100.00% (1/1)	100.00%	100.00%	
10	ws-separa	VYRINWITGGIAIAM	15	100.00% (1/1)	100.00%	100.00%	
11	ws-separa	YRINWITGGIAIAMA	15	100.00% (1/1)	100.00%	100.00%	

Fig no: 19 MHC Class II (HTL): High-ranking 15-mer epitopes located in conserved regions of the M-protein were chosen to support robust helper T-cell activation and long-term immune memory.

V. POPULATION COVERAGE ANALYSIS

A population coverage analysis was done to check the selected epitopes can provide broad immune protection in different human populations. The potential usefulness of these epitopes was assessed by estimating how well they cover diverse global populations based on HLA allele frequencies, helping predict vaccine effectiveness across different ethnic and geographical groups.

Population Coverage for SARS-CoV-2 (M-Protein)

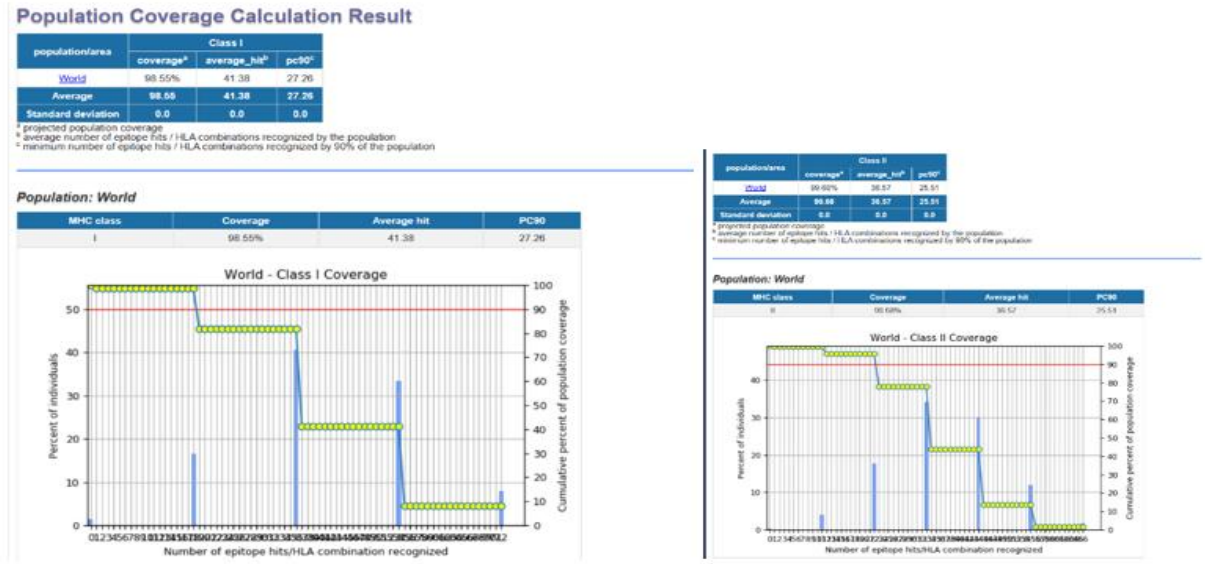


Fig no: 20 MHC Class I: CTL epitopes also demonstrated high worldwide coverage, enabling strong $CD8^+$ T-cell responses and MHC Class II: SARS-CoV-2 HTL epitopes showed broad global coverage, supporting effective $CD4^+$ T-cell activation across diverse populations.

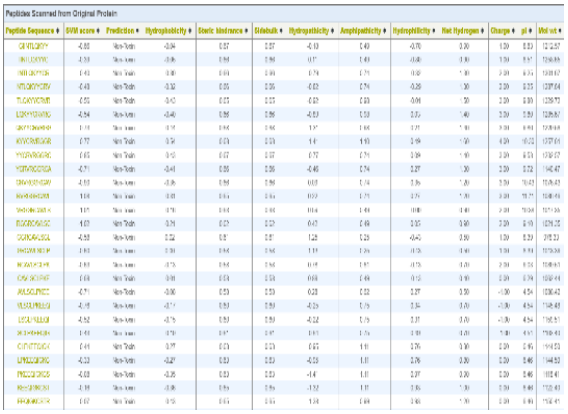
Construction and Physicochemical Evaluation of the Final Vaccine

The final vaccine construct was created by combining all selected epitopes into a single protein sequence. At the beginning to enhance the immune response a β-defensin adjuvant was added. Different linkers were inserted between epitopes to keep them properly separated and functioning as intended. Specifically, EAAAK was used to connect the adjuvant to the vaccine sequence, while AAY, GPGPG, and KK linkers were placed between CTL, HTL, and B-cell epitopes, respectively, to prevent unwanted interactions between neighboring sequences.

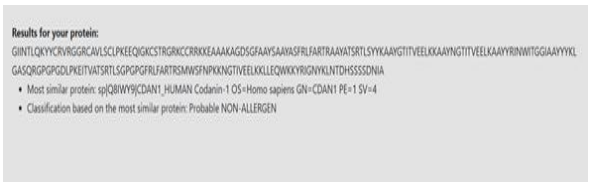
SARS-CoV-2 epitope vaccine

GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKC
STRGRKCCRRKKEAAAKAGDSGFAAYS
ASFRLFARTR ATSRTLSYYK
GTITVEELKKAAYNGTITVEELKAAY
YRINWITGGI
YYKLGASQRGPGPGDLPKEITVATSRTLSFRLF
ARTRSMWSFNPKKNGTIVEELKKLLEQWYRIG
NYKLNTDHSSSSDNIA

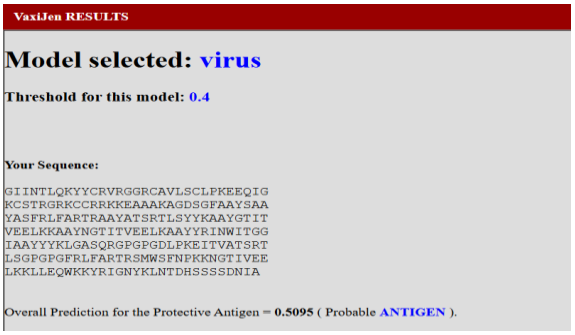
Immunological Re-validation After assembling the full fusion constructs SARS-CoV-2 vaccines were re-evaluated to confirm safety and immunogenicity. The final constructs remained antigenic, non-toxic, and non-allergenic, indicating their suitability for vaccine development.



(a)



(b)



(c)

Fig no: 21 (A) Toxicity evaluation for SARS-CoV-2 multi-epitope vaccine by Toxinpred (b)Evaluation of allergenicity of SARS-CoV-2 multi-epitope vaccine using AllerTOP v2.0(c)Evaluation of antigenicity of SARS-CoV-2 multi-epitope vaccine using VaxiJen

Physicochemical Profiling ProtParam analysis showed that both constructs possess favourable stability and formulation properties. High aliphatic index values indicated good thermostability, while negative GRAVY scores confirmed hydrophilicity and good solubility. The predicted molecular weight and isoelectric point further support feasibility for purification and large-scale production.

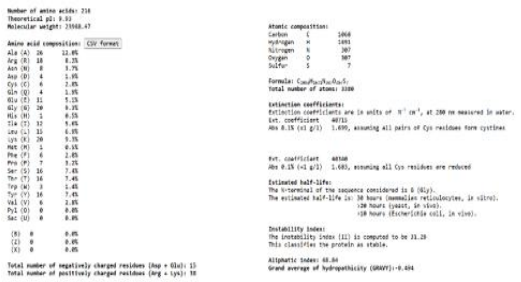


Fig no: 22 Physicochemical analyses of SARS-CoV-2 multi-epitope

5.7 Structural Modelling, Refinement, and Validation
The structural integrity of the designed multi-epitope vaccines was evaluated to ensure that the integrated epitopes are correctly folded and accessible for immune recognition.

Secondary Structure Analysis
The secondary structures of the SARS-CoV-2 M-protein constructs were predicted using the PSIPRED server. The results showed a balanced mix of structural elements.

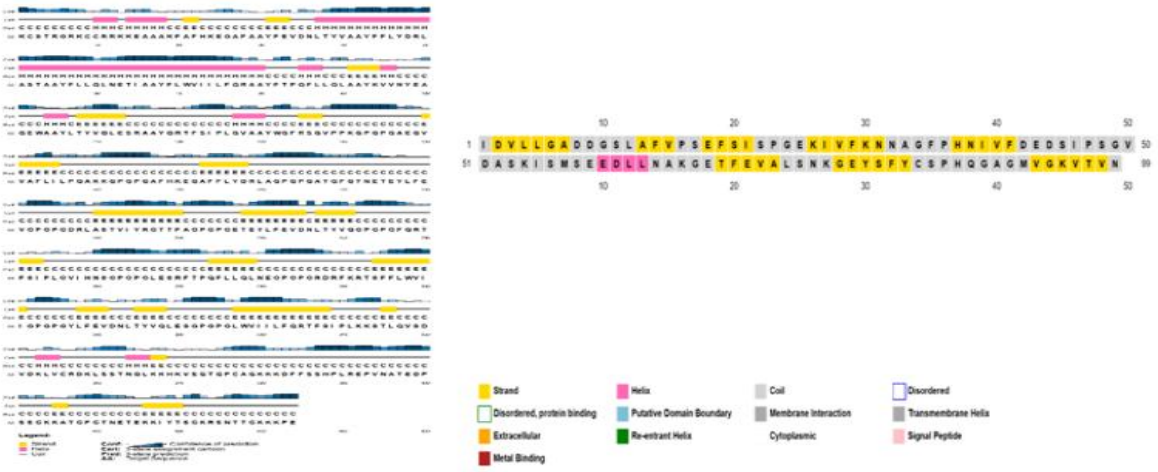


Fig no: 23 Secondary Structure prediction of SARS-CoV-2 Vaccine Designing

Tertiary Structure Prediction
Structural modelling and refinement of the vaccine constructs were conducted to evaluate their three-dimensional conformation and stability. The 3D structures were initially predicted For SARS-CoV-2: The predicted structure displays a combination of compact helical regions and extended loops. The lower confidence in certain loop regions is expected for multi-epitope vaccines, as these areas are designed to be flexible and "disordered" to allow better access for immune cell receptors.

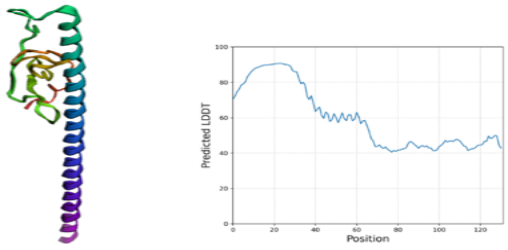


Fig no: 24 3D vaccine of SARS-CoV-2 and predicted per-residue pLDDT profile for Membrane (M) protein, residues 1-120. The plot highlights a confident N-terminal domain (pLDDT >80) transitioning to disordered C-terminal regions (pLDDT <50), derived from trRosetta prediction.

Structural Refinement

The initial confidence levels in the structure were low. By the help of Galaxy Refinement (Heo et al., 2013), the quality of structures was improved stereo-chemically by relaxation of structure and rebuilding of repeated sidechain. The refined models were then checked using Ramachandran plots to examine backbone angles and confirm that the structures were energetically reasonable.

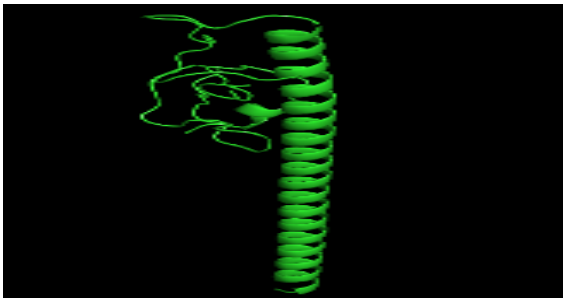


Fig no: 25 Galaxy Refinement-optimized three-dimensional structure of the vaccine construct.

Receptor-ligand docking

Afterall, protein-protein docking was performed using the ClusPro server (Ashizawa et al., 2024) to Analyse the interaction between the vaccine constructs and human immune receptorsand the SARS-CoV-2 vaccine was docked against the 2A0Z receptor. The complexes with the lowest energy scores and largest cluster sizes were selected to determine the binding affinity and stability of the vaccine-receptor interface. Visualization in PyMOL showed that the SARS-CoV-2 epitope–receptor complexes form clear and well-organized interaction interfaces. Several salt bridges and hydrophobic contacts were observed, which help stabilize these protein–protein interactions.

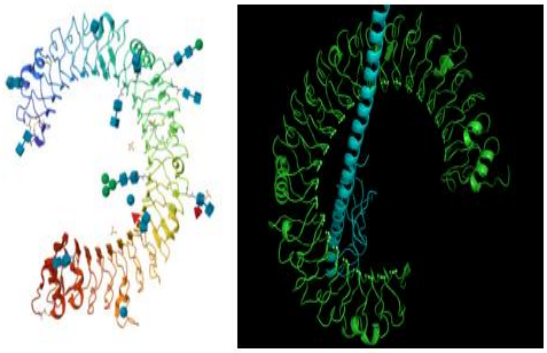


Fig no: 26 Structure of 2A0Z receptor to dock against vaccine construct of SARS-CoV-2

VI. MOLECULAR SIMULATION RESULT

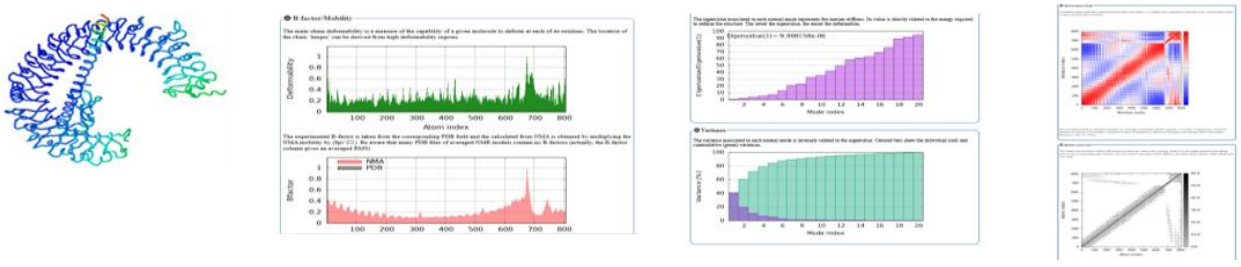


Fig no 27: Result of Immune Simulation by using NMA method

Normal Mode Analysis by using iMODS revealed a low eigenvalue (9.998×10^{-6}), it indicating favorable deformability and structural flexibility of the vaccine construct. The covariance matrix demonstrated coordinated residue motions, while the elastic network confirmed strong structural integrity, suggesting that the designed vaccine possesses adequate stability and dynamic behavior.c-imm-simulation Result:

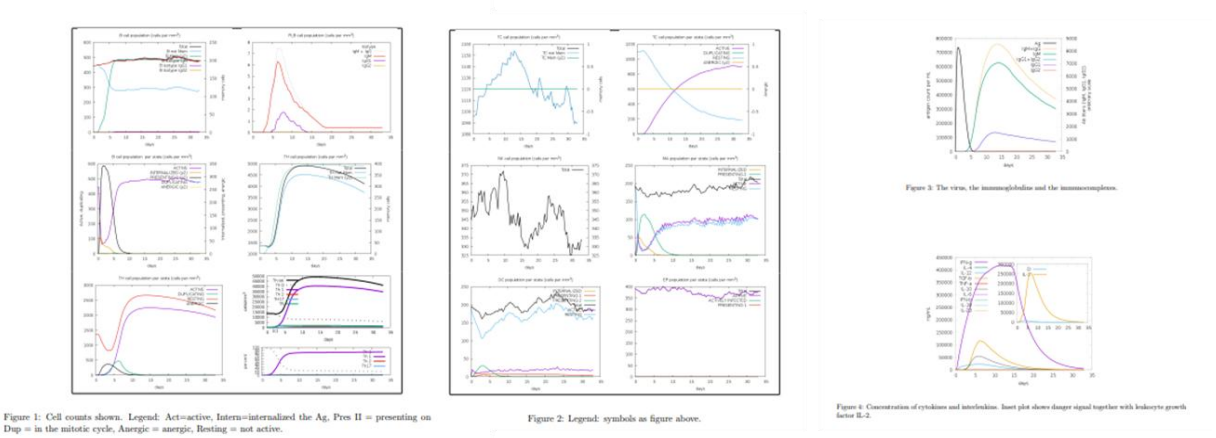


Fig no 28: Immune of Simulation by c-immsim for B epitopes, MHC-I peptides, MHC-II peptides

The C-ImmSim simulation demonstrated strong activation of B cells, helper T cells, cytotoxic T cells, macrophages, and dendritic cells, accompanied by substantial production of IgM/IgG antibodies and elevated IFN-γ levels. The generation of immune memory cells and efficient antigen clearance suggest that the vaccine construct is capable of eliciting a robust and long-lasting immune response.

VII. CONCLUSION

In this study, a computational multi-epitope vaccine design approach was employed to develop a vaccine candidate targeting the Membrane (M) protein of SARS-CoV-2. The membrane protein was selected due to its high conservation among viral variants, strong antigenic potential, important immunological roles, and non-allergenic nature. B-cell, MHC Class I (CTL), and MHC Class II (HTL) epitopes were identified to stimulate both humoral and cellular immune responses. The predicted epitopes were screened based on antigenicity, allergenicity, toxicity, population coverage, and sequence conservancy to ensure safety, immunogenicity, and broad-spectrum protection against viral mutations.

The selected epitopes were assembled into a multi-epitope vaccine construct using suitable adjuvants and linkers to enhance immune stimulation, facilitate efficient antigen processing and presentation, and maintain structural stability. The designed vaccine candidate was further evaluated through physicochemical characterization, structural modelling, molecular docking, molecular dynamics simulation, and immune simulation analyses.

The results demonstrated that the designed vaccine construct possesses favorable immunological, structural, and physicochemical properties, supporting its potential as a promising vaccine candidate against SARS-CoV-2. Although the present findings are based on computational analyses, they provide a strong foundation for future in vitro and in vivo experimental validation and may contribute to the development of effective and broad-spectrum vaccines against SARS-CoV-2 and its emerging variants.

REFERENCE

[1] N. N. Agbiji and M. O. Ekong, “An overview of antigen properties, classifications and mechanisms of action,” *Asian Journal of Immunology*, vol. 7, no. 1, pp. 201–208, Oct. 2024, doi: 10.9734/AJI/2024/V7I1144.

[2] B. Ahmad, U. A. Ashfaq, M. Rahman, M. S. Masoud, and M. Z. Yousaf, “Conserved B and T cell epitopes prediction of Ebola virus glycoprotein for vaccine development: An immuno-informatics approach,” *Microbial*

Pathogenesis, vol. 132, pp. 243–253, 2019, doi: 10.1016/j.micpath.2019.05.010.

[3] H. R. Ansari and G. P. S. Raghava, "Identification of conformational B-cell epitopes in an antigen from its primary sequence," *Immunome Research*, vol. 6, no. 6, 2010, doi: 10.1186/1745-7580-6-6.

[4] M. A. de La Vega, G. Wong, G. P. Kobinger, and X. Qiu, "The multiple roles of sGP in Ebola pathogenesis," *Viral Immunology*, vol. 28, no. 1, pp. 3–9, Feb. 2015, doi: 10.1089/vim.2014.0068.

[5] S. O. Elshafei, N. A. Mahmoud, and Y. A. Almofti, "Immunoinformatics, molecular docking and dynamics simulation approaches unveil a multi epitope-based potent peptide vaccine candidate against avian leukosis virus," *Scientific Reports*, vol. 14, Art. no. 2870, 2024, doi: 10.1038/s41598-024-53048-6.

[6] L. Heo, H. Park, and C. Seok, "GalaxyRefine: Protein structure refinement driven by side-chain repacking," *Nucleic Acids Research*, vol. 41, no. Web Server issue, pp. W384–W388, Jul. 2013, doi: 10.1093/nar/gkt458.

[7] S. Jain, S. Khaiboullina, E. Martynova, S. Morzunov, and M. Baranwal, "Epidemiology of Ebolaviruses from an etiological perspective," *Pathogens*, vol. 12, no. 2, Art. no. 248, 2023, doi: 10.3390/pathogens12020248.

[8] K. Koelle, M. A. Martin, R. Antia, B. Lopman, and N. E. Dean, "The changing epidemiology of SARS-CoV-2," *Science*, vol. 375, no. 6585, pp. 1116–1121, Mar. 2022.

[9] D. Lara-Ramírez, C. E. Santacruz-Tinoco, E. Ramón-Gallegos, and J. E. Muñoz-Medina, "In silico design of Ebola virus glycoprotein antigenic peptides as vaccine candidates," *PLoS ONE*, vol. 20, no. 3, Art. no. e0319496, Mar. 2025.

[10] L. J. McGuffin, K. Bryson, and D. T. Jones, "The PSIPRED protein structure prediction server," *Bioinformatics*, vol. 16, no. 4, pp. 404–405, Apr. 2000, doi: 10.1093/bioinformatics/16.4.404.

[11] L. Dai, Z. Lu, R. Zhang, Q. Zhang, Y. Ming, Y. Jia, H. Qi, J. He, X. Qin, W. Ke, F. Zhang, F. Liu, and L. Lei, "Vaccine," 2025, doi: 10.1016/j.vaccine.2025.128073.

[12] Y. A. Malik, "Properties of coronavirus and SARS-CoV-2," *Malaysian Journal of Pathology*, vol. 42, no. 1, pp. 3–11, 2020.

[13] D. Malvy, A. K. McElroy, H. de Clerck, S. Günther, and J. van Griensven, "Ebola virus disease," *The Lancet*, vol. 393, no. 10174, pp. 936–948, 2019.

[14] X. Maoe, "Antigens and their impact on immune responses and transform modern healthcare strategies," *Immunome Research*, vol. 21, Art. no. 296, 2025.

[15] M. C. Jespersen, B. Peters, M. Nielsen, and P. Marcatili, "BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes," *Nucleic Acids Research*, vol. 45, no. W1, pp. W24–W29, Jul. 2017, doi: 10.1093/nar/gkx346.

[16] M. C. Jespersen, B. Peters, M. Nielsen, and P. Marcatili, "BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes," *Nucleic Acids Research*, vol. 45, no. W1, pp. W24–W29, Jul. 2017, doi: 10.1093/nar/gkx346.

[17] Normalina Irine *et al*., "Characterization of the spike glycoprotein and construction of an epitope-based vaccine candidate against Indonesian SARS-CoV-2: In silico study," *Systematic Reviews in Pharmacy*, vol. 11, no. 7, pp. 404–413, 2020.

[18] B. Peters, J. Sidney, P. Bourne, H. H. Bui, S. Buus, G. Doh, W. Fleri, M. Kronenberg, R. Kubo, O. Lund, D. Nemazee, J. V. Ponomarenko, M. Sathiamurthy, S. P. Schoenberger, S. Stewart, P. Surko, S. Way, S. Wilson, and A. Sette, "The design and implementation of the immune epitope database and analysis resource," *Immunogenetics*, vol. 57, no. 5, pp. 326–336, Jun. 2005, doi: 10.1007/s00251-005-0803-5.

[19] E. Raoufi, M. Hemmati, S. Eftekhari, K. Khaksaran, Z. Mahmodi, M. M. Farajollahi, and M. Mohsenzadegan, "Epitope prediction by novel immunoinformatics approach: A state-of-the-art review," *International Journal of Peptide Research and Therapeutics*, vol. 26, no. 2, pp. 1155–1163, 2020, doi: 10.1007/s10989-019-09918-z.

[20] R. Ashizawa, S. Kotelnikov, O. Khan, S. X. Li, E. Glukhov, X. Cao, M. Lazou, A. Bekar-

Cesaretli, D. Hailegeorgis, V. Averkava, Y. Zhu, G. Jones, H. Yu, D. Kalitin, D. Stepanenko, K. Koirala, T. Patsahan, D. Beglov, M. Lukin, D. Joseph-McCarthy, C. Simmerling, A. Tropsha, E. Coutsias, K. A. Dill, D. Padhorny, S. Vajda, and D. Kozakov, "Reference details incomplete."

[21] M. Waqas, A. Haider, M. Sufyan, S. Siraj, and S. A. Sehgal, "Determine the potential epitope-based peptide vaccine against novel SARS-CoV-2 targeting structural proteins using immunoinformatics approaches," *Frontiers in Molecular Biosciences*, vol. 7, Art. no. 227, 2020, doi: 10.3389/fmolb.2020.00227.

[22] S. F. Ahmed, A. A. Quadeer, and M. R. McKay, "Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies," *Viruses*, vol. 12, no. 3, Art. no. 254, 2020.

[23] F. Amanat and F. Krammer, "SARS-CoV-2 vaccines: Status report," *Immunity*, vol. 52, no. 4, pp. 583–589, 2020.

[24] F. Anwar *et al*., "A multi-epitope vaccine against SARS-CoV-2: An immunoinformatic approach," *Vaccines*, vol. 8, no. 3, Art. no. 444, 2020.

[25] L. R. Baden *et al*., "Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine," *New England Journal of Medicine*, vol. 384, no. 5, pp. 403–416, 2021.

[26] M. Bhattacharya *et al*., "Development of epitope-based peptide vaccine against novel coronavirus," *Scientific Reports*, vol. 10, Art. no. 14068, 2020.

[27] E. Callaway, "The race for coronavirus vaccines," *Nature*, vol. 580, no. 7805, pp. 576–577, 2020.

[28] C. Chakraborty, A. R. Sharma, and S. S. Lee, "SARS-CoV-2 vaccines in advanced clinical trials," *Human Vaccines & Immunotherapeutics*, vol. 17, no. 10, pp. 3165–3176, 2021.