Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

International Journal of Biological Macromolecules 242 (2023) 124893

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac

A novel mutation-proof, next-generation vaccine to fight against upcoming SARS-CoV-2 variants and subvariants, designed through AI enabled approaches and tools, along with the machine learning based immune simulation: A vaccine breakthrough

Manojit Bhattacharya [a,1], Abdulrahman Alshammari [b], Metab Alharbi [b], Kuldeep Dhama [c], Sang-Soo Lee [d], Chiranjib Chakraborty [e,*,1]

[a] Department of Zoology, Fakir Mohan University, Vyasa Vihar, Balasore 756020, Odisha, India
[b] Department of Pharmacology and Toxicology, College of Pharmacy, King Saud University, Post Box 2455, Riyadh 11451, Saudi Arabia
[c] Division of Pathology, ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly 243122, Uttar Pradesh, India
[d] Institute for Skeletal Aging & Orthopaedic Surgery, Hallym University-Chuncheon Sacred Heart Hospital, Chuncheon-si 24252, Gangwon-do, Republic of Korea
[e] Department of Biotechnology, School of Life Science and Biotechnology, Adamas University, Kolkata, West Bengal 700126, India

ARTICLE INFO

Keywords:
Multi-epitopic peptide vaccine
Mutation-proof
SARS-CoV-2 variants and subvariants
Immune simulation

ABSTRACT

Emerging SARS-CoV-2 variants and subvariants are great concerns for their significant mutations, which are also responsible for vaccine escape. Therefore, the study was undertaken to develop a mutation-proof, next-generation vaccine to protect against all upcoming SARS-CoV-2 variants. We used advanced computational and bioinformatics approaches to develop a multi-epitopic vaccine, especially the AI model for mutation selection and machine learning (ML) strategies for immune simulation. AI enabled and the top-ranked antigenic selection approaches were used to select nine mutations from 835 RBD mutations. We selected twelve common antigenic B cell and T cell epitopes (CTL and HTL) containing the nine RBD mutations and joined them with the adjuvants, PADRE sequence, and suitable linkers. The constructs' binding affinity was confirmed through docking with TLR4/MD2 complex and showed significant binding free energy (−96.67 kcal mol^{-1}) with positive binding affinity. Similarly, the calculated eigenvalue (2.428517e-05) from the NMA of the complex reveals proper molecular motion and superior residues' flexibility. Immune simulation shows that the candidate can induce a robust immune response. The designed mutation-proof, multi-epitopic vaccine could be a remarkable candidate for upcoming SARS-CoV-2 variants and subvariants. The study method might guide researchers in developing AI-ML and immunoinformatics-based vaccines for infectious disease.

1. Introduction

The outbreak of SARS-CoV-2 caused the pandemic and is responsible for the death of millions of people worldwide [1–3]. This virus belongs to the β coronaviruses family, and is closely related with SARS-CoV and MERS-CoV viruses [4–7]. During the last two years, after the gradual waning of the significant infectious waves, the situation was under control in some countries, like China, India, Australia, etc. Conversely, the condition of a few countries, the USA, Brazil, and Mexico, is prolonged suffering in the period mentioned [8]. COVID-19 flows worldwide with various effects regarding infections and death [9]. However, these extended infectious conditions arose in different countries owing to the emergence of new mutational variants of SARS-CoV-2.

For protection against SARS-CoV-2 and its variants, the researcher and pharmaceutical companies developed a successful vaccine candidate. Numerous research and development were performed in this direction. Previously, we designed a multi-epitopic peptide-based vaccine construct against the wild strain of SARS-CoV–2 using an immunoinformatics approach [10]. Subsequently, we also developed a next-generation vaccine construct using the alternative epitopes of significant emerging variants of SARS-CoV-2 [11]. Simultaneously, Dong et al. (2020) developed a vaccine using 44 epitopes from the SARS-CoV-2 S-

* Corresponding author.
E-mail address: drchiranjib@yahoo.com (C. Chakraborty).
[1] Authors contributed equally.

https://doi.org/10.1016/j.ijbiomac.2023.124893
Received 29 March 2023; Received in revised form 27 April 2023; Accepted 12 May 2023
Available online 17 May 2023

Table 1
All significant mutations in the SARS-CoV-2 genome (structural and non-structural protein).

Sl. no.	Part of the SARS-CoV-2 genome	Mutations	Reference
1.	nsp1	R24C, E37D, E37K, H45Y, D48G, V56I, E57G, V60I, D75E, D75G, M85V, E87D, L92F, H110Y, K120N, R124C, N126S, D144A, S166G	[68]
2.	nsp2	G220D, R218C, H237R/Y, L180S, N133D, A302V, A306V, G313D, K384N, Q427H, T434I, G438C, P624L, G584C, V485I, T44I, T85I, G339S, G285V, Q496P, K500N, G313D, D144G, G165S, V381A, I513V, T265I	[69,70]
3.	nsp3	F106F, T1198K, S1197R, P1213L, K38R, V1069I, Δ1265, L1266I, A1892T	[70–72]
4.	nsp4	F308Y, M2796I, T492I	[71,72]
5.	nsp5	K236R, N142L, K90R, A7V, L75F, C22N, H246Y, I43V, P132H	[73]
6.	nsp6	L37F, L3606F, Δ105-107, A189V	[71,74]
7.	nsp7	L71F	[71]
8.	nsp8	T4159K	[72]
9.	nsp9	–	
10.	nsp10	A20V	[75]
11.	nsp11	–	
12.	nsp12	P323L, A97V, D135Y, G671S	[71,76]
13.	nsp13	P504L, Y541C	[70,77]
14.	nsp14	I42V	
15.	nsp15	–	
16.	nsp16	A116S, P7S, P236L, L111F, N285D	[75]
17.	Spike (S)	A67V, Δ69-70, T95I, G142D, Δ143-145, Δ211, L212I, ins214EPE,T547K, D614G, H655Y, N679K, P681H, N764K, R346S, D796Y, N856K, Q954H, N969K, L981F, G339D, S371L, S373P, S375F, K417N, D427N, N440K, Y449S, G446S, S477N, T478K, F490S, E484A, Q493K, G496S, Q498R, N501Y, S494P, Y505H, A522S, K417T, L452Q, N439K, E484K, N501Y, L18F, T20N, P26S, D138Y, R190S, H655Y, R357K, L452R, T1027I, T19R, R158G, P384L, E484Q, N501T, A520S, P681R, D950N, E156del, F157del, L425R, R346K, L452Q, F490S, HV69-G196V, Q57H, G251V, S253P, L53F, S26L, Q38P, L95F, R195I	[78–80]
18.	ORF 3A		[71,81,82]
19.	E	T9I	[83]
20.	M	V10A, A194V, D3G, Q19E, A63T, I82T	[72]
21.	ORF 6	I33T	[71]
22.	ORF 7A	T28I, V82A, T120I, Δ115	[82]
23.	ORF 7B	F19L	[84,85]
24.	ORF 8	L84S, E92K, ins28269-28273	[71]
25.	N	S197L, R203K, G204R, P13L, S194L, I292T, R202K, G204R, R40C, Δ31-33, P80R, D63G, R203M, D377Y	[70,72]
26.	ORF 10	F11S, F11L, L16P, S23P, S24F, R24C, I4L	[85]

protein, membrane, and envelope protein [12]. Kalita et al. (2021) also designed a multi-peptide subunit-based epitope vaccine against SARS-CoV-2 which was developed through 33 highly antigenic epitopes using three diverse proteins [13]. Aasim et al. (2022) also developed a vaccine construct against the Omicron variants using immunoinformatics approaches applying several bioinformatics servers and tools [14]. Several researchers have given immense efforts to develop a successful vaccine construct to combat the pandemic virus. A total of >33 peptide-based vaccine constructs have been reported by researchers against the SARS-CoV-2 virus using computational tools and techniques [15]. At the same time, 12 COVID-19 vaccines have been approved worldwide against the SARS-CoV-2 infection, such as BBIBPCorV, Covaxin, RBD-Dimer, CoronaVac, CoviVac, Sputnik V, Oxford–AstraZeneca vaccine (ChAdOx1 nCoV-19), Convidicea, Johnson & Johnson vaccine, and EpiVacCorona [16]. However, it has been noted that no such vaccine candidate delivers broad protection against the

Table 2
Significant mutations within the S-protein of four SARS-CoV-2 variants; Alpha (B.1.1.7), Beta (B.1.351), Delta (B.1.617.2), Gamma (P.1), Lambda (C.37), Omicron (B.1.1.529) which are used for the selection of mutations through AI (Artificial intelligence).

Variants name	Mutations in Spike (S) glycoprotein	
	Other than RBD region	RBD region
Alpha (B.1.1.7)	HV 69–70 deletion, Y144 deletion, A570D, P681H, T716I, S982A, D1118H	E484K, N501Y, A522S
Beta (B.1.351)	L18F, D80A, D215G, R246I, D614G A701V	K417N, E484K, N501Y
Delta (B.1.617.2)	G142D, T19R, R158G, D614G, P681R, D950N, E156del, F157del	P384L, V401L, A411S, L452R, T478K, E484Q, F490S
Gamma (P.1)	L18F, T20N, P26S, D138Y, R190S, H655Y, T1027I, D614G,	K417T, E484K, N501Y, A520S
Lambda (C.37)	G75V, T76I, Δ246-252, D614G T859N	L452Q, F490S
Omicron (B.1.1.529)	A67V, Δ69-70, T95I, G142D, Δ143-145, Δ211, L212I, ins214EPE,T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F	G339D, R346S, S371L, N501T, S373P, S375F, K417N, D427N, N439K, L452R, N440K, G446S, Y449S, S494P, T470N, A475V, S477N, T478K, E484A, Q493K, G496S, Q498R, N501Y, Y505H, R357K

SARS-CoV-2 variants, especially Omicron and its subvariants which were generated due to different mutations. Therefore, there is an urgent need for a broad range of protection against the infection of all kinds of SARS-CoV-2 variants and subvariants.

Numerous research studies have identified vital non-synonymous or point mutations and nucleotide deletions arising in the emerging new variants of SARS-CoV-2. These variants and their possible impacts reflect on the inclusive structure and the function of SARS-CoV-2 viral encoded proteins. Thus, any structural change of spike glycoprotein due to mutation affects function, and it is responsible for the hindrance in protection by vaccination programs. The newly emerging variants alter its characteristic features and boost cellular invading power. Therefore a new mutation-proof modified vaccine candidate is urgently needed that mostly impels with significant SARS-CoV-2 variants. We also previously urged a mutation-proof vaccine construct that can provide robust immune protection against all the forthcoming variants and subvariants [17].

Considering such facts, our study aims to design a multi-epitopic peptide-based vaccine candidate considering all the previous emerging variants' mutations, especially the RBD regions' mutations. Here, one of the objectives is that the vaccine candidate should efficiently stimulate adaptive and innate immune responses in the body of host cells. Successively, this mutation-proof vaccine should act as a broad-range immune booster for protecting against all the future variants and subvariants of SARS-CoV-2.

In the present study, we developed a mutation-proof, single construct containing multi-epitopic peptide sequences based on advanced computational and bioinformatics approaches. In the first part of our study, we selected the CTL epitopes, HTL epitopes, and B cell epitopes from the spike glycoprotein (S-protein), especially from the RBD region of all emerging variants to develop the vaccine construct considering the massive number of RBD mutations. The specific epitopes (9–mer) with selected mutations were used for vaccine development. We used AI-enabled approaches and the top-ranked antigenic score selection method for mutation selection. In the second part of our study, the selected top-ranked, most promising, common, antigenic B cell and T cell epitope (CTL and HTL) are joined with the adjuvants and PADRE sequence within the N-terminal of vaccine construct, which was added by the suitable linker.

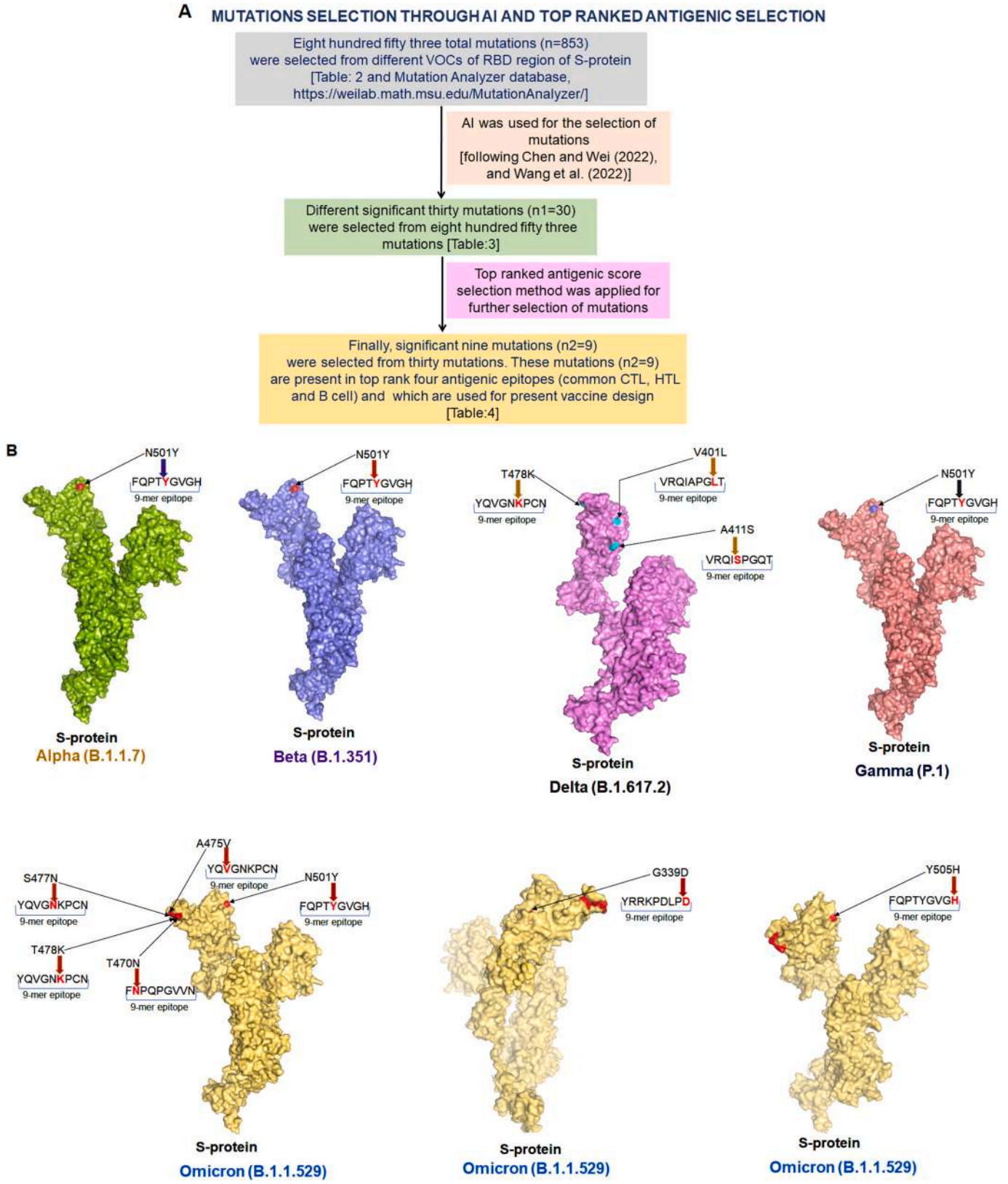


Fig. 1. Different methodologies were adopted (AI models, machine learning in silico techniques, and immunoinformatics approaches), from mutation selection to designing the next-generation multi-epitopic peptide vaccine candidate of S-protein SARS-CoV-2. (A) A flowchart depicts the methodologies adopted for the mutation selection used in the epitopes for our vaccine design. (B) A 3D model demonstrated all the finally selected nine mutations and their corresponding epitopes in the RBD region of S-protein of different SARS-CoV-2 variants (Alpha, Beta, Delta, Gamma, and Omicron). (C) A flowchart illustrates the methodologies adopted (AI models, machine learning in silico techniques, and immunoinformatics approaches) from mutation selection, multi-epitopic peptide vaccine construct design, and its characterization of several properties such as physicochemical properties, antigenicity, allergenicity, etc.

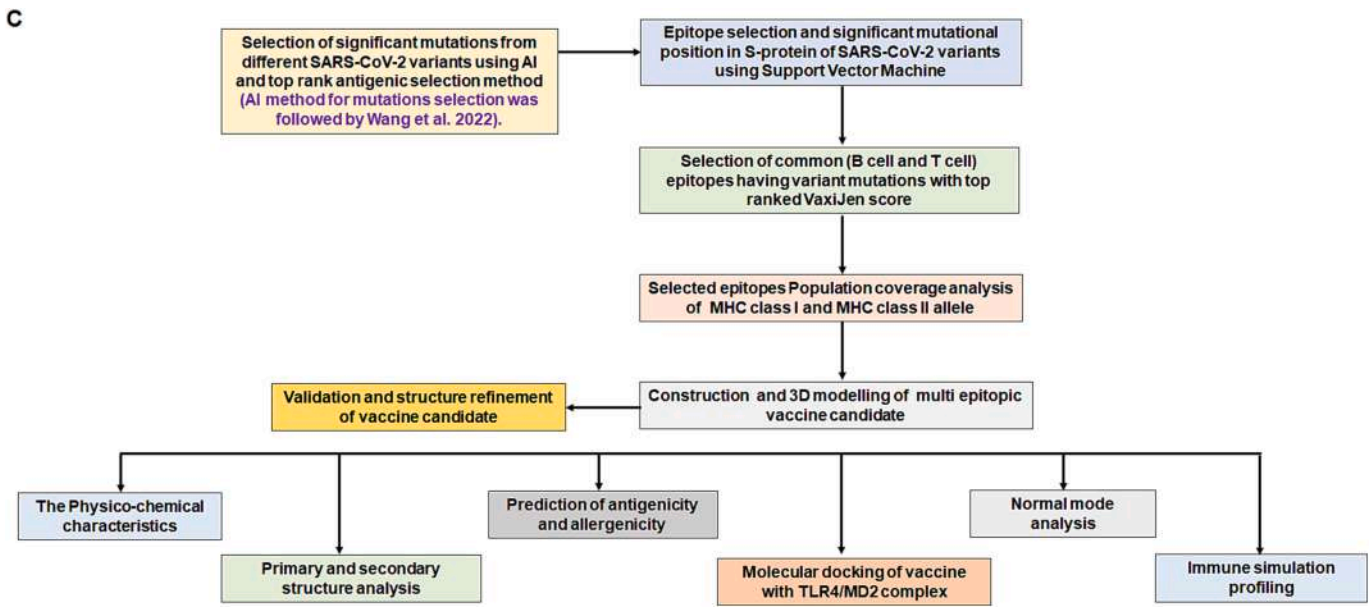


Fig. 1. (*continued*).

Table 3
It listed top mutations (n2 = 30), selected through AI-enabled approaches from the RBD region of S-protein (according to using the Mutation Analyzer database).

Sl. no.	Mutation with amino acid position	Global rank	Worldwide count
1.	N501Y	3	936,596
2.	L452R	2	1,990,939
3.	T478K	1	2,053,531
4.	E484K	5	125,704
5.	K417T	19	67,533
6.	S477N	4	126,383
7.	N439K	20	19,531
8.	K417N	6	102,746
9.	F490S	23	6928
10.	S494P	24	6225
11.	N440K	15	87,781
12.	E484Q	21	8852
13.	L452Q	27	3970
14.	A520S	26	5080
15.	N501T	35	2681
16.	R357K	38	2307
17.	A522S	32	3251
18.	R346K	12	7814
19.	V367F	42	1737
20.	N440S	49	1389
21.	P384L	40	1919
22.	Y449S	54	1212
23.	D427N	50	1376
24.	R346S	51	1327
25.	A475V	43	1647
26.	G339D	8	90,853
27.	V401L	105	279
28.	A411S	37	2401
29.	T470N	70	657
30.	Y505H	12	90,504

In the third part of our study, several properties of the vaccine construct were analyzed, such as physicochemical properties, antigenicity, allergenicity, etc. Simultaneously, we have modeled the 3D structure of the vaccine candidate and docked it alongside the TLR-4/MD2 complex to confirm its efficacy for inducing an immune response. Finally, immune simulation was performed using the machine learning-based tool to understand its effect after administration into the human body.

Table 4
Significant mutations within the RBD of S-protein of some of the SARS-CoV-2 variants of Alpha (B.1.1.7), Beta (B.1.351), Delta (B.1.617.2), Gamma (P.1), Omicron (B.1.1.529) which are having in S-protein (9-mer) epitopic regions and also used for the multi-epitopic vaccine development. The table also shows the mutations' corresponding epitopes used in the vaccine design.

Variants name	Mutations in RBD region of Spike (S) glycoprotein	Mutations and their corresponding epitopes (9-mer) used for our vaccine designing
Alpha (B.1.1.7)	N501Y	FQPTYGVGH
Beta (B.1.351)	N501Y	FQPTYGVGH
Delta (B.1.617.2)	V401L, A411S, T478K	VRQIAPGLT, VRQISPGQT, YQVGNKPCN
Gamma (P.1)	N501Y	FQPTYGVGH
Omicron (B.1.1.529)	G339D, T470N, A475V, S477N, T478K, N501Y, Y505H	YRRKPDLPD, FNPQPGVVN, YQVGNKPCN, FQPTYGVGH

2. Material and methods

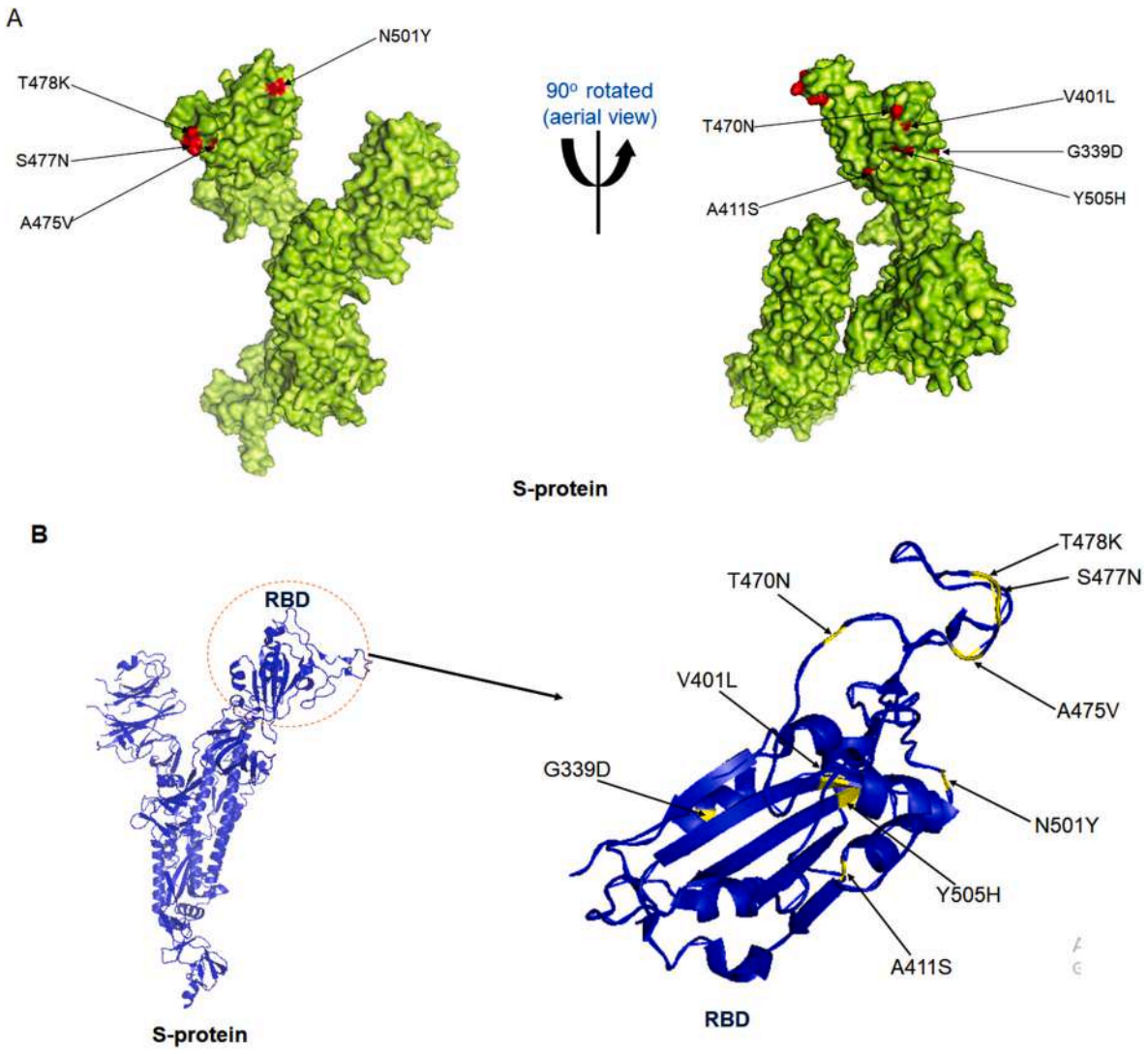
We followed subsequent steps using advanced computational tools, servers, and programs to rationalize a mutation-proof multi-epitopic peptide vaccine. We have used AI-based tools and servers for mutation selection. The machine learning approaches were used for immune simulation. Every epitope selected from the S-protein sequence of SARS-CoV-2 variants is verified within their mutational position.

2.1. Selection of significant mutations from SARS-CoV-2 variants using AI enabled approaches, and top ranked antigenic score selection method

We have studied and collected all significant mutations in the SARS-CoV-2 genome (structural and non-structural protein coding). The reported mutations are also enlisted in Table 1.

Here, we considered the S-protein for vaccine design because S-protein contains the main antigen components compared to all the structural proteins of this virus [18,19].

These significant mutations were found within the S-protein of five SARS-CoV-2 variants, namely Alpha (B.1.1.7), Delta (B.1.617.2), Beta (B.1.351), Omicron (B.1.1.529), Gamma (P.1), and other variants. These variants are used for the selection of mutations through using AI



(caption on next page)

Fig. 2. The schematic diagram 3D model selected RBD mutations for our vaccine construct. (A) A 3D surface model shows the location of selected RBD mutations within the S-protein. (B) Another ribbon-like model shows the location of selected RBD mutations within the ribbon diagram of the RBD region of the S-protein. (C) A model depicted using a crisscross line diagram demonstrates the relatedness between our selected RBD mutations and the occurrence of SARS-CoV-2 variants. (D) A heatmap was described to understand the positions of the different ranks of our chosen mutations which were used to develop the vaccine construct. Here we used three global ranks (worldwide observed rank, BFE change rank, and antibody disruption rank) retrieved from the mutation analyzer database and depicted the heatmap.

(Artificial Intelligence) -based tools and top-ranked antigenic score selection method, which is recorded in Table 2. The RBD mutations are more immunodominant and also responsible for vaccine escape compared to other domains of the S-protein of SARS-CoV-2 [20–22]. Therefore, we finally concentrate on RBD mutations of different VOCs and VOIs for the study. In this study, we consider all the RBD mutations reported till today, and these mutations were collected from the SARS-CoV-2 Mutation Analyzer database [23] and also from the Table 2.

For mutation selection, we have performed the following steps: First, eight hundred fifty-three total mutations (n = 853) were considered from the RBD of S-protein of different VOCs and VOIs (RBD mutations mentioned in Mutation Analyze database and Table 2). Following Chen and Wei (2022), and Wang et al. (2022) [24,25], AI was used for a further selection of mutations. We selected thirty mutations (n1 = 30) from eight hundred fifty-three mutations (Table 2). Again, the top-ranked antigenic score selection method was used for another selection of mutations, and we finally decided on nine mutations (n2 = 9) for the current vaccine design (Fig. 1A). Now, the number of mutation-selection can describe as follows:

$$n \rightarrow n1 \rightarrow n2$$

Where n, n1, and n2 were the different numbers of mutations selected in the various steps. And finally, we chose the n2 number of mutations in the current vaccine design.

Considering SNP in the S-protein of RBD, a mutations database was developed by Wei and his colleagues. There is a vast BFE change reported in SARS-CoV-2. Considering BFE changes, a database was developed by those researchers consisting of 853 mutations. The datasets of the databases have been used for mutation selection using AI/machine learning model. Chen and Wei (2022) used the same model to design a mutation-proof COVID-19 monoclonal antibody using mathematical AI [24]. From 853 mutations, we selected 35 mutations which are listed in Table 3. Similarly, Wang et al. (2022) have prepared a list of top-ranked 25 mutations from all significant emerging variants [25]. Our selected 35 mutations contain all these 25 mutations. At the same time, we found 9-mer, top-rank 12 common T cell and B cell antigenic epitopes (Table 6). Our study noted nine mutations present in four 9-mer epitopes. We used these four 9-mer epitopes for vaccine design. We used eight other top-rank antigenic epitopes for vaccine design along with these four epitopes (with nine mutations). Therefore, a total of 12 antigenic epitopes were used for vaccine design.

2.2. Epitope selection with significant mutational positions in S-protein of SARS-CoV-2 variants

The common B and T cell epitopes against MHC allele (class I and class II) from the S-protein of the SARS-CoV-2 virus is screened out. From our previous study targeted the potent epitope prediction, we used the S-protein sequence of the Wuhan strain of the SARS-CoV-2 virus [10]. We used the BCPRED server to predict linear B cell epitopes within the selected antigenic S-protein. This server performed Support Vector Machine (SVM) algorithm to identify and classify the 9-mer epitopes with a default specificity of 75 % for B cell receptors by using a FASTA (amino acids sequence) as input [26]. The best-scoring epitope was selected for the advanced development of the multi-epitope vaccine.

For the MHC class I binding prediction, the IEDB server was used to calculate the interaction with different MHC I alleles [27]. This server used numerous methods for calculating the binding affinity of a specific amino acid sequence to a precise MHC class I molecule. In contrast, the ANN (artificial neural network) method was used to compute the half-maximal inhibitory concentration (IC_{50}) values of targeted peptide binding to MHC I molecules. The alleles with an IC_{50} binding affinity of 100 nM or less were selected for further study. At the same time, the variable was set up to 0.5. Similarly, the conditions of the specificity and sensitivity setup were considered as 0.94 and 0.89, correspondingly [28]. For the T cell epitope predictions targeted to develop a multi-epitopic vaccine, the specially designed computational server (NetCTL1.2 36) is intended to detect human CTL epitopes within a target protein sequence. The MHC class I immunogenicity was confirmed of the selected epitopes by using at IEDB v2.24 server.

From the outputs of BCPRED and IEDB v2.24 webserver for the recognized B cell and T cell epitopes, we further categorised common 9-mer epitopes, both non-antigenic and antigenic types. Later, we selected the twelve common epitopes (antigenic) with the uppermost antigenic score (VexiJen score).

Simultaneously, the overlapping mutational regions (for nine mutations) are also calculated and listed in Table 3 after analyzing the epitopes and the position of amino acids within the S-protein of SARS-CoV-2 variants. Further, we also compared the antigenicity score of these selected epitopes due to mutational changes in the specified region of S-protein of different significant SARS-CoV-2 variants.

A schematic diagram demonstrated all the finally selected nine mutations and their corresponding epitopes in the RBD region of S-protein of different SARS-CoV-2 variants (Alpha, Delta, Beta, Gamma, and Omicron) (Fig. 1B).

2.3. Construction of mutation proof multi-epitopic peptide vaccine candidate

The epitope cluster analysis tool (IEDB) was engaged at a lowest sequence identity threshold of 100 % to search out the most identical peptides among the top B and T cell epitopes. The clusters and exclusive epitopes were employed and recombined to make the multi-epitope vaccine molecules. The top-ranked epitopes were placed first in each serial design, followed by the adjuvant. Generally, the adjuvant helps the vaccine candidate become more immunogenic. It allows the vaccine candidate to interact with the TLRs, which assists in generating more steady and robust immunological responses within the host cell [29]. Subsequently, the selected epitopes are mixed with an effective adjuvant and connected with the appropriate peptide linkers. The peptide linker performed a vital role in refining the epitope separation and support in epitope presentation towards the MHC class I and class II receptors and subsequent immunological processing [30]. The TLR4 agonist was utilized as an essential adjuvant because the viral glycoproteins detect TLR4, and adjuvants are required for overcoming the protein translation and synthesis restrictions [31]. Therefore, the adjuvant (CTxB) was added to the N-terminal part of the peptide and tried to improve the immunogenicity of the vaccine construct [32]. The stated adjuvant was further connected to the vaccination construct front using the bi-functional linker (EAAAK). It also breaks separately into two domains with feebly interacting contacts within a wide range of peptide chain lengths. Conversely, the GPGPG and AAY peptide linkers were used to attach the chosen common epitopes of B cell and T cell.

The 'junctional epitope' is avoided using the GPGPG peptide linker, which enables immune processing. Whereas the bi-lysine KK linker aids in holding the vaccine construct in distinct immunogenic features.

Table 5
List of common B cell and T cell epitopes, the position of epitopic regions, and antigenicity score of SARS-CoV-2.

Serial no	Common B and T cell epitopes	MHC-I alleles	MHC-II alleles	Positions	Antigenicity score
1.	VRQIAPGQT	HLAB*2702 HLAB*2705	DRB1_0101 DRB1_0102 DRB1_1104 DRB1_1106 DRB1_1107 DRB1_1307 DRB1_1311 DRB5_0101 DRB5_0105	407–415	0.8675 (Probable ANTIGEN)
2.	YQAGSTPCN	HLAB*2702 HLAB*5401 HLA-B*51	DRB1_0305 DRB1_0309	473–481	0.4992 (Probable ANTIGEN)
3.	FQPTNGVGF	HLAB*2702 HLAB*2705 HLAB*3501 HLAB*4403 HLAB*5301 HLAB*5401 HLA-B*51 HLA-B62 HLACw*0702 MHC-Db MHC Db revised MHC-Dd	DRB1_0701 DRB1_0703 DRB5_0101 DRB5_0105	497–505	0.5711 (Probable ANTIGEN)
4.	ILPDPSKPS	MHC-Dd	DRB1_0301 DRB1_0305 DRB1_0306 DRB1_0307 DRB1_0308 DRB1_0309 DRB1_0311 DRB1_0401 DRB1_0421 DRB1_0426 DRB1_1107	805–813	1.2217 (Probable ANTIGEN)
5.	VRQIAPGVT	HLA-B*2702 HLA-B*2705	DRB1_0101 DRB1_0102 DRB1_1104 DRB1_1106 DRB1_1107 DRB1_1307 DRB1_1311 DRB5_0101 DRB5_0105	394–402	0.8675 (Probable ANTIGEN)
6.	LTPRSVRSV	HLA-B*51 HLA-B*5801 HLA-B61 MHC-Dd	DRB1_0402 DRB1_0801 DRB1_0802 DRB1_0804 DRB1_0806 DRB1_1102 DRB1_1120 DRB1_1121 DRB1_1301 DRB1_1302 DRB1_1322 DRB1_1327 DRB1_1328	745–753	1.1039 (Probable ANTIGEN)
7.	LGNSTGIDF	HLA-B*5301 HLA-B*51 HLA-B*5801 HLA-B62 MHC-Dd MHC-Ld	DRB1_0701 DRB1_0703	1223–1231	0.9917 (Probable ANTIGEN)
8.	YRRKPDLPN	HLA-B*2702 HLA-B*2705 HLA-B*5401 HLA-B*51	DRB1_0801 DRB1_0802 DRB1_0817	331–339	1.1352 (Probable ANTIGEN)
9.	FTPQPGAVS	HLA-A20 Cattle HLA-B*3902 HLA-B*5401 HLA-B*51 HLA-B*5801 MHC-Dd	DRB1_0309 DRB5_0101 DRB5_0105	469–477	1.3768 (Probable ANTIGEN)
10.	VKSGSPGDS	HLA-A20 Cattle	DRB1_0301	531–539	1.0477 (Probable ANTIGEN)

(*continued on next page*)

Table 5 (*continued*)

Serial no	Common B and T cell epitopes	MHC-I alleles	MHC-II alleles	Positions	Antigenicity score
			DRB1_0305 DRB1_0309 DRB1_1107		
11.	VRPRNSSDN	HLA-B*2702 HLA-B*2705 MHC-Db	DRB1_0402 DRB1_0801 DRB1_0802 DRB1_0804 DRB1_0806 DRB1_0813 DRB1_0817	745–753	0.9475 (Probable ANTIGEN)
12.	ITAYDPRSC	HLA-B*5801	DRB1_0802 DRB1_0804 DRB1_1102 DRB1_1121 DRB1_1322	525–533	0.8468 (Probable ANTIGEN)

Finally, an extra PADRE sequence (13 amino acids chain) was added at the C-terminal end of vaccine constructs to increase the immunogenicity in host cells.

2.4. Antigenicity and allergenicity prediction of the vaccine candidate

The antigenicity of the multi-epitopic vaccine construct has been predicted through the VaxiJen webserver. Based on the ACC (Auto Cross Covariance) alteration of targeted peptide sequences, this server trailed a new alignment-independent method specifically for antigen prediction [33]. This server is a steady and consistent server for predicting the protective antigens using the level of peptides antigenicity containing a threshold score in default mode targeted for viruses and other studied organisms.

Consequently, the AllerTop 2.0 web server has been used to recognize the vaccine construct allergenicity. This specialized web server-based on ACC method and transforms the chain of amino acids sequence as an equal-length vector [34].

2.5. Primary, secondary structure and physiochemical analysis of vaccine candidate

The Expasy ProtParam tool has been applied to predict and access the primary structure of the multi-epitopic vaccine construct. Subsequently, different physiochemical properties of the input peptide sequence were also characterized, including aliphatic index, isoelectric point (pI), numbers of amino acids, GRAVY (grand average of hydropathicity), molecular weight, net charges, instability index, and assessed half-life [35]. The specialized tool "ProtParam" is employed, which relies on the "N-end rule" and relates the half-life of a protein to the identity of its N-terminal residue; the prediction is given for three model organisms (human, yeast, and *E. coli*). The N-end rule originated from the observations that the identity of the N-terminal residue of a protein plays a vital role in determining its stability in vivo models. For analysis of the secondary structure (transmembrane helices, globular regions, bend regions, random coil, and coiled-coil regions) of the proposed vaccine candidate, the SOPMA (Secondary Structure Prediction Method) computational tools and PSIPRED 4.0 server are used [36,37].

2.6. Three-dimensional (3D) modeling and structure refinement of the vaccine candidate

Understanding of molecular functions at the atomic level in three-dimensional (3D) protein shapes provides essential information. Here, we used the I-TASSER web server to predict the 3D structure of the developed vaccine [38]. This server performed an integrated platform targeted to the prediction of automated protein function and structure-based approach on the paradigm of sequence-to-structure-to-function. It developed a fine-quality model comprising accurate coordinate formation equated to other known protein structures. Furthermore, the GalaxyWEB server was used to refine the predicted 3D model of the vaccine candidate [39].

2.7. Structural validations of peptide-based vaccine construct

Two critical web servers have been employed to precisely validate the vaccine's predicted 3D model protein structure: ProSA and PROCHECK. The ProSA server was applied to confirm the protein tertiary structure, where the overall quality scoring was considered by the local quality model and the 'Z' plot [40]. Whereas the PROCHECK web server studies the stereochemical quality of protein structure generates several PostScript plots, and analyzes its residue-by-residue geometry by forming the Ramachandran plot [41].

2.8. Molecular docking and binding affinity prediction of vaccine candidate with the human TLR4/MD2 complex

For eliciting proper immune responses against microbial infection, the human TLR4/MD2 complex performed a crucial role in the host cell. Therefore, the interaction of the epitopic peptide-based vaccine with the targeted immune cell proteins is much more critical for creating a suitable immune response. To perform the molecular docking of the vaccine candidate and human TLR4/MD2 complex (PDB ID: 4G8A), we employed the HawkDock server [42]. The HawkDock server calculates binding interactions pattern with the source of computing mechanisms of MM/GBSA free energy based on the top-ranked model. Later, these interactive molecular patterns were selected as per the primary structure at the atomic level for commencing the interactive molecular scores (atomic contact energy) [43].

Additionally, for calculating protein-protein binding interactions of our proposed vaccine construct with the TLR4/MD2 complex, we used the cross-platform molecular graphic tool, PyMoL [44]. This proprietary molecular visualization system eliminates the undesired elements from the accurate coordinates of the PDB structure. The binding affinity in the form of Gibbs free energy (ΔG) determined the definite form of interaction pattern and dissociation constant (KD) values of vaccine construct with the TLR4/MD2 complex has been predicted at 37 °C temperature by using the PRODIGY web server [45].

2.9. Normal mode analysis of multi-epitopic peptide-based vaccine

The iMODS server has been applied to validate the cooperative motion of peptide-based multi-epitopic vaccine over the normal mode within the inner coordinates as normal mode analysis (NMA) [46]. Subsequently, an added program of the iMODs, identified as the essential dynamics simulation (EDM), was also applied to recognize the stability and the macromolecular mobility of the vaccine and TLR4/MD2 docked complex. The evaluated parameters of this docked complex are

Table 6
List of common B cell and T cell epitopes, the position of epitopic regions, comparative antigenicity score due to the mutation, and mutations of SARS-CoV-2 variants in S-protein selected for vaccine development.

Serial no	Common B and T cell epitopes after mutation	MHC-I alleles	MHC-II alleles	Positions	Comparison of antigenicity before and after mutation		Mutation
					Antigenicity score without mutation	Antigenicity score due to mutation	
1.	VRQISPGQT	HLAB*2702 HLAB*2705	DRB1_0101 DRB1_0102 DRB1_1104 DRB1_1106 DRB1_1107 DRB1_1307 DRB1_1311 DRB5_0101 DRB5_0105	407–415	0.8675 (Probable antigen)	1.1298 (Probable antigen)	A411S
2.	YQVGNKPCN	HLAB*2702 HLAB*5401 HLA-B*51	DRB1_0305 DRB1_0309	473–481	0.4992 (Probable antigen)	0.6648 (Probable antigen)	A475V S477N T478K
3.	FQPTYGVGH	HLAB*2702 HLAB*2705 HLAB*3501 HLAB*4403 HLAB*5301 HLAB*5401 HLA-B*51 HLA-B62 HLACw*0702 MHC-Db MHC Db revised MHC-Dd	DRB1_0701 DRB1_0703 DRB5_0101 DRB5_0105	497–505	0.5711 (Probable antigen)	0.691 (Probable antigen)	N501Y Y505H
4.	ILPDPSKPS	MHC-Dd	DRB1_0301 DRB1_0305 DRB1_0306 DRB1_0307 DRB1_0308 DRB1_0309 DRB1_0311 DRB1_0401 DRB1_0421 DRB1_0426 DRB1_1107	805–813	1.2217 (Probable antigen)		
5.	VRQIAPGLT	HLA-B*2702 HLA-B*2705	DRB1_0101 DRB1_0102 DRB1_1104 DRB1_1106 DRB1_1107 DRB1_1307 DRB1_1311 DRB5_0101 DRB5_0105	394–402	0.8675 (Probable antigen)	1.1175 (Probable antigen)	V401L
6.	LTPRSVRSV	HLA-B*51 HLA-B*5801 HLA-B61 MHC-Dd	DRB1_0402 DRB1_0801 DRB1_0802 DRB1_0804 DRB1_0806 DRB1_1102 DRB1_1120 DRB1_1121 DRB1_1301 DRB1_1302 DRB1_1322 DRB1_1327 DRB1_1328	745–753	1.1039 (Probable antigen)		
7.	LGNSTGIDF	HLA-B*5301 HLA-B*51 HLA-B*5801 HLA-B62 MHC-Dd MHC-Ld	DRB1_0701 DRB1_0703	1223–1231	0.9917 (Probable antigen)		
8.	YRRKPDLPD	HLA-B*2702 HLA-B*2705 HLA-B*5401 HLA-B*51	DRB1_0801 DRB1_0802 DRB1_0817	331–339	1.1352 (Probable antigen)	1.5874 (Probable antigen)	G339D
9.	FNPQPGVVN	HLA-A20 HLA-B*3902 HLA-B*5401	DRB1_0309 DRB5_0101 DRB5_0105	469–477	1.3768 (Probable antigen)	1.3319 (Probable antigen)	T470N A475V S477N

(*continued on next page*)

Table 6 (*continued*)

Serial no	Common B and T cell epitopes after mutation	MHC-I alleles	MHC-II alleles	Positions	Comparison of antigenicity before and after mutation		Mutation
					Antigenicity score without mutation	Antigenicity score due to mutation	
		HLA-B*51 HLA-B*5801 MHC-Dd					
10.	VKSGSPGDS	HLA-A20 Cattle	DRB1_0301 DRB1_0305 DRB1_0309 DRB1_1107	531–539	1.0477 (Probable antigen)		
11.	VRPRNSSDN	HLA-B*2702 HLA-B*2705 MHC-Db	DRB1_0402 DRB1_0801 DRB1_0802 DRB1_0804 DRB1_0806 DRB1_0813 DRB1_0817	745–753	0.9475 (Probable antigen)		
12.	ITAYDPRSC	HLA-B*5801	DRB1_0802 DRB1_0804 DRB1_1102 DRB1_1121 DRB1_1322	525–533	0.8468 (Probable antigen)		

considered the possible motions marked by the specialized terms of deformability plot, elastic model, B-factors, covariance, and eigenvalue.

2.10. Immune simulation profiling of peptide-based multi-epitope vaccine candidate using machine learning approaches

We used the C-ImmSim server to interpret the immune response profile of our proposed vaccine candidate [47]. This server defines the humoral and cellular response within the mammalian immune system, in the presence of immunogenic elements of the pathogen (bacteria, virus, protozoa, etc.), within the sub-cellular localization level (mesoscopic scale). It also evaluated the definite antigen-mediated immune response by the "position-specific scoring matrix," and short-coded machine learning approaches. Some additional essential parameters for this simulation server were adopted, such as random speed (12345), simulation volume (30), and simulation step (1000). Simultaneously, we have employed the four-week double injection intervals planned to perform the proposed Immune simulation. Nevertheless, the rest of the additional parameters were engaged as the default state [48].

A flow chart was depicted to demonstrate the comprehensive methodologies to complete our work (Fig. 1C).

3. Result

3.1. Selection of significant mutations from SARS-CoV-2 variants using AI enabled approaches, and top ranked antigenic score selection method

We have documented all significant mutations in the different protein-coding parts of the SARS-CoV-2 variant genome. From the 26 types of the structural and non-structural genes of SARS-CoV-2, several significant mutations are reported (Table 1). From these total mutations in the genome, the important mutations within the S-protein (both in RBD and non-RBD regions) of six SARS-CoV-2 variants, namely Alpha, Delta, Beta, Lambda, Gamma, and Omicron, were listed in Table 2. These mutations are used to select mutations through AI (Artificial Intelligence) following Chen and Wei (2022), and Wang et al. (2022) [24,25] along with the SARS-CoV-2 Mutation Analyzer database [23]. After using the AI-enabled approaches, a total of thirty (n1 = 30) mutations were selected from the RBD region of the S-protein (Table 3).

Again, from thirty mutations, we selected nine mutations (n2 = 9) within the RBD of S-protein using the top-ranked antigenic score selection method (Table 4). These nine mutations are present in the five SARS-CoV-2 variants (Alpha, Beta, Delta, Gamma, and Omicron) and other emerging variants. Some selected mutations are common, such as N501Y, which is present in four variants: Alpha, Beta, Gamma, and Omicron.

Again, two types of 3D models were generated to demonstrate the location of our chosen RBD mutations for our vaccine construct. The models are the 3D surface model (Fig. 2A) and the 3D ribbon model (Fig. 2B). At the same time, another model was depicted using a criss-cross line diagram to illustrate the relatedness between our selected RBD mutations and the occurrence of SARS-CoV-2 variants (Fig. 2C). Furthermore, we represented a heatmap to describe the positions of the different ranks of our chosen mutations for the vaccine construct (Fig. 2D). Here we used three global ranks (Worldwide observed rank, BFE change rank, and Antibody disruption rank) using the Mutation Analyzer database and depicted through the heatmap using Heatmapper [49].

3.2. Epitope selection and significant mutational position in S-protein of SARS-CoV-2 variants

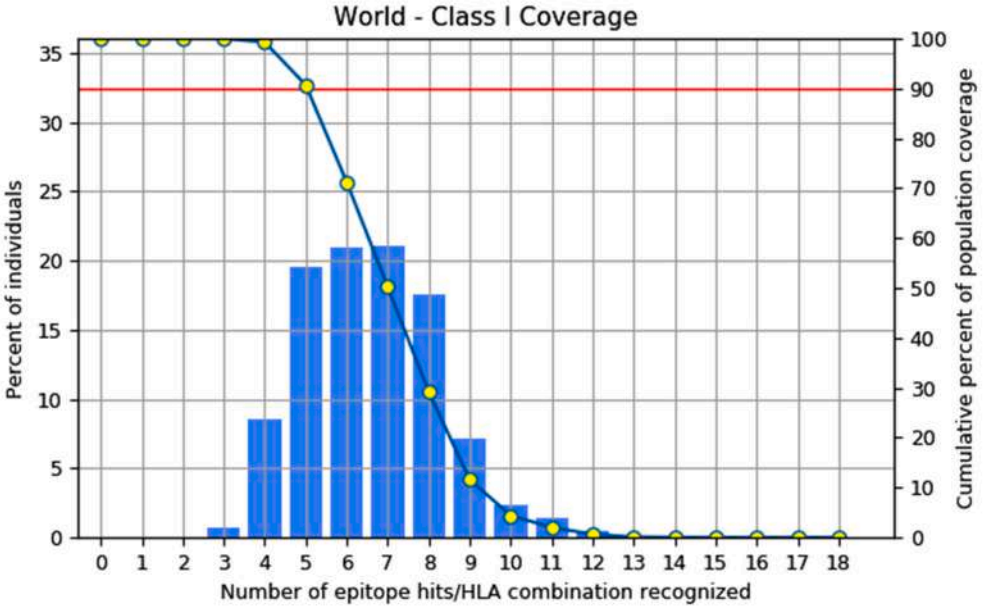
The highly antigenic common (B cell and T cell) epitopes selection is essential for possible vaccine design. To identify MHC class I and MHC class II epitopes, we retrieved 12 nos. epitopic sequences (9-mer) along with the highest antigenic score. Considering the position of amino acids of these epitopic regions within the S-protein of the SARS-CoV-2 virus (Wuhan strain) (Table 5). We have again recalculated the antigenicity of these 12 nos. epitopes antigenicity after the integration of specific mutations overlapped within the amino acids positions. Table 4 enlisted the particular mutations of four significant SARS-CoV-2 variants within the selected common epitopic (B cell and T cell) part. Table 6 shows the comparative antigenicity score and mutations of SARS-CoV-2 variants in S-protein chosen for vaccine development with mutations and without mutations in the epitopic region. The table describes the antigenicity score of the epitopes before mutations and after mutations in the epitopic region. In most cases, the mutations shown in six nos. of epitopes with increased antigenicity scores due to altered amino acids. Therefore, these mutant epitopes performed potent elements for multi-epitopic peptide-based vaccine development.

3.3. Prediction of epitopes population coverage

The population coverage has been developed based on the parallel allele-specific (MHC class I and MHC class II) population coverage (epitope-based) analysis. The used web server defined the population coverage for epitopes of both alleles. The population coverage regarding 12 nos. top-ranked mutations overlapped filtrated epitopes for both the

A *Population: World*

MHC class	Coverage	Average hit	PC90
I	100.0%	6.59	5.04



B *Population: World*

MHC class	Coverage	Average hit	PC90
II	100.0%	10.53	7.23

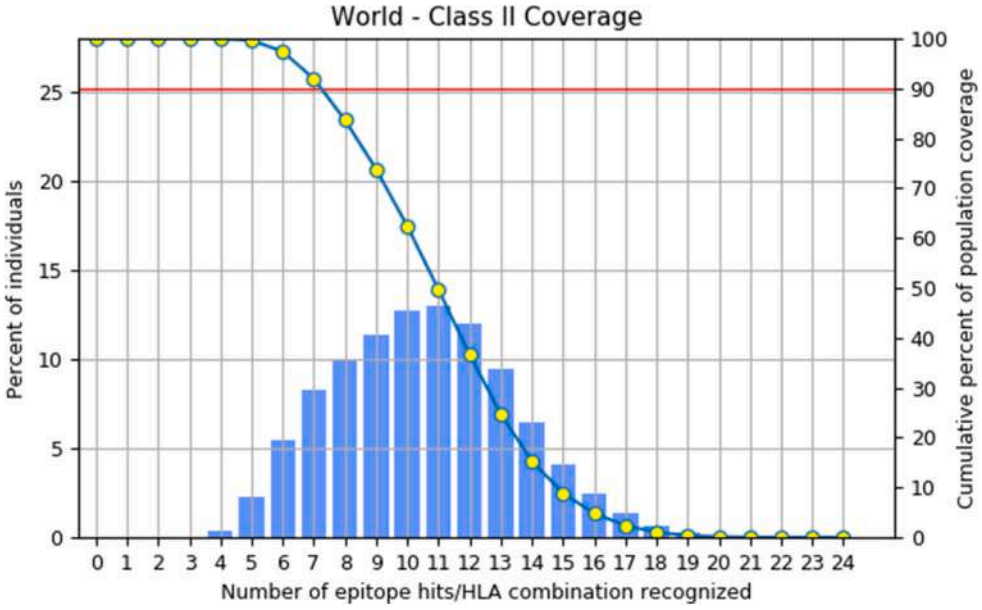


Fig. 3. Population coverage by selected CTL epitopes and their particular MHC binding alleles of designed vaccine construct. We consider two critical alleles for our work (MHC class I and MHC class II) (A) The figure shows the worldwide population coverage of the MHC class I allele, (B) It represents the worldwide population coverage of the MHC class II allele.

Table 7
Evaluated different physicochemical properties of the peptide sequence of the multi-epitopic vaccine candidate (antigenicity, allergenicity, solubility, number of amino acids, molecular weight, theoretical isoelectric point (pI), total number of atoms, estimated half-life, instability index, GRAVY, etc.).

Sl. no.	Features	Assessment	Remark
1.	Antigenicity	0.5703 (antigenic score)	Probable antigen
2.	Allergenicity	Confirmed by AllerTOP v.2.0 server	Probable non-allergen
3.	Solubility	0.569	Soluble
4.	Number of amino acids	205	Suitable
5.	Molecular weight	19,974.55 Da	Average
6.	Theoretical isoelectric point (pI)	9.33	Basic
7.	Total number of atoms	2805	Suitable
8.	Formula	$C_{888}H_{1392}N_{264}O_{259}S_2$	–
9.	Estimated half-life	1 h (mammalian reticulocytes, in vitro) 30 min (yeast, in vivo) >10 h (*Escherichia coli*, in vivo)	–
10.	Instability index	24.39	Stable
11.	Aliphatic index	48.10	Thermostable
12.	Grand average of hydropathicity (GRAVY)	0.632	Hydrophilic

MHC class I and MHC class II alleles are shown in Fig. 3A and 3B.

3.4. Antigenicity and allergenicity prediction of the vaccine candidate

The assessment of the antigenicity of our vaccine construct showed it is a potent antigen with an antigenicity score of 0.5703 (Table 7). This score support that the designed multi-epitope peptide-based construct with the PADRE sequence and adjuvant could boost the immune response within the host body. The illustration shows the developed vaccine construct in Fig. 4.

The allergenicity calculation of the probable vaccine candidate was the predicted result and noted that the vaccine candidate was a safe vaccine.

3.5. Primary, secondary structure and physiochemical analysis of vaccine candidate

The analysis of the secondary structural architecture of the structural properties of our vaccine, such as the alpha helix (16.59 %), extended strand (7.32 %), beta-turn (2.93 %), and random coil (73.17 %) (Fig. 5A and B). The outputs designate an advanced ranking quality of the peptide-based vaccine containing the secondary structure (strand, helix, and coil) (Fig. 6).

We measure twelve crucial physicochemical properties of our vaccine candidate (Table 7). The theoretical isoelectric point (pI) was 9.33, molecular weight (MW) of 19,974.55 Da. The grand average of hydropathicity (GRAVY) was calculated at 0.632. The GRAVY score as positive designated the vaccine candidate protein should be hydrophobic. This hydrophobic nature of the peptide chain might help in the smooth purification and improved design of a future vaccine.

Subsequently, the Aliphatic index (thermo-stability) of the peptide construct was calculated as 48.10. The half-life of this proposed vaccine was calculated as 30 min in yeast (in vivo), 1 h in mammalian reticulocytes (in vitro), and >10 h in *E. coli*. Simultaneously, the instability index was calculated as 24.39. This value states S-protein's stable multi-epitopic vaccine construct derived from SARS-CoV-2.

3.6. Structural validation of peptide-based vaccine-construct

The structural validation of a peptide-based vaccine construct is a crucial step in vaccine development. In this study, two web servers (PROCHECK and ProSA) accurately validated the peptide-based construct structure. These tools evaluate the overall quality of the vaccine structure. The refined vaccine model from the predicted Ramachandran plot showed that the residues exist in four specific states as favored (58.6 %), generously allowed (7.8 %), additionally allowed (27.1 %), and disallowed region (6.5 %) (Fig. 7A, Table 8). The 'Z-score' of our developed vaccine in the form of an advanced model was calculated as −2.86 (Fig. 7B). The negative 'Z-score' specifies the superior 3D model of the developed protein. Afterward, we developed a further configuration to study the local quality model by applying the ProSA webserver. This server is depicted by a significant plot to show the overall quality of the acceptable peptide model (Fig. 7C). The 3D structure of our designed vaccine protein is shown in Fig. 4B (surface model) and 4C (ribbon model).

3.7. Molecular docking and binding affinity prediction of vaccine candidate with the human TLR4/MD2 complex

The molecular docking resulted in the TLR4/MD2 and vaccine construct docked complex (the protein-protein complex) and indicated the binding affinity. The protein-protein docked complex pointed out that this complex is among the top-ranked ten models based on ACE values. The docking cluster containing the highest molecular binding score (−319.56) was selected to visualize and interpret the binding interactions. Also, the HawkDock tool program was engaged to offer a molecular docking complex by a specified machine learning approach, trained with the MM/GBSA. It also noted that the binding free energy (BFE) associated with the molecular docked complex was −96.67 kcal mol^{-1}. This crucial score established a greater binding affinity between the developed multiepitoic vaccine model (ligand molecule) and the human TLR4/MD2 complex (receptor element) (Fig. 8A).

The docking score (HADDOCK score) (a.u.) is noted as -157.9 ± 8.6, and it specifies a suitable interaction between the vaccine construct and the human TLR4/MD2 complex. The score showed that buried surface area (BSA) is 2864.3 ± 99.1 Å2, revealed as contiguity and a less water-exposed area of protein binding surface. Whereas the small RMSD value of docked complex (generated through HADDOCK) identifies, the development of good quality docked complex. The calculated van der Waals energy, desolvation energy, restraints violation energy, electrostatic energy, and Z-Score values are shown in Table 9. At the same time, Table 10 lists the hydrogen bond between both interactive protein molecules. Subsequently, the dissociation constant (Kd) values and the binding affinity of Gibbs free energy (ΔG) of docked complex (analyzed through the PRODIGY web service) indicated the results, which are 4.5E-09 M and −12.7 kcal mol^{-1} (Table 9). It also noted that the negative value of the molecular docking complex had specified the thermodynamically stable protein-protein docking structure.

3.8. Normal mode analysis of multi-epitopic peptide-based vaccine

We analyzed the protein-protein interactions machinery and flexibility in biological macromolecules through the Normal Mode Analysis (NMA). The molecular motion and structural flexibility are computed and combined with the coordinates of the molecular docked complex. The results show the protein domains' mobility, which signifies by the colored affine arrow that directs the clusters. Finally, it depicts the NMA mobility (Fig. 8B). At the same time, the complex deformability primarily depends on the distortion of individual residues (Cα atom), which is also presented through colored hinges within the chain of complex (Fig. 8C).

Furthermore, the server also calculated the eigenvalue as 2.428517e-05 (Fig. 8E). Likewise, the eigenvalue and variance are oppositely linked to each NMA (Fig. 8D). Additionally, the B-factor graph displayed the average RMS value, and, at the same time, the B-factor plot signifies the steady structure of docked molecules presented in Fig. 8D. The

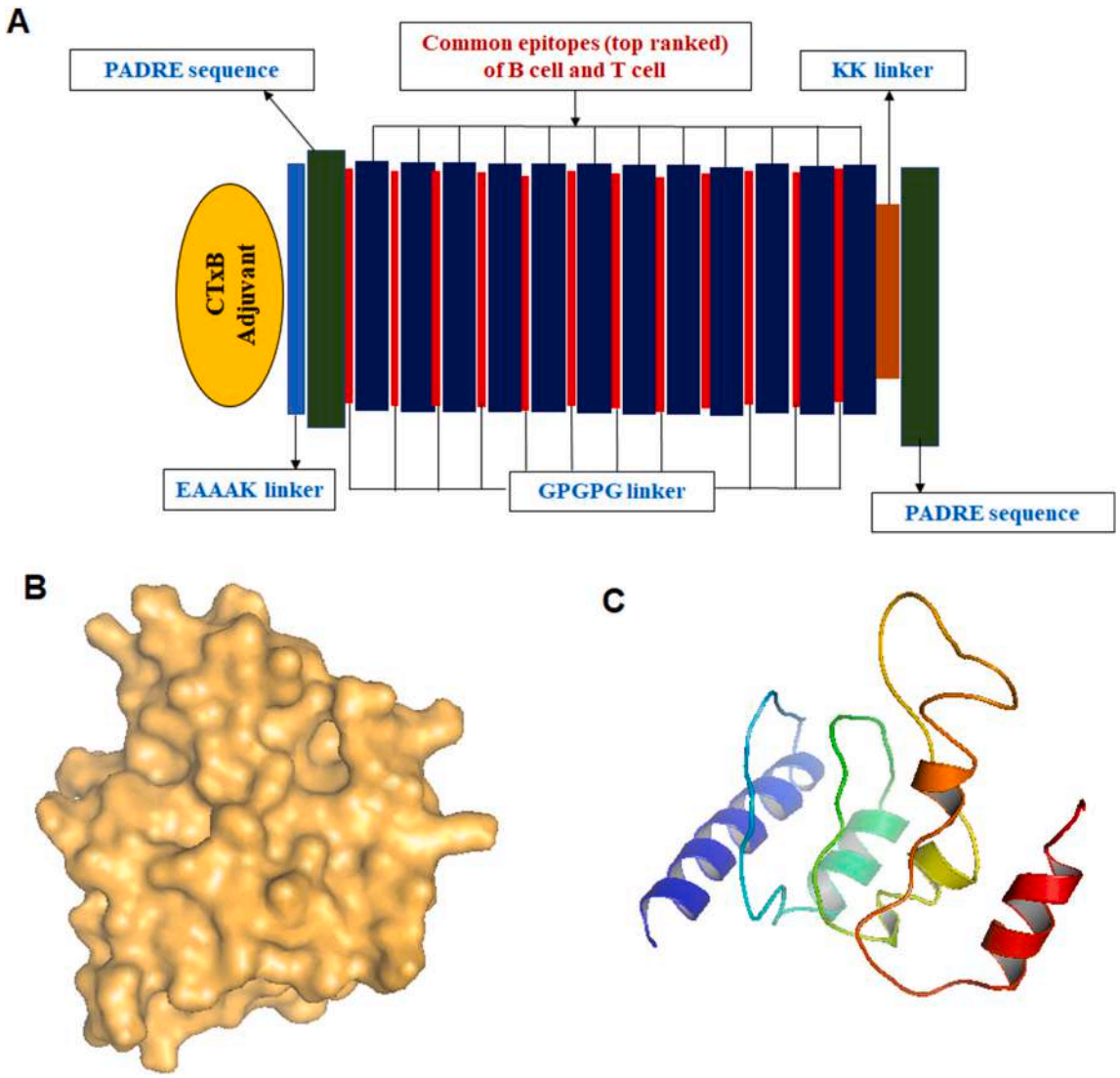


Fig. 4. Peptide vaccine construct and its refined 3D structure from the S-proteins of SARS-CoV-2 and its mutational epitopes. (A) The graphical diagram shows the final vaccine construct from antigenic mutational epitopes of S-proteins of SARS-CoV-2 using different peptide linkers such as AAY and GPGPG. (B) Surface structure model. (C) The ribbon model shows the vaccine construct.

covariance matrix characterizes by individual colors graph where correlated (red), uncorrelated (white), and anti-correlated (blue) (Fig. 8F). The elastic network model display docked protein molecule (Cα) atoms that are interconnected with "springs" of particular strengths (the darker greys symbol of the stiffer springs) (Fig. 8G).

3.9. Immune simulation profiling of peptide-based, multi-epitopic vaccine candidate using machine learning approaches

The in silico technique for immune simulation profiling of the developed vaccine candidate supported primary and secondary immune responses by activating the immune system, comprising the memory cells, HTL, CTL, and associated immune cells (NK cells, DC cells, etc.). The levels of the specific immunoglobulins (IgM + IgG) were detected by drastically increasing after administrating the proposed vaccine construct. The extended immune responses were successfully described against the SARS-CoV-2 through the high titers value of IgG and IgM (Fig. 9A). During the resting and acting phase, the innate immunity in the B cell population and host body was increased with IgM, B-memory cells, and its isotypes, which also scaled up at 680–740 cells/mm^3 (Fig. 9A–C). Similarly, the CTL cells also raised and extended a maximum of 1127 cells/~8 days after immunization applied by vaccine and gradually reduced after 24 days (Fig. 9D).

Furthermore, during the active and resting phase, the T cells showed the highest level of the T cell population (Fig. 9E). Likewise, the elevated HTL cells (both in active and resting phases) achieved a critical role in generating the adaptive immune response against the infections of SARS-CoV-2. The elevated cells can generate the maximum memory cells (Fig. 9F–G). Memory cells show a significant role in preventing and regulating viral infection and reinfection in a chance of self-memorization by encountering pathogens. Therefore, the effective administration of this designed vaccine construct efficiently elevated the added regulators elements of the immune system (e.g., cytokines, interleukins, DC cells, and NK cells) (Fig. 9F–L). Hence, such consequences specify the proposed multi-epitopic peptide vaccine could be considered an effective next-generation peptide-based vaccine for provoking a robust immune response to counteracting the infection by the SARS-CoV-2 virus.

4. Discussion

Last three years, global public health has been under significant

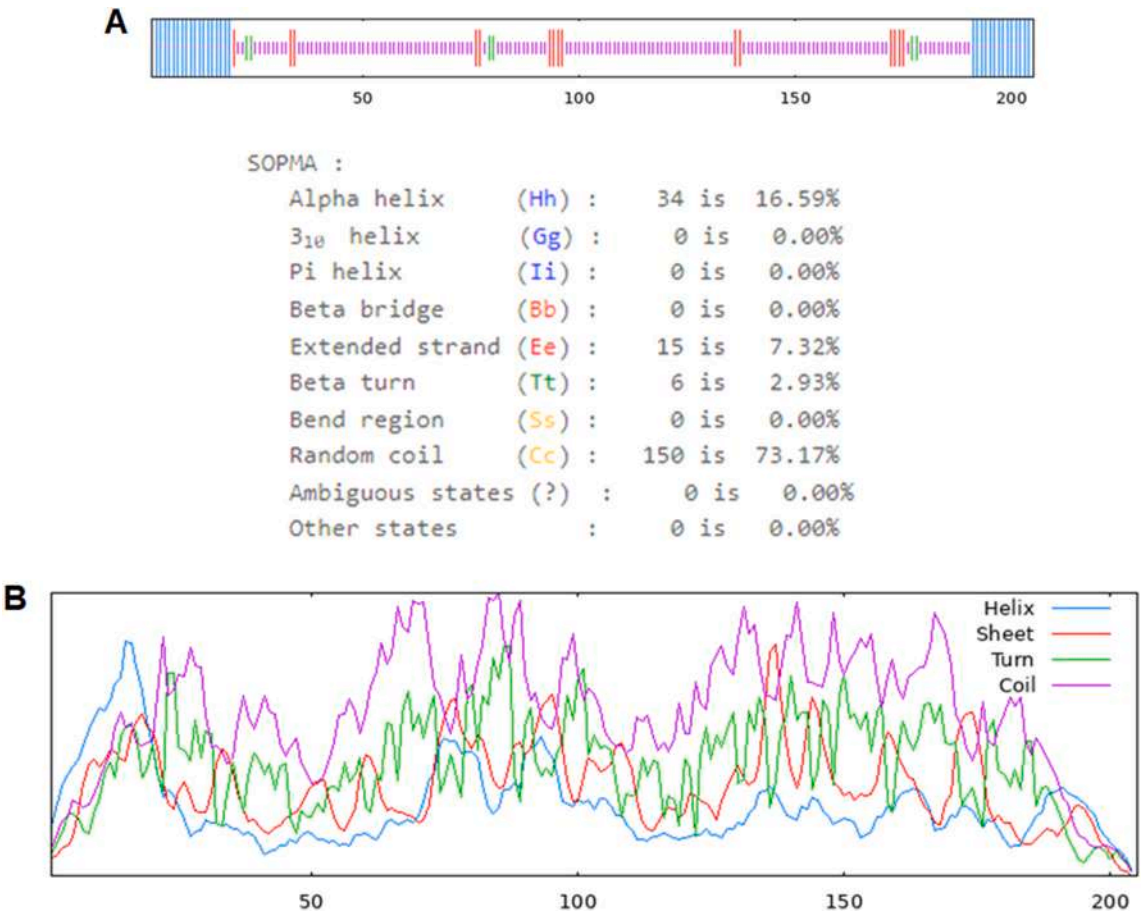


Fig. 5. The secondary structure prediction plot of our vaccine construct. The alpha helices were marked in blue color, while extended strands and beta turns were shown in red and green colors, respectively. (A) It demonstrates the visualization of the result of the study and (B) The score curves for each predicted state of peptide chain analysis.

threat due to the SARS-CoV-2 infection. Numbers of drugs and vaccines are already available to resist infection and boost up human immune system [50]. However, until today, no vaccine can fight against the disease caused by diverse mutational variants of SARS-CoV-2. The number of mutations present in the S-protein of these virus variants ultimately caused vaccine escape [51–56]. Therefore, the work aims to develop a mutation-proof COVID-19 vaccine candidate with a broad range of immune protection against Omicron and its forthcoming subvariants. In this direction, the current research work designed a multi-epitope, next-generation peptide vaccine candidate with significant mutations overlapping in epitopes of SARS-CoV-2 S-protein. These variant-specific mutations co-exist within the epitopes, and the vaccine candidate should be a strong mutation-proof vaccine candidate. The vaccine candidate was formulated through advanced computational approaches and tools such as AI-based approaches, machine learning, and programs that can induce an immunological response within the host cell [11]. Using AI-enabled strategies, and top-ranked antigenic scores, we selected nine top-ranked RBD mutations from 853 total mutations (n = 853) following the method of Chen and Wei (2022), Wang et al. (2022) and Doytchinova and Flower (2022) [24,25,33]. We considered the RBD mutations. Because most cases, the RBD mutations are responsible for immune and vaccine escape [57]. Several other people have also tried to develop the vaccine considering the RBD region. Recently, using trimeric RBD of SARS-CoV-2 (mos-tri-RBD), Zhang et al. developed a vaccine that can protect the SARS-CoV-2 Omicron and several other variants [58]. However, in this study, we consider the diverse number of RBD mutations to develop the mutation-proof SARS-CoV-2 vaccine, which can protect SARS-CoV-2 Omicron variants and all forthcoming subvariants. Therefore, in this point of view, our vaccine candidate is significant. At the same time, our study's importance is that it used an AI model for mutation selection and a machine-learning approach for Immune simulation.

Therefore, we have identified the common (B cell and T cell) epitopes (nine amino acids window length) from the S-protein of the SARS-CoV-2 virus. The epitopes (9-mer) twelve nos. having the highest antigenic scores are selected for future analysis, which contains nine top-ranked mutations. All these mutations are selected from the significant mutations located in RBD of S-Protein in the significant SARS-CoV-2 VOIs and VOCs. We consider a total of 853 RBD mutations (n = 853) for mutation selection, which is a vast number of mutations. However, Chen and Wei (2022), and Wang et al. (2022) initiated the task of mutation selection through AI (Artificial Intelligence) model from mutation-based "big data" analysis [24,25]. Finally, twelve numbers of the highest antigenic epitopes with some altered amino acids were considered for vaccine design. Suitable peptide linkers and adjuvants added these epitopes with mutations to develop the final vaccine construct. It also combines the antigenic T cell and B cell epitopic sequences with overlapping mutation (altered amino acids) following a chemical process that ultimately stimulates specific production of immunoglubulins [59,60].

The distribution patterns and expression profile of HLA (MHC class I and MHC class II) alleles may differ worldwide between ethnicities and regions specific. Successively targeting the formulation of a functional vaccine, it is crucial to perform the standard assessment of the allelic distribution of HLA inside the whole world population [61,62]. In our study, we have found the worldwide population coverage of the MHC

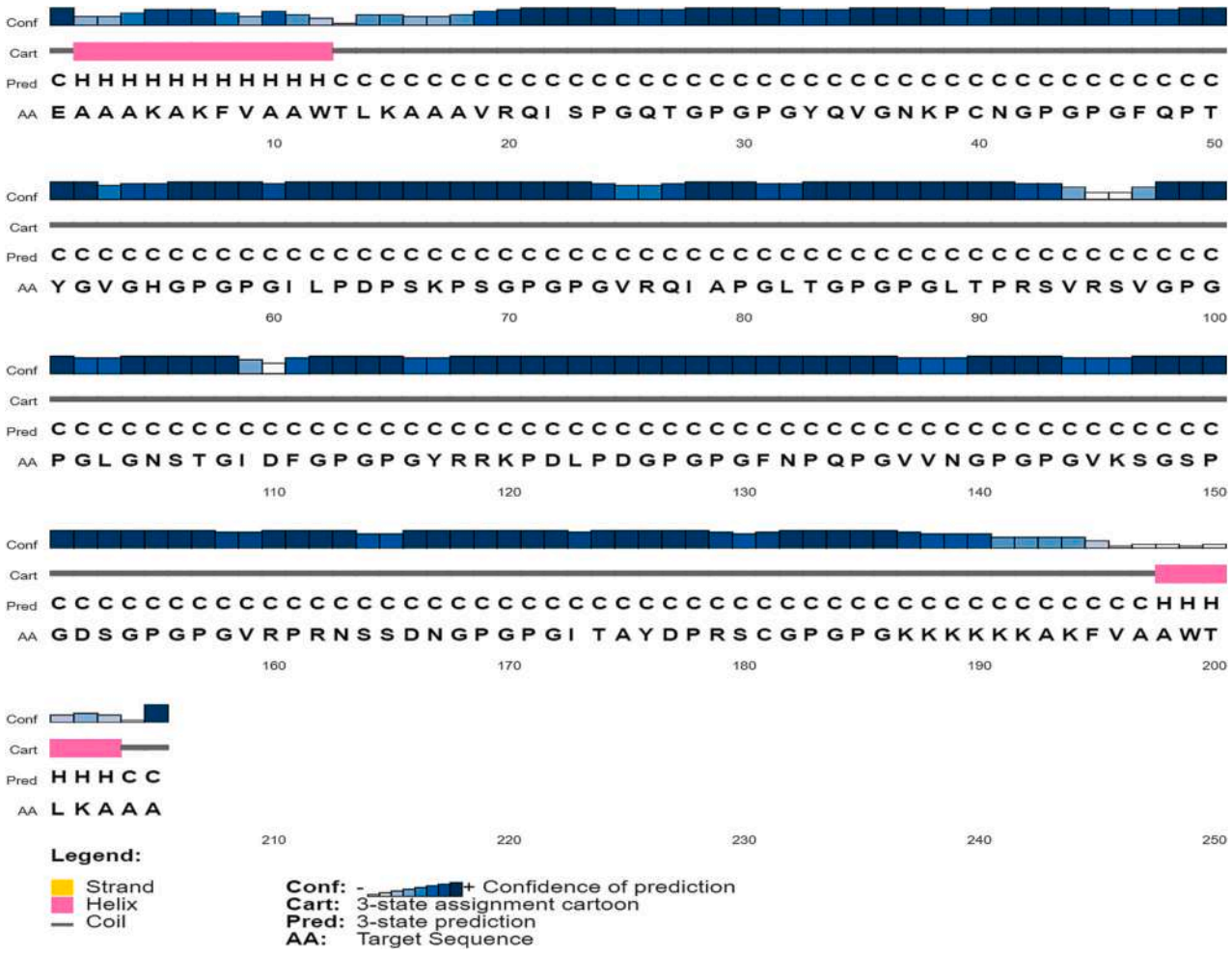


Fig. 6. The secondary structure analysis of the proposed multi-epitopic peptide vaccine using the PESIPRED webserver.

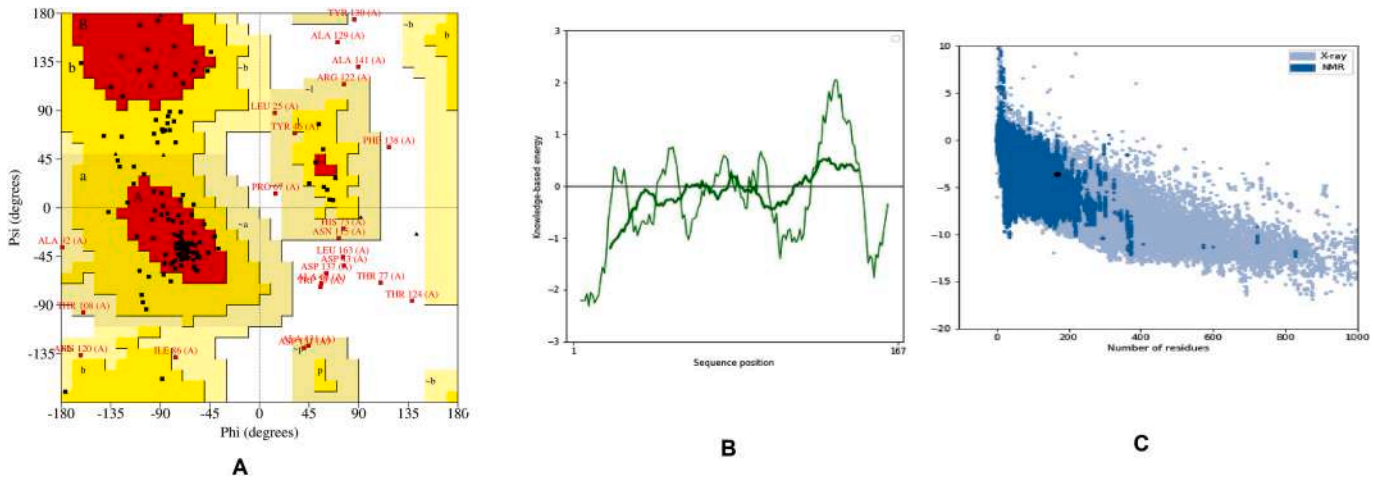


Fig. 7. Structural quality validation of our vaccine construct (A) Structure was validated through the Ramachandran plot for all amino acids residues of the vaccine construct, and it displays the result of the validation (B) Showing the energy plot of amino acids of vaccine candidate (C) The 'Z' score of vaccine candidate (black dot) within the 'Z' score range of experimentally validated peptide structure.

class I and MHC class II alleles in respect of twelve top-ranked antigenic epitopes from the S-protein comprising significant mutations of SARS-CoV-2 variants. The cumulative percentage (%) coverage for the MHC class I allele is 21.7 % of twelve nos. epitopes in the case of the entire world population. At the same time, only two numbers epitopes have 21 % coverage which was noted as the highest coverage in our study. One type of 9-mer epitope provides 18 % coverage, and, another single 9-mer epitope has shown 6 % coverage. Similarly, two 9-mer epitopes have 8 %, and three epitopes have 2 % coverage. In the case of MHC class II, we found four categories of epitope coverage on a whole world scale (four

Table 8
Distribution of amino acid residues showing in Ramachandran plot of the peptide-based vaccine construct.

Amino acid position	Residue number	Percentage (%)	Total
Favored region	94	58.6 %	157
Additional allowed region	44	27.1 %	
Generously allowed region	16	7.8 %	
Disallowed region	13	6.5 %	
End-residues			18
Glycine residues			12
Proline residues			9
Total residues			205

epitopes of >10 %, five epitopes of >5 %, and four epitopes in 2 % coverage) (Fig. 3).

The common (B cell and T cell) antigenic epitopes containing altered amino acids were characterized for developing the multi-epitope peptide-based vaccine. We considered the highest score bearing antigenic epitopes, and therefore, it can provide a robust immunological response against SARS-CoV-2 S-protein in the body of host cells. Our vaccine candidate contains twelve nos. epitopes with 205 nos. amino acids, two PADRE sequences, one CTxB adjuvant, and an essential peptide linker. Inside the N-terminal end of the vaccine, the adjuvant elements joined with the EAAAK peptide linkers. This peptide compound merged for the dynamic separation of the bifunctional fusion protein components [63]. Moreover, the EAAAK peptide linker was linked by the PADRE sequence. This sequence acts as an HTL epitope for raising the CTL immune response in varied classes of antigens.

The consequences of several advanced computational servers and tools recognized that the ultimate vaccine candidate was basic and highly stable in the normal physiological pH range in the human body. Likewise, the evaluated aliphatic index value stated that the vaccine candidate is also thermostable [64]. Hence, the GRAVY value in the positives core also shows its hydrophobic nature, without any viable interface by the water or non-solubility in the water medium [65].

Within our research, we have used the I-TASSER web server design to create a 3D structure of mutation containing multi-epitopic peptide vaccine candidate. The generated 3D structure of the vaccine was also carefully validated by PROCHECK and ProSA web servers. Here, the local energy plot model and 'Z' score were accurately accessed to recognize the correct folding of the protein chain. The ProSA web server predicted the 'Z' score as a negative value (−2.86). It suggests the standing signal for the protein structural configuration quality assessment [66]. Additionally, the analysis of the local quality model of residues is considered a consistent model.

The novel peptide-based vaccine candidate can be recognized by the human TLR4/MD2 complex (cell surface protein) as a target element and is safe with the required functions. The subsequent analyses of the molecular interaction of our multi-epitopic peptide vaccine construct intensely interact with the human TLR4/MD2 complex. It was also evident that TLR4/MD2 performed an imperative role in recognizing pathogens and stimulating the suitable innate immune response against infectious agents [67]. In our study, we implemented the molecular docking between the peptide vaccine construct and the TLR4/MD2

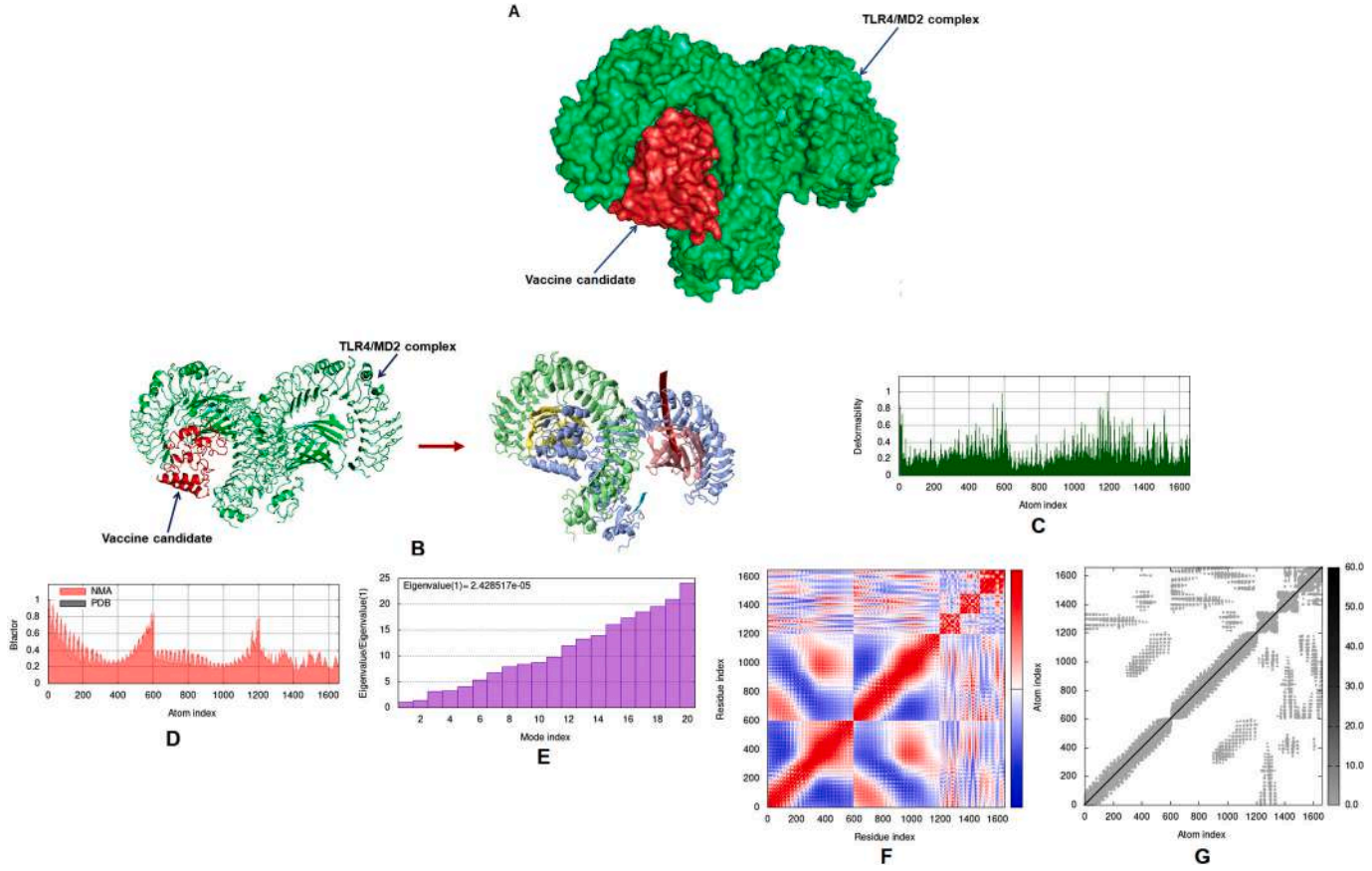


Fig. 8. The complex structure shows the molecular docked complex. The consequence of our vaccine construct with TLR4/MD2 complex analyzed through Normal Mode Analysis (NMA) (A) Docked complex (peptide vaccine candidate docked with TLR4/MD2 complex) illustrates surface model shown vaccine construct (red), (B) The mobility of the docking complex of vaccine and TLR4/MD2 specified with colored affine-arrow, (C) The deformability plot of the docked complex, (D) The calculated B-factor of NMA and PDB B-factor of the proposed vaccine candidate. (E) The eigenvalue showing for the molecular docking complex, (F) The covariance matrix map of atomic pair of interacting residues (amino acid) of the molecular docked complex, (G) The connection spring map of the elastic network model of the docked complex.

Table 9
The statistical analysis of vaccine construct and TLR4/MD2 docking complex. The negative docking (HADDOCK score) signifies a strong protein interaction expressed in arbitrary units (a.u.) and binding affinity score.

Vaccine construct and TLR4/MD2	
Docking score (HADDOCK score (a.u.))	-157.9 ± 8.6
Cluster size	9
RMSD from the overall lowest-energy structure	1.9 ± 0.7
Van der Waals energy (kcal mol^{-1})	-73.6 ± 1.5
Electrostatic energy (kcal mol^{-1})	-316.4 ± 32.7
Desolvation energy (kcal mol^{-1})	-13.5 ± 2.8
Restraints violation energy (kcal mol^{-1})	3.4 ± 2.6
Buried surface area ($Å^2$)	2864.3 ± 99.1
Z-Score	-2.8
Binding score of the docked protein-protein complex (PRODIGY binding score)	Gibbs free energy (ΔG) -12.7 kcal mol^{-1} and dissociation constant (Kd) 4.5E-09 M

Table 10
Interacting hydrogen bond (distance, A°) between the vaccine construct and TLR4/MD2 complex.

Vaccine construct (atom)	TLR4/MD2 (atom)	Distance (A°)
H:11305	O:15341	3.17
H:14316	H:12133	3.54
C:14163	H:20742	4.12
O:16837	H:21355	4.71
H:15312	C:12415	4.86
O:5241	H:12656	3.19
H:4872	H:22131	4.01

complex using the HwkDock and HADDOCK integrated web servers' platform. The results in the negative docking score confirmed that the vaccine candidate contains high binding interface features. It also supports that the potent vaccine construct can trigger the TLR4 activation and boosts the proper immune response against the S-protein of SARS-CoV-2 variants. The vaccine candidate and TLR4/MD2 and docked complex execute the binding free energy (BFE) of -96.67 kcal mol^{-1}. This BFE value identifies a consistent binding affinity between the immune sensing (TLR4/MD2) cellular receptor protein and the proposed vaccine candidate.

The NMA study was carried out to examine the relative deformability, B factor, and molecular motion of the docked complex of protein PDB. Our investigation has revealed the eigenvalue of the docked peptide complex sequence as 2.428517e-05. The noted eigenvalue can be elected for the greater flexibility of the ultimate multi-epitopic vaccine construct.

Consequently, the immune simulation profiling of the vaccine candidate established that the vaccine candidate could elicit cellular and humoral immune responses against the infection of SARS-CoV-2 and its significant variants. Likewise, the secondary immune response generated by the proposed vaccine was vividly greater than the primary immune response sets. As a final point, our proposed multi-epitopic vaccine construct exhibited satisfactory results in its structure and functional properties as assessed through different in silico analyses. The 'Z' score in negative value (-2.86) represents a reliable protein model for structural validation. The docked complex exhibits free binding energy of -96.67 kcal mol^{-1}, reflecting a consistent binding affinity leading to the immune sensing by interactions of molecules. Through the NMA analysis, the eigenvalue of TLR4/MD2 and vaccine peptide sequence (2.428517e-05) presented greater flexibility, lesser protein deformability, and good molecular motion. Subsequently, the immune simulation profiling using the computational program and tools of vaccine construct also established the assembly of abundant diversity, especially in the T cell population, and the upsurge of HTL cells within the immune system mammalian body. Thus, all the experimental properties based on computational tools and techniques specify that our multi-epitopic, peptide-based vaccine construct is an ideal next-generation vaccine and might be mutation-proof against the significant variant of SARS-CoV-2.

5. Limitations of the study

The study has some limitations. The study developed a mutation-proof, multi-epitopic peptide vaccine against SARS-CoV-2 containing a single construct. We use advanced computational and bioinformatics approaches such as artificial intelligence (AI) for mutation selection and machine learning (ML) based immune simulation to design the vaccine construct. Therefore, the vaccine construct remains very promising. However, the immunoinformatics-based vaccine construct needs successive in vivo and in vitro validation. Therefore, it needs to synthesize the vaccine construct and needs further validation. Finally, at molecular, cellular, and animal levels, successive validation will validate the proof of concepts. We urge researchers for future validation of our vaccine construct for the benefit of our society.

6. Conclusion

Mutations in the S-protein are reported as one of the primary causes of vaccine escape. The current work is to design a next-generation, multi-epitopic peptide-based potential vaccine candidate against the SARS-CoV-2 S-protein considering significant top-ranked mutations using selected through the AI model, top-ranked antigenic score selection method. At the same time, several computational and bioinformatics approaches are used for vaccine construct design. The peptide vaccine was designed from the S-protein of SARS-CoV-2 with many mutations bearing variants (Alpha, Beta, Delta, Gamma, and Omicron) in overlapping epitopic (9-mer) regions. Our immunoinformatics-based vaccine construct showed rational characterized properties regarding different physicochemical features, antigenicity, and non-allergenicity which were analyzed through bioinformatics and computational methods. So, different in silico analyses showed the designed vaccine construct is effective and safe for use against the infection of SARS-CoV-2 and its variants. However, the vaccine contract needs to synthesize and needs further validation through successive in vivo and in vitro experiments, which will provide proof of concepts.

Considerably this next-generation epitopic vaccine candidate might help in the formulation of a novel antiviral vaccine against SARS-CoV-2 and its significant variants. Apart from that, the structural quality validation of this vaccine and its binding affinity to the immune receptor were also monitored by the residues-oriented molecular docking and NMA of protein-protein molecules and these analyses also show good results.

The immune simulation informed us that the designed vaccine construct could provoke a robust and significant immune response. In this study, we used the immune simulation with advanced immunoinformatics and machine learning approaches. It also provides a broader range of protection against the infection of SARS-CoV-2, and the vaccine construct probably protect from all the forthcoming significant variants and subvariants. However, in vitro and in vivo validation of the vaccine candidate is undoubtedly required to confirm the effectiveness before administration in the human body.

Authors' contributions

Conceptualization was done by C·C; Writing—original draft preparation, methodology, investigation was performed by M.B.; Validation and formal analysis done by A.A and M.A; Writing—review and editing was done by C·C.; Visualization was done by K·D and S·S-L; Supervision and project administration was done by C.C; All authors have read and agreed to the published version of the manuscript.

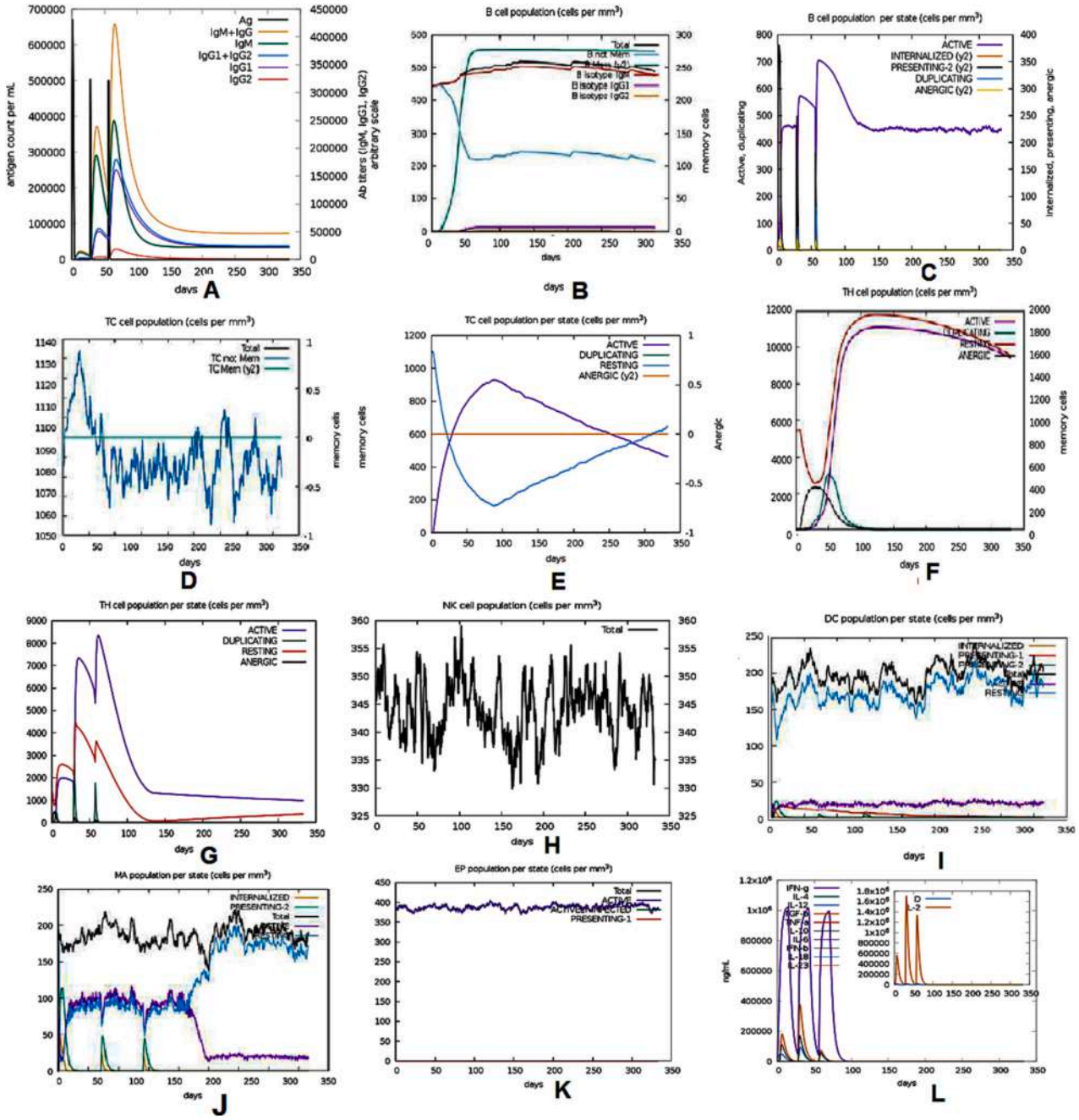


Fig. 9. The specific immune response of human host cells after injection of our designed peptide-based vaccine construct. The study illustrates diverse results from the in silico simulation and machine learning approaches. (A) The simulation graph shows the elevated level of immunoglobulins. Such elevation of immunoglobulins at different concentrations gradients of antigens, (B) It indicates the population of B cells secretory (IgG1, IgG2, and IgM) subsequently three injections of designed vaccine construct (C) The graph showing the analysis consequence of the population per entity-state (presenting on class-II, counts for the active state, internalized the Ag, duplicating antigenic by the diverse color variant). (D) The figure notifies the CTL population in the time (days) gaps after injection of the designed vaccine candidate. (E) The graph illustrates the CTL population in different situations, active and resting in the time (days) gaps after injection of the stated vaccine construct. (F) It depicts the total count (TC) of TH cell populations and the memory cells sub-divided into isotypes of IgG1, IgG2, and IgM after the injection of the developed vaccine. (G) It shows the population count per entity-state of HTL cell count in the active and resting states after the injection of our vaccine, (H) The behavior of the NK cell populations. (I) The graph shows the performance of the DC population in the resting and active states after injection of our designed vaccine. (J) The macrophages cell population after only the peptide vaccination, (K) It shows the concentration of cytokines and interleukins components with Simpson index [D], (L) The graph shows the total count of EC cells that is altered to virus-infected, active, and presenting on MHC class-I molecules.

Ethical approval and consent to participate

Not applicable.

Declaration of competing interest

The authors declare that they have no competing interests.

Data availability statement

All data included within the manuscript.

Acknowledgement

Authors are thankful to the Researchers Supporting Project number (RSP2023R491), King Saud University, Riyadh, Saudi Arabia.

References

[1] T. Kuiken, Escriou N. van der WS, J.C. Manuguerra, K. Stohr, J.S. Peiris, A. D. Osterhaus, et al., Newly Discovered Coronavirus as the Primary Cause of Severe Acute Respiratory Syndrome 362, 2003, pp. 263–270.
[2] H. Zhu, et al., The Novel Coronavirus Outbreak in Wuhan, China 5, 2020, pp. 1–3 (1).
[3] C. Chakraborty, et al., SARS-CoV-2 causing pneumonia-associated respiratory disorder (COVID-19): diagnostic and proposed therapeutic options, Eur. Rev. Med. Pharmacol. Sci. 24 (7) (2020) 4016–4026.
[4] C. Chakraborty, et al., The 2019 novel coronavirus disease (COVID-19) pandemic: a zoonotic prospective, Asian Pac. J. Trop. Med. 13 (6) (2020) 242.
[5] K. Dhama, et al., COVID-19, An Emerging Coronavirus Infection: Advances and Prospects in Designing and Developing Vaccines, Immunotherapeutics, and Therapeutics 16, 2020, pp. 1232–1238 (6).
[6] L. Wang, et al., Review of the 2019 Novel Coronavirus (SARS-CoV-2) Based on Current Evidence 55, 2020, p. 105948 (6).
[7] R.B. Viana, C.A.B. de Lira, Exergames as coping strategies for anxiety disorders during the COVID-19 quarantine period, Games Health J. 9 (3) (2020) 147–149.
[8] H. Tegally, et al., Emergence and Rapid Spread of a New Severe Acute Respiratory Syndrome-Related Coronavirus 2 (SARS-CoV-2) Lineage with Multiple Spike Mutations in South Africa, 2020.
[9] A. Bhuyan, Covid-19: India Sees New Spike in Cases Despite Vaccine Rollout 372, British Medical Journal Publishing Group, 2021, p. n854.
[10] M. Bhattacharya, et al., Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): immunoinformatics approach, J. Med. Virol. 92 (6) (2020) 618–631.
[11] M. Bhattacharya, et al., A next-generation vaccine candidate using alternative epitopes to protect against Wuhan and all significant mutant variants of SARS-CoV-2: an immunoinformatics approach, Aging Dis. 12 (8) (2021) 2173–2195.
[12] R. Dong, et al., Contriving multi-epitope subunit of vaccine for COVID-19: Immunoinformatics approaches, Front. Immunol. 11 (2020) 1784.
[13] P. Kalita, et al., Design of a peptide-based subunit vaccine against novel coronavirus SARS-CoV-2, Microb. Pathog. 145 (2020), 104236.
[14] Aasim, et al., Identification of vaccine candidate against Omicron variant of SARS-CoV-2 using immunoinformatic approaches, Silico Pharmacol. 10 (1) (2022) 12.
[15] C. Chakraborty, et al., Lessons learned from cutting-edge immunoinformatics on next-generation COVID-19 vaccine research, Int. J. Pept. Res. Ther. 27 (4) (2021) 2303–2311.
[16] C. Chakraborty, M. Bhattacharya, K. Dhama, SARS-CoV-2 vaccines, vaccine development technologies, and significant efforts in vaccine development during the pandemic: the lessons learned might help to fight against the next pandemic, Vaccines 11 (3) (2023) 682.
[17] C. Chakraborty, et al., Immediate need for next-generation and mutation-proof vaccine to protect against current emerging omicron sublineages and future SARS-CoV-2 variants: an urgent call for researchers and vaccine companies - correspondence, Int. J. Surg. 106 (2022), 106903.
[18] Y. Huang, et al., Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19, Acta Pharmacol. Sin. 41 (9) (2020) 1141–1149.
[19] L. Du, et al., The spike protein of SARS-CoV–a target for vaccine and therapeutic development, Nat. Rev. Microbiol. 7 (3) (2009) 226–236.
[20] W.T. Harvey, et al., SARS-CoV-2 variants, spike mutations and immune escape, Nat. Rev. Microbiol. 19 (7) (2021) 409–424.
[21] M. Bhattacharya, et al., Omicron variant (B.1.1.529) of SARS-CoV-2: understanding mutations in the genome, S-glycoprotein, and antibody-binding regions, Geroscience 44 (2) (2022) 619–637.
[22] C. Yi, et al., Jigsaw puzzle of SARS-CoV-2 RBD evolution and immune escape, Cell. Mol. Immunol. 19 (7) (2022) 848–851.
[23] J. Chen, et al., Revealing the threat of emerging SARS-CoV-2 mutations to antibody therapies, J. Mol. Biol. 433 (18) (2021), 167155.
[24] R. Wang, et al., Emerging vaccine-breakthrough SARS-CoV-2 variants, ACS Infect Dis. 8 (3) (2022) 546–556.
[25] J. Chen, G.W. Wei, Mathematical artificial intelligence design of mutation-proof COVID-19 monoclonal antibodies, Commun. Inf. Syst. 22 (3) (2022) 339–361.
[26] Y. El-Manzalawy, D. Dobbs, V. Honavar, Predicting linear B-cell epitopes using string kernels, J. Mol. Recognit. 21 (4) (2008) 243–255.
[27] R. Vita, et al., The immune epitope database (IEDB): 2018 update, Nucleic Acids Res. 47 (D1) (2019) D339–D343.
[28] P. Wang, et al., A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach, PLoS Comput. Biol. 4 (4) (2008), e1000048.
[29] A. Rana, Y. Akhter, A multi-subunit based, thermodynamically stable model vaccine using combined immunoinformatics and protein structure based approach, Immunobiology 221 (4) (2016) 544–557.
[30] M. Kavoosi, et al., Strategy for selecting and characterizing linker peptides for CBM9-tagged fusion proteins expressed in Escherichia coli, Biotechnol. Bioeng. 98 (3) (2007) 599–610.
[31] J. Olejnik, A.J. Hume, E. Muhlberger, Toll-like receptor 4 in acute viral infection: too much of a good thing, PLoS Pathog. 14 (12) (2018), e1007390.
[32] M. Sadraeian, et al., Cloning and Expression of CtxB-StxB in Esherichia Coli: A Challenge for Improvement of Immune Response Against StxB, 2011.
[33] I.A. Doytchinova, D.R. Flower, VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines, BMC Bioinformatics 8 (2007) 4.
[34] I. Dimitrov, et al., AllerTOP v.2–a server for in silico prediction of allergens, J. Mol. Model. 20 (6) (2014) 2278.
[35] M.R. Wilkins, et al., Protein identification and analysis tools in the ExPASy server, Methods Mol. Biol. 112 (1999) 531–552.
[36] C. Geourjon, G. Deleage, SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, Comput. Appl. Biosci. 11 (6) (1995) 681–684.
[37] L.J. McGuffin, K. Bryson, D.T. Jones, The PSIPRED protein structure prediction server, Bioinformatics 16 (4) (2000) 404–405.
[38] A. Roy, A. Kucukural, Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction, Nat. Protoc. 5 (4) (2010) 725–738.
[39] J. Ko, et al., GalaxyWEB server for protein structure prediction and refinement, Nucleic Acids Res. 40 (Web Server issue) (2012) W294–W297.
[40] M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, Nucleic Acids Res. 35 (Web Server issue) (2007) W407–W410.
[41] R.A. Laskowski, et al., PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (2) (1993) 283–291.
[42] Y. Yan, et al., The HDOCK server for integrated protein-protein docking, Nat. Protoc. 15 (5) (2020) 1829–1852.
[43] G. Weng, et al., HawkDock: a web server to predict and analyze the protein-protein complex based on computational docking and MM/GBSA, Nucleic Acids Res. 47 (W1) (2019) W322–W330.
[44] L. Schrodinger, The PyMOL molecular graphics system, Version 1 (5) (2010) 0.
[45] L.C. Xue, et al., PRODIGY: a web server for predicting the binding affinity of protein-protein complexes, Bioinformatics 32 (23) (2016) 3676–3678.
[46] J.R. Lopez-Blanco, et al., iMODS: internal coordinates normal mode analysis server, Nucleic Acids Res. 42 (Web Server issue) (2014) W271–W276.
[47] N. Rapin, et al., Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system, PLoS One 5 (4) (2010), e9862.
[48] F. Castiglione, et al., How the interval between prime and boost injection affects the immune response in a computational model of the immune system, Comput. Math. Methods Med. 2012 (2012), 842329.
[49] S. Babicki, et al., Heatmapper: web-enabled heat mapping for all, Nucleic Acids Res. 44 (W1) (2016) W147–W153.
[50] R.P. Saha, et al., Repurposing drugs, ongoing vaccine, and new therapeutic development initiatives against COVID-19, Front. Pharmacol. 11 (2020) 1258.
[51] C. Chakraborty, M. Bhattacharya, A.R. Sharma, Emerging mutations in the SARS-CoV-2 variants and their role in antibody escape to small molecule-based therapeutic resistance, Curr. Opin. Pharmacol. 62 (2022) 64–73.
[52] C. Chakraborty, et al., A comprehensive analysis of the mutational landscape of the newly emerging omicron (B.1.1.529) variant and comparison of mutations with VOCs and VOIs, Geroscience 44 (5) (2022) 2393–2425.
[53] M. Bhattacharya, et al., Delta variant (B.1.617.2) of SARS-CoV-2: current understanding of infection, transmission, immune escape, and mutational landscape, Folia Microbiol. (Praha) 68 (1) (2023) 17–28.
[54] C. Chakraborty, et al., Omicron (B.1.1.529) - a new heavily mutated variant: mapped location and probable properties of its mutations with an emphasis on S-glycoprotein, Int. J. Biol. Macromol. 219 (2022) 980–997.
[55] C. Chakraborty, et al., A detailed overview of immune escape, antibody escape, partial vaccine escape of SARS-CoV-2 and their emerging variants with escape mutations, Front. Immunol. 13 (2022) 801522.
[56] C. Chakraborty, et al., Evolution, mode of transmission, and mutational landscape of newly emerging SARS-CoV-2 variants, MBio 12 (4) (2021), e0114021.
[57] A.V. Kudriavtsev, et al., Immune escape associated with RBD omicron mutations and SARS-CoV-2 evolution dynamics, Viruses 14 (8) (2022).
[58] J. Zhang, et al., A mosaic-type trimeric RBD-based COVID-19 vaccine candidate induces potent neutralization against omicron and other SARS-CoV-2 variants, elife (2022) 11.
[59] R.H. Meloen, et al., Synthetic peptide vaccines: unexpected fulfillment of discarded hope? Biologicals 29 (3–4) (2001) 233–236.
[60] M. Bhattacharya, et al., Immunoinformatics approach to understand molecular interaction between multi-epitopic regions of SARS-CoV-2 spike-protein with TLR4/MD-2 complex, Infect. Genet. Evol. 85 (2020), 104587.

[61] H.H. Bui, et al., Predicting population coverage of T-cell epitope-based diagnostics and vaccines, BMC Bioinformatics 7 (2006) 153.
[62] P. Oyarzun, B. Kobe, Computer-aided design of T-cell epitope-based vaccines: addressing population coverage, Int. J. Immunogenet. 42 (5) (2015) 313–321.
[63] H.J. Kim, et al., Intranasal vaccination with peptides and cholera toxin subunit B as adjuvant to enhance mucosal and systemic immunity to respiratory syncytial virus, Arch. Pharm. Res. 30 (3) (2007) 366–371.
[64] A. Ikai, Thermostability and aliphatic index of globular proteins, J. Biochem. 88 (6) (1980) 1895–1898.
[65] M. Ali, et al., Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection, Sci. Rep. 7 (1) (2017) 9232.
[66] M. Bhattacharya, et al., TN strain proteome mediated therapeutic target mapping and multi-epitopic peptide-based vaccine development for mycobacterium leprae, Infect. Genet. Evol. 99 (2022), 105245.
[67] M. Bhattacharya, et al., Designing, characterization, and immune stimulation of a novel multi-epitopic peptide-based potential vaccine candidate against monkeypox virus through screening its whole genome encoded proteins: an immunoinformatics approach, Travel Med. Infect. Dis. 50 (2022), 102481.
[68] K. Mou, et al., Emerging mutations in Nsp1 of SARS-CoV-2 and their effect on the structural stability, Pathogens 10 (10) (2021) 1285.
[69] M. Gupta, et al., CryoEM and AI Reveal a Structure of SARS-CoV-2 Nsp2, a Multifunctional Protein Involved in Key Host Processes, bioRxiv, 2021.
[70] H. Banoun, Evolution of SARS-CoV-2: review of mutations, role of the host immune system, Nephron 145 (4) (2021) 392–403.
[71] Á. Nagy, S. Pongor, B. Győrffy, Different mutations in SARS-CoV-2 associate with severe and mild outcome, Int. J. Antimicrob. Agents 57 (2) (2021), 106272.
[72] S. Rehman, et al., Identification of novel mutations in SARS-COV-2 isolates from Turkey, Arch. Virol. 165 (12) (2020) 2937–2944.
[73] N. Yashvardhini, A. Kumar, D.K. Jha, Analysis of SARS-CoV-2 mutations in the main viral protease (NSP5) and its implications on the vaccine designing strategies, Vacunas 23 (2021) S1–S13.
[74] H. Akkiz, Implications of the novel mutations in the SARS-CoV-2 genome for transmission, disease severity, and the vaccine development, Front. Med. (2021) 8.
[75] G.K. Azad, Identification of novel mutations in the methyltransferase complex (Nsp10-Nsp16) of SARS-CoV-2, Biochem. Biophys. Rep. 24 (2020), 100833.
[76] S. Vilar, D.G. Isom, One year of SARS-CoV-2: how much has the virus changed? Biology 10 (2) (2021) 91.
[77] C. Cao, et al., Molecular epidemiology analysis of early variants of SARS-CoV-2 reveals the potential impact of mutations P504L and Y541C (NSP13) in the clinical COVID-19 outcomes, Infect. Genet. Evol. 92 (2021), 104831.
[78] L. Guruprasad, Human SARS CoV-2 spike protein mutations, Proteins: Struct., Funct., Bioinf. 89 (5) (2021) 569–576.
[79] A.K. Padhi, T. Tripathi, Can SARS-CoV-2 accumulate mutations in the S-protein to increase pathogenicity? ACS Pharmacol. Transl. Sci. 3 (5) (2020) 1023–1026.
[80] J. Chen, et al., Mutations strengthened SARS-CoV-2 infectivity, J. Mol. Biol. 432 (19) (2020) 5212–5226.
[81] P. Majumdar, S. Niyogi, ORF3a mutation associated with higher mortality rate in SARS-CoV-2 infection, Epidemiol. Infect. (2020) 148.
[82] L. van Dorp, et al., No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2, Nat. Commun. 11 (1) (2020) 1–8.
[83] M.S. Rahman, et al., Mutational insights into the envelope protein of SARS-CoV-2, Gene Rep. 22 (2021), 100997.
[84] Jungreis, I., R. Sealfon, and M. Kellis, SARS-CoV-2 Gene Content and COVID-19 Mutation Impact by Comparing 44 Sarbecovirus Genomes.
[85] S.S. Hassan, P.P. Choudhury, B. Roy, Rare mutations in the accessory proteins orf6, orf7b, and orf10 of the sars-cov-2 genomes, Meta Gene 28 (2021), 100873.

Travel Medicine and Infectious Disease 50 (2022) 102481

Contents lists available at ScienceDirect

Travel Medicine and Infectious Disease

journal homepage: www.elsevier.com/locate/tmaid

Designing, characterization, and immune stimulation of a novel multi-epitopic peptide-based potential vaccine candidate against monkeypox virus through screening its whole genome encoded proteins: An immunoinformatics approach

Manojit Bhattacharya [a], Srijan Chatterjee [b], Sagnik Nag [c], Kuldeep Dhama [d], Chiranjib Chakraborty [b,*]

[a] Department of Zoology, Fakir Mohan University, Vyasa Vihar, Balasore, 756020, Odisha, India
[b] Department of Biotechnology, School of Life Science and Biotechnology, Adamas University, Kolkata, West Bengal, 700126, India
[c] Department of Biotechnology, School of Biosciences & Technology, Vellore Institute of Technology (VIT), Tamil Nadu, 632014, India
[d] Division of Pathology, ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly, 243122, Uttar Pradesh, India

ARTICLE INFO

Keywords:
Monkeypox virus (MPXV)
Multi-epitope
Peptide-based vaccine
Immunoinformatics
Molecular docking

ABSTRACT

Background: The current monkeypox virus (MPXV) spread in the non-epidemic regions raises global concern. Presently, the smallpox vaccine is used against monkeypox with several difficulties. Conversely, no next-generation vaccine is available against MPXV. Here, we proposed a novel multi-epitopic peptide-based in-silico potential vaccine candidate against the monkeypox virus.

Methods: The multi-epitopic potential vaccine construct was developed from antigen screening through whole genome-encoded 176 proteins of MPXV. Afterward, ten common B and T cell epitopes (9-mer) having the highest antigenicity and high population coverage were chosen, and a vaccine construct was developed using peptide linkers. The vaccine was characterized through bioinformatics to understand antigenicity, non-allergenicity, physicochemical properties, and binding affinity to immune receptors (TLR4/MD2-complex). Finally, the immune system simulation of the vaccine was performed through immunoinformatics and machine learning approaches.

Results: The highest antigenic epitopes were used to design the vaccine. The docked complex of the vaccine and TLR4/MD2 had shown significant free binding energy (−98.37 kcal/mol) with a definite binding affinity. Likewise, the eigenvalue (2.428517e-05) from NMA analysis of this docked complex reflects greater flexibility, adequate molecular motion, and reduced protein deformability, and it can provoke a robust immune response.

Conclusions: The designed vaccine has shown the required effectiveness against MPXV without any side effects, a significant milestone against the neglected disease.

1. Introduction

During the present post-COVID-19 pandemic state, the cases of monkeypox virus (MPXV) infection started appearing in multiple non-endemic countries beyond Africa in May 2022, as reported by the WHO. Afterward, the cases of monkeypox increased to a higher rate with considerable levels. With such a continuous rapid spread, the current monkeypox outbreak has become the largest in non-African countries. Though travel links from endemic to non-endemic countries have not been found to be associated with the current MPX outbreak, the virus infection might have spread initially through a traveler—a patient with a travel history from Nigeria to the UK and back [1]. The MPXV is a viral-zoonotic disease showing similar features as observed in the smallpox virus infection, even though it is physiologically less acute [2]. The MPXV belongs to the *Orthopoxvirus* genus under the family of Poxviridae. It is also noted that the Poxviruses have a strong propensity to emerge outer of their consistent ecological range by spreading to a naive community [3]. The MPXV obtained global attention during its first appearance in the Western Hemisphere in 2003, which caused a high number of cases in the Midwest of the USA [4].

* Corresponding author.
E-mail address: drchiranjib@yahoo.com (C. Chakraborty).

https://doi.org/10.1016/j.tmaid.2022.102481
Received 21 July 2022; Received in revised form 11 September 2022; Accepted 13 October 2022
Available online 17 October 2022

From January 1 to September 11, 2022, 57,527 confirmed cases and 18 deaths has been reported from 103 countries/territories [5]. The Central African Republic, the Democratic Republic of the Congo, Gabon, and Ghana are the countries where the monkeypox was endemic. Gradually the MPXV infection scenario is rising, and additional cases of MPXV infection are also detected as the disease outbreak progresses [6].

During the current outbreak in African countries, the genome sequence of the MPXV was first published in NCBI online database, and several sequences from other infected countries have also followed. Subsequently, the discovered MPXV strain was not so linked to the viral strain predominant in West Africa, which causes mild symptoms with a lesser death rate. MPXV genome contains a linear dsDNA (double-strand DNA), where the ITR (inverted terminal repeats) are explicitly composed of the hairpin loops like structure, tandem repeats sequence, and the ORFs (open reading frames) at end parts. The MPXV resides in the cytoplasmic part of the virally infected cells and encodes the proteins required for DNA replication, virion assembly, RNA expression, and viral escape [7].

Currently, no specific approved drugs or vaccines are available to treat the human MPXV infection; only the Dryvax (smallpox vaccine) has been used against MPXV and smallpox virus. During the vaccination, the skin is needed to be punctured many times by a bifurcated needle holding a small quantity of vaccine element. The technique is quite painful. On the other hand, stockpiling vaccines for longer days might also cause to reduce the vaccine efficiency [8]. However, various adverse effects were noted among the vaccinated individuals [9,10]. Therefore, there is an urgent need for a next-generation vaccine for monkeypox. In this direction, we have gone through the entire genome-encoded proteins (176 protein chains) of MPXV to develop a next-generation vaccine. We have screened all the common antigenic Bcell and T cell epitopes. The common epitopes with high antigenic scores were considered for designing a next-generation peptide vaccine against the MPXV.

The principal idea for all vaccination types is the ability of a vaccine to create an effective immune response faster than the infecting virus itself. However, classical vaccines are developed on several biochemical experiments and elicit antibodies in vaccinated individuals. These vaccines are also expensive, time-taking, and allergic, and the process requires *in vitro* culture of harmful microbes and faces severe safety concerns. Conversely, peptide-based epitopic vaccine production is extremely safe and less cost-effective than the traditional vaccine candidate [11,12]. Technical requisites for effective and safe vaccines are too crucial.

In this study, intensive research was performed to identify the common (B and T cell) 9-mer antigenic epitopes from the whole genome encoded proteins of MPXV. Afterward, the high-rank epitopes with the utmost antigenic score were selected from these epitopes for vaccine construct development. Furthermore, multiple *in silico* tools characterized and validated the potential vaccine candidate, and studies were performed to understand the molecular interaction with the human TLR4/MD2 complex. Finally, we have tested the immune simulation profiling of this novel peptide-based vaccine in the mammalian immune system using a machine learning approach. This novel vaccine candidate revealed protection against MPXV without any side effects and has the potential for usage in the coming days.

2. Methods

We followed the steps of muti-epitopic peptide-based vaccine development. However, every step of this implemented methodology is important and promising for designing an effective peptide-based multi-epitopic vaccine against MPXV.

2.1. The study of the whole genome, genome-encoded proteins, and retrieval of all of the genome encoded-protein sequences from the monkeypox genome

We studied the whole genome of the dsDNA of the MPXV and its all genome-encoded proteins. We found there are about 176 gene-coding proteins in the monkeypox virus. Simultaneously, all the genome-encoded proteins of the virus were studied, and the FASTA sequences of all its genome-coded protein sequences were retrieved from the NCBI. A total number of 176 gene-coding peptide sequences were collected for epitope screening and further analysis.

2.2. Identification of the 9-mer B cell epitopes

The IEDB (Bepipred 2.0) recommended 2020.09 methods were employed to predict linear B cell (9-mer) epitopes based on sequence characteristics of the antigen using amino acid scales and Hidden Markov models (HMMs) [13]. All gene-coded proteins from the genome were submitted in the mentioned online tool for identifying the 9-mer B cell epitopes.

2.3. Identification of the 9-mer T cell epitopes

The Immune Epitope Database (IEDB) v2.24 webserver was used to predict the 9-mer T cell epitope(s) in protein sequence(s) containing MHC class I epitopes. The server is based on sequence alignment using artificial neural networks to predict T cell epitopes [14]. We have selected all the suitable parameters for identifying 9-mer T cell epitopes and each peptide sequence derived from the MPXV genome are shown in Table 1.

2.4. Identification of the common 9-mer epitopes from B cell and T cell epitopes

From the outputs of Bepipred 2.0 and IEDBv2.24 server for the identified B cell and T cell epitopes, we further characterized common 9-mer epitopes, both antigenic and non-antigenic. Afterward, we selected ten common antigenic epitopes with the highest antigenic score (VexiJen score).

2.5. Determination of epitopes population coverage

We have identified the considerable number of HLA binding antigenic epitope peptides, which provide the population coverage information by its interaction patterns. The IEDB population coverage tool is used to estimate the population coverage of the potent epitopes of MPXV that can interact with the HLA alleles (MHC-I and MHC-II) [15].

2.6. Construction of multi-epitopic peptide-based potential vaccine candidate

The complete vaccine candidate was constructed by joining common B cell and T cell-derived antigenic epitopes. Peptide linkers such as GPGPG and AAY were taken to fuse these 9-mer antigenic epitopes. As noted, the peptide linker performed a critical role in refining the epitope separation and helped epitope presentation towards MHC class I and II receptors and immunological processing purposes [16]. Additionally, for immunity boosting, the adjuvant (CTxB) was added to the N-terminal part of the developed chimeric peptide sequences by incorporating the EAAAK linker into the MPXV vaccine construct. The CTxB is considered a nontoxic piece of 124 amino acids sequence of cholera toxin, which can develop the human body's humoral and cell-mediated immune responses [17]. Lastly, 13 amino acids long chain, an extra PADRE sequence (AKFVAAWTLKAAA) was added at the construct's C-terminal end to expand the host immunogenicity [18].

Table 1
Common B cell & T cell (9-mer) epitope of MPXV with the highest antigenic score (threshold for VaxiJen score = 0.4).

SL. No.	Gene name	Protein name	Antigenic score of gene	Common B cell & T cell epitope	Start (aa)	End (aa)	Antigenic score of epitope
1.	MV-6	ankyrin-like protein	0.3705	FDLSVKCEN	127	135	2.2910
2.	MV-12	IL-1 receptor antagonist	0.4645	RFNDMTITD	76	84	1.6117
3.	MV-13	complement binding	0.6228	LDIGGVDFG	95	103	2.2501
4.	MV-20	serine protease inhibitor-like protein	0.5702	NLRKRDLGP	74	82	2.6689
5.	MV-28	major membrane protein	0.5696	VSHINYTSW	63	71	1.6572
6.	MV-42	poly polymerase large subunit	0.4419	IRYGDIDIL	198	206	2.2883
7.	MV-147	hypothetical protein	1.0183	ITIDSKIGN	6	14	1.7127
8.	MV-150	toll/IL1-receptor [TIR]-like protein	0.4593	IRHRNTISG	99	107	1.6052
9.	MV-158	ankyrin-like protein	0.4212	ANIDSVDFN	238	246	1.5435
10.	MV-170	ankyrin-like protein	0.4222	ADISLKTDD	539	547	1.5014

2.7. Prediction of antigenicity and allergenicity of the potential vaccine candidate

We have predicted the antigenicity of the vaccine candidate through the VaxiJen webserver. This server followed a new alignment-independent method specifically for antigen prediction, based on Auto Cross Covariance (ACC) transformation of targeted peptide sequences [19]. The VaxiJen server is a consistent and reliable server for assuming the protective antigens using antigenicity of peptides with a threshold value (0.4) in default mode for viruses and other organisms.

At the same time, the AllerTop 2.0 server has been applied to understand the vaccine allergenicity. This web server-based ACC method transforms the amino acids sequence as equal-length vectors [20].

2.8. Primary and secondary structure analysis

Using the ExpasyProtParam tool, we have predicted and assessed the primary structure of the proposed potential vaccine candidate [21]. We have also evaluated the set of chemical and physical features of the input peptide sequence, numbers of amino acids, isoelectric point (pI), molecular weight, grand average of hydropathicity (GRAVY), net charges, aliphatic index, instability index, total atoms number and assessed half-life [21]. Subsequently, to analyze the secondary structure of the vaccine construct, we used the PSIPRED 4.0 and SOPMA (Secondary Structure Prediction Method) tools [22,23]. We also accessed the significant properties of our vaccine construct (e.g., globular regions, transmembrane helices, random coil, bend regions, and coiled-coil regions).

2.9. Three-dimensional (3D) modeling of the potential vaccine candidate and model refinement

The protein architecture in three-dimensional (3D) shapes offer important information for understanding molecular functions at the atomic level. The 3D structure of the developed vaccine was predicted by the I-TASSER web server [24]. This server is an integrated platform for the prediction of automated protein function and structure based on the sequence-to-structure-to-function paradigm. It makes a good quality model containing accurate coordinate formation compared to other known proteins. The GalaxyWEB server was employed to refine the predicted 3D model [25].

2.10. Validation of peptide-based potential vaccine structure

Accurate validation of the predicted protein structure 3D model is also important. We applied two significant web servers to validate the peptide-based vaccine structure of MPXV (PROCHECK and ProSA). The PROCHECK server examines the stereochemical quality of any protein structure and creates several PostScript plots which analyze its overall and residue-by-residue geometry by Ramachandran plot [26]. In contrast, the ProSA server was employed to confirm the protein tertiary structure. The overall quality scoring was calculated by the 'Z' plot and local quality model [27].

2.11. Molecular docking of MPXV vaccine construct with the human TLR4/MD2 complex

The interactions of antigenic peptide-based vaccines with the targeted immune cell proteins are critical for generating an appropriate immune response. The human TLR4/MD2 complex has a significant role in eliciting immune responses against microbial infection. For the calculation of protein-protein binding interactions of the MPXV vaccine constructs with the human TLR4/MD2 complex (PDB ID: 4G8A), we used the PyMoL tool [28] to eliminate the undesired elements from the coordinates of PDB.

For molecular docking, we employed the HawkDock server [29,30]. The Hawkdock server computes its binding interactions pattern along with the basis of calculating mechanisms of MM/GBSA free energy by the top-ranked model. Hence, these molecular species were selected as per the primary structure for commencing the molecular dynamic simulation.

2.12. Normal mode analysis of peptide-based vaccine

For normal mode analysis (NMA), the iMODS server has been used to demonstrate the cooperative motion of peptide-based vaccine through normal mode within the inner coordinates [31]. We have applied the essential dynamics simulation (EDM), an added program of the iMODS, to understand the macromolecular mobility and stability of the docked complex. Here, the evaluated parameters of the docked complex are the probable motions in the specialized terms of B-factors, deformability, covariance, elastic model, and eigenvalues.

2.13. Immune simulation profiling of our peptide-based vaccine using machine learning approaches

The C-ImmSim server has been used to portray the immune response profile of the designed vaccine construct [32], which defines cellular and humoral response within the mammalian immune system in the presence of antigenic or immunogenic components of bacteria, virus, etc., within the sub-cellular level (mesoscopic scale). The specific server evaluated the specific antigen-mediated immune response using the "position-specific scoring matrix" and the machine learning approaches. Other important parameters like the simulation volume, random seed, and simulation step were adopted at 30, 12345, and 1000 correspondingly. In addition, we have applied double injection intervals of four-week to execute the proposed Immune simulation. However, all other parameters retained in the default state [33].

The typical methodologies applied to accomplish our work are also listed in a flowchart (Fig. 1).

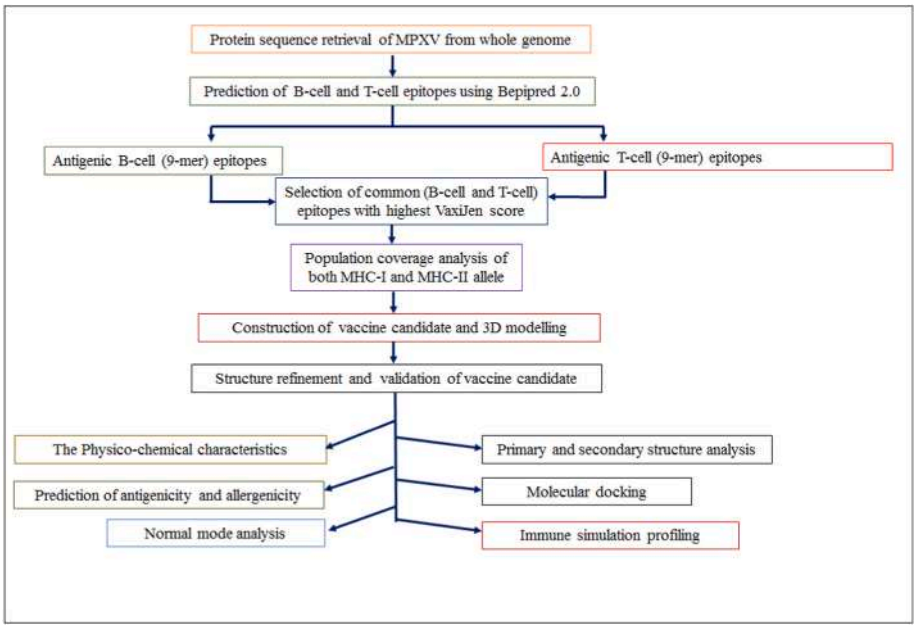


Fig. 1. Adopted methodologies (*in silico* techniques, proteome, immunoinformatics, and machine learning approaches) for designing next-generation multi-epitopic potent vaccine candidate screened from whole-genome encoded proteins of MPXV.

3. Results

3.1. Identification and prediction of the 9-mer common B cell and T cell epitopes from monkeypox whole genome encoded protein

In an antigen, the epitopic part is considered to be the key stimulator of the immune system. T cells or B cells, the shared peptide-based epitopic vaccine construct, may achieve a dual purpose of boosting the host immune system. Therefore, we designated 75 common 9-mer B cell and T cell epitopes from MPXV genome-encoded proteins. In contrast, the ten significant 9-mer common B cell and T cell epitopes having maximum antigenicity with highest VaxiJen score were selected for the final vaccine development (Table 1).

3.2. Population coverage analysis

We have interpreted the population coverage based on the population coverage (epitope-based) analysis of the parallel alleles of MHC-I and MHC-II. This web server clearly defined the population coverage for MHC-I and MHC-II epitopes.

The population coverage in respect to MPXV whole genome encoded protein-derived filtrated epitopes for both the MHC-I and MHC-II alleles is shown in Fig. 2.

3.3. Prediction of antigenicity and allergenicity of MPXV vaccine candidate

The exploration of vaccine construct antigenicity by the VaxiJen v2.0 tool exhibited that it is a potent antigen having an antigenicity score of 1.1013 (Table 2). This score also proposes that the designed peptide construct of multi-epitopes, along with the PADRE sequence and adjuvant, might enhance the immune response in the host body. The diagram shows the constructed vaccine in Fig. 3A.

The AllerTOP 2.0 webserver was applied for the allergenicity calculation of the potential vaccine candidate for safe future use.

This server calculated the significant physicochemical properties of the MPXV vaccine candidate (Table 2), which was likely to be basic. The developed vaccine's theoretical Isoelectric point (pI) was 6.33, with molecular weight (MW) of 18108.33 Da. The Grand average of hydropathicity (GRAVY) was calculated at 0.069, the positive GRAVY score specifying the potential vaccine candidate protein might be hydrophobic. The hydrophobic nature of the peptide chain should help in the smooth purification and better design of the vaccine.

The Aliphatic index (thermo-stability) of the designed MPXV vaccine was found to be 88.08. Additionally, the half-life of the potential vaccine candidate was calculated as 1 h in mammalian reticulocytes (*in vitro*), 30 min in yeast (*in vivo*), and >10 h in *E. coli.* At the same time, the instability index was calculated as 18.96. It specifies the stable MPXV vaccine construct of protein.

3.4. Primary and secondary structure analysis of MPXV vaccine construct

The SOPMA study projected the secondary structural architecture values such as the alpha helix (47.90%), extended strand (11.98%), beta-turn (0.60%), and random coil (39.52%) in the case of our vaccine (Fig. 4A and B). This server considered the window width value as 17. The similarity threshold value was fixed as 8 and measured the four numbers of states. Additionally, the feature of the vaccine secondary structure was also predicted through the PSIPRED server. This result indicates a higher ranking quality of the peptide vaccine secondary structure (Fig. 5).

Furthermore, the Protein-Sol server was used to calculate the solubility of our MPXV vaccine construct and was found as soluble (0.549). The scaled solubility value (QuerySol) is shown in Fig. 4C.

3.5. Validation of peptide-based vaccine structure

The validation of protein structure was completed by ProSA and PROCHECK web tools. These assess the overall quality of MPXV vaccine protein structure. The refined vaccine model from the Ramachandran plot showed that the residues exist in four specific states as favored (42.7%), additionally allowed (33.1%), generously allowed (19.1%), and disallowed (5.1%) (Fig. 6A Table 3). The 'Z-score' of the MPXV vaccine advanced model was calculated as −2.78 (Fig. 6B). At the same time, the negative score specifies that it is a superior 3D protein model. Subsequently, we generated an additional local quality model using the ProSA server. It is portrayed by a substantial plot, whereas it showed the overall quality of the peptide model as acceptable (Fig. 6C). The 3D

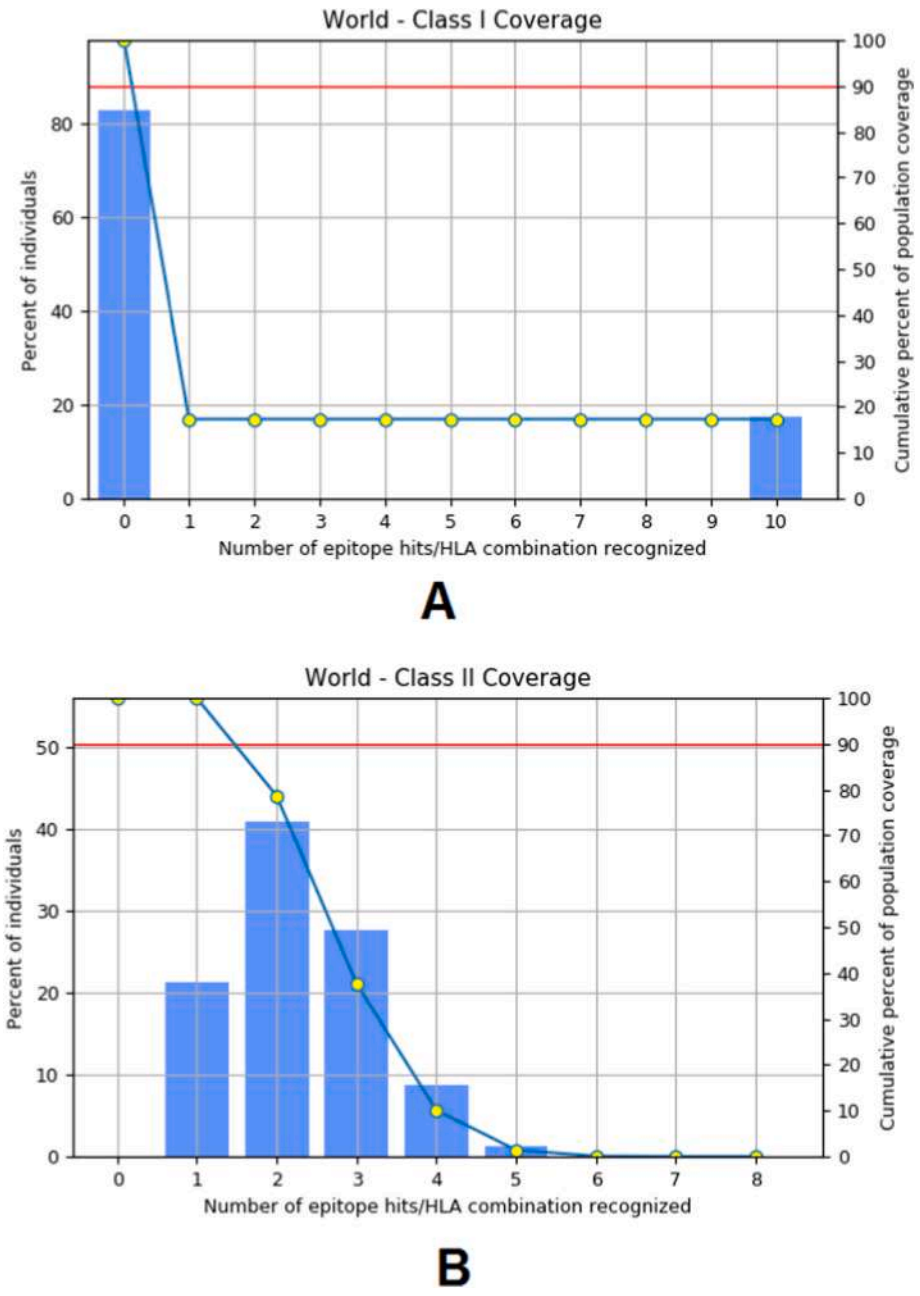


Fig. 2. Population coverage by selected CTL epitopes and their respective MHC binding alleles of our vaccine construct. We consider two alleles for our work (MHC-I and MHC-II) (A) It shows the worldwide population coverage of the MHC-I allele, (B) It shows the worldwide population coverage of the MHC-II allele.

structure of the MPXV vaccine protein is shown in Fig. 3C (ribbon model) and 3D (surface model).

3.6. Molecular docking of vaccine construct with the human TLR4/MD2 complex

Molecular docking was executed by the HDOCK server with vaccine construct and human TLR4/MD2 complex (PDB code: 4G8A). This TLR4/MD2 complex can activate proinflammatory cascades within the human body. Within the protein-protein docked complex among the top 10 models, the top-rank docking cluster with the highest binding score (−309.78) was selected to visualize binding interactions. Furthermore, the HawkDock web tool was employed to provide a molecular docking complex by machine learning approach, followed by the MM/GBSA. We also found the binding free energy of the docked complex was −98.37 kcal/mol. This score supports a superior binding affinity between the MPXV vaccine model as a ligand molecule and TLR4/MD2 as the receptor (Fig. 7A).

3.7. Normal mode analysis of peptide-based MPXV vaccine and TLR4/MD2 complex

Outputs of the normal mode analysis of the MPXV vaccine construct and the TLR4/MD2 are shown in Fig. 7B. The peak points of the deformability graph represent the parallel regions having deformability in the docked protein complex (Fig. 7C). The B-factor plot in the docking complex reflects an accurate visualization of the assessment among the PDB model and NMA (Fig. 7D). Fig. 7E shows the eigenvalue ($2.428517e\text{-}^{05}$) of the designed vaccine and TLR4/MD2 complex. Furthermore, the covariance map of the complex offers the interacting motion between the two molecules. In the present study, the correlated motion among the pair of amino acid residues is specified by red color,

Table 2
Antigenicity, allergenicity, solubility, and other physicochemical property evaluations of the primary protein sequence of the multi-epitopic potent vaccine candidate.

Sl. No.	Features	Assessment	Remark
1.	Antigenicity	1.1013 (antigenic score)	Probable antigen
2.	Allergenicity	Confirmed by AllerTOP v.2.0 server	Probable non-allergen
3.	Solubility	0.549	Soluble
4.	Number of amino acids	167	Suitable
5.	Molecular weight	18108.33 Da	Average
6.	Theoretical Isoelectric point (pI)	6.33	Slightly acidic
7.	Total number of atoms	2524	Suitable
8.	Formula	$C_{825}H_{1240}N_{214}O_{243}S_2$	–
9.	Estimated half-life	1 h (mammalian reticulocytes, *in vitro*) 30 min (yeast, *in vivo*) > 10 h (*Escherichia coli*, *in vivo*)	–
10.	Instability index	18.96	Stable
11.	Aliphatic index	88.08	Thermostable
12.	Grand average of hydropathicity (GRAVY)	0.069	Hydrophilic

and uncorrelated motion is marked as white color. In contrast, the anti-correlated motion is denoted as blue (Fig. 7F). The specialized elastic map of the vaccine and TLR4/MD2 complex signifies the assembly between the atoms of the larger molecule, and the stiffer regions are indicated by the darker grey color (Fig. 7G).

3.8. Immune simulation profiling of peptide-based potential vaccine

The immune stimulation by *in silico* technique of the potential vaccine candidate was carried out through primary and secondary immune responses by activating the immune system, containing HTL, CTL, viable memory cells, and other associated immune cells (DC cells, NK cells, etc.). Levels of the immunoglobulins (IgM + IgG) were observed to be drastically increasing after administering the vaccine construct. Subsequently, extended immune responses were reported against the MPXV through the high titers value of IgM and IgG (Fig. 4A). In the acting and resting phase, the innate immunity within the host body and B cell population was augmented with IgM, B isotype, and B-memory cells, which might be scaled up at 650–750 cells/mm^3 (Fig. 8A–C). Likewise, CTL cells were elevated and reached a maximum of 1117 cells/next ~8 days of immunization by vaccine and progressively reduced after 24 days (Fig. 8D).

Moreover, during the resting and active phase, the T cells exhibited abundant diversity within the T cell population (Fig. 8E). It was also noted that the raised HTL cells (in resting and active phases) performed a crucial role in creating adaptive immunity against the infections of MPXV. The raised cells produce the highest memory cells (Fig. 8F–G).

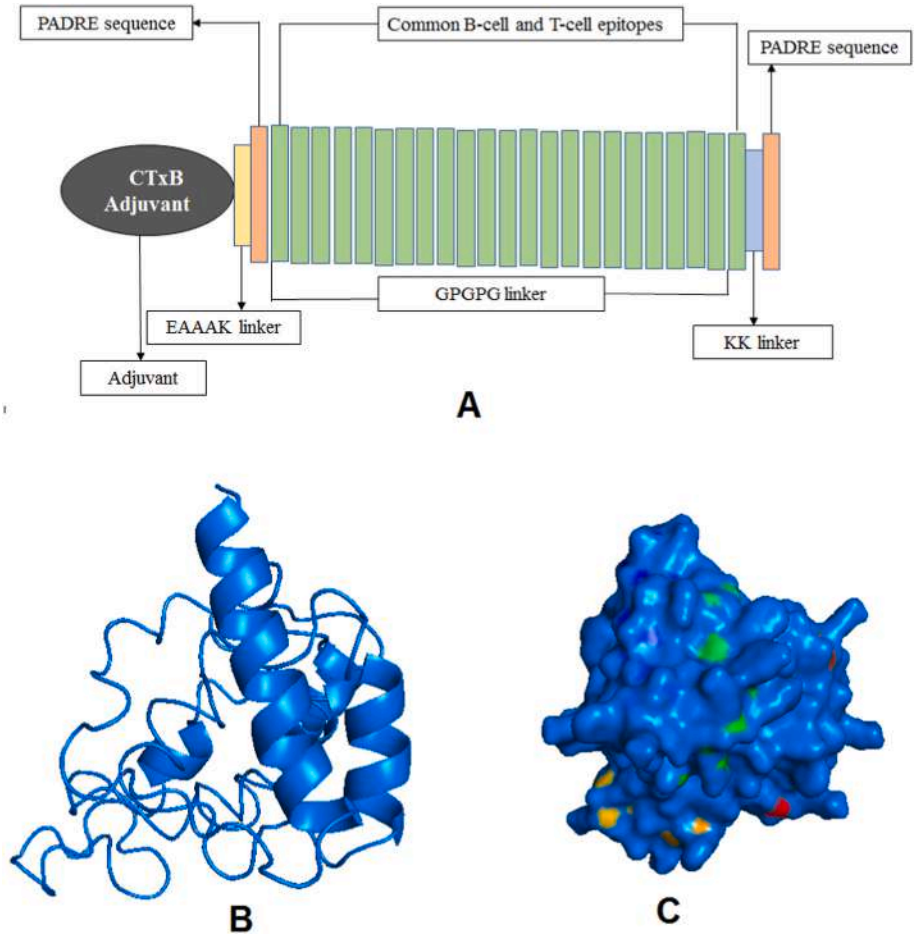


Fig. 3. Vaccine construct and its refined 3D structure from the MPXV full genome encoding proteins. (A) The graphical diagram depicts the final vaccine construct from antigenic epitopes of MPXV full genome encoding proteins using different linkers such as GPGPG and AAY (B) Ribbon model, (C) The model shows the Surface structure of the vaccine construct.

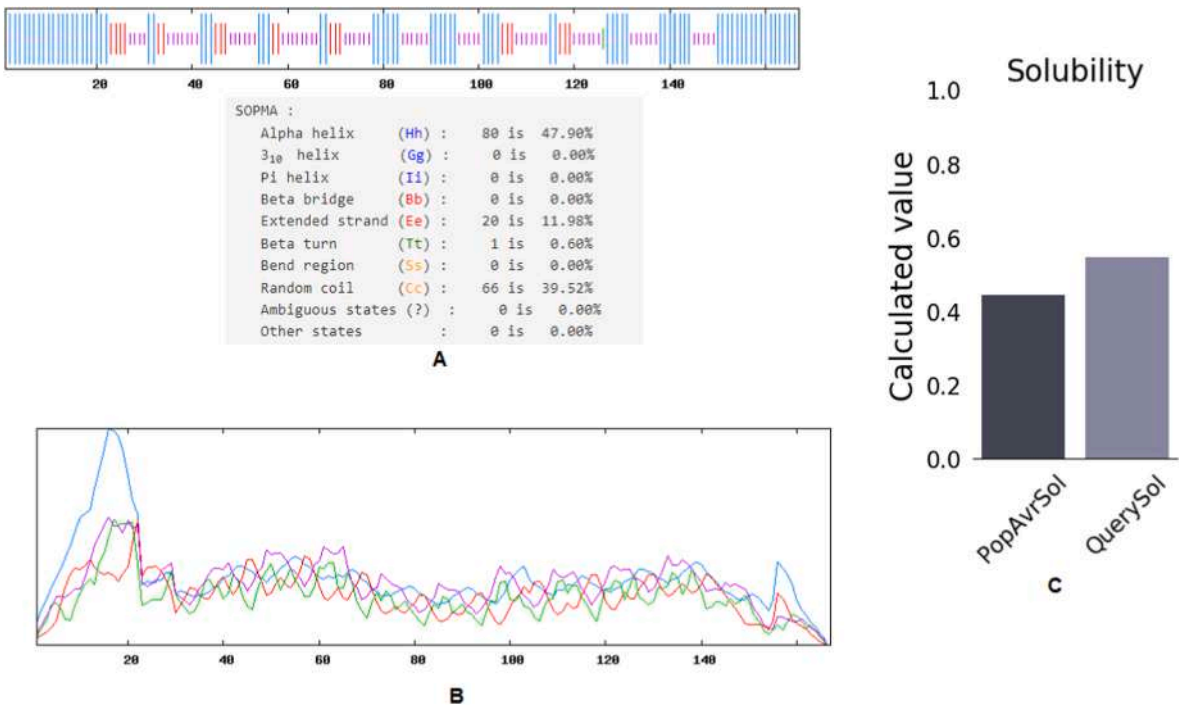


Fig. 4. Secondary structure prediction plot of the MPXV vaccine construct. Here the alpha helices were shown in blue color, while extended strands and beta turns were shown in red and green colors, correspondingly. (A) It shows the visualization of the result of the study and (B) The score curves for each predicted state of study (C) Protein-Sol predicted water solubility test of the potent vaccine candidate. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

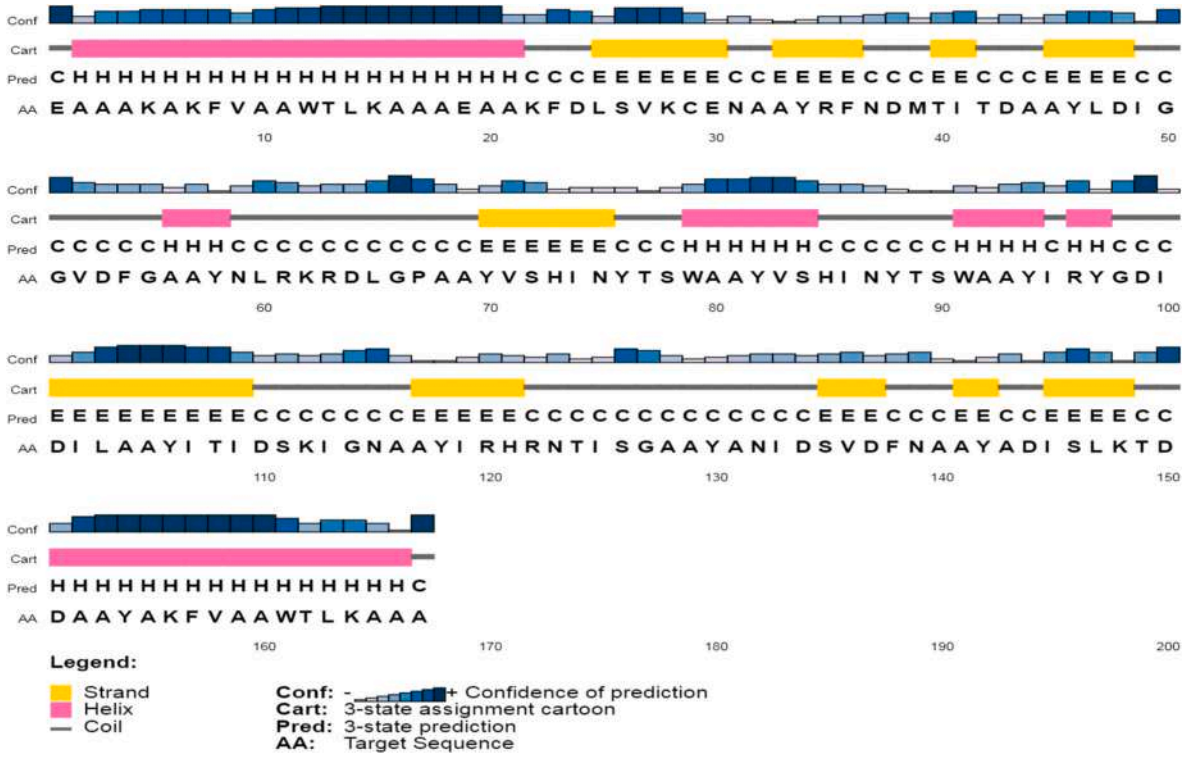


Fig. 5. The secondary structure analysis of the result of the multi-epitopic vaccine by the PESIPRED webserver.

Memory cells play a curial role in preventing and regulating viral infection and reinfection over the self-memorization by encountering pathogens. The successful administration of the vaccine candidate effectively elevated added regulatory elements of the immune system (e.g., interleukins, cytokines, and NK cells) (Fig. 8F-L). These consequences indicate that the designed MPXV vaccine could be considered a potent peptide-based next-generation vaccine for provoking a robust immune response to counteract MPXV infection.

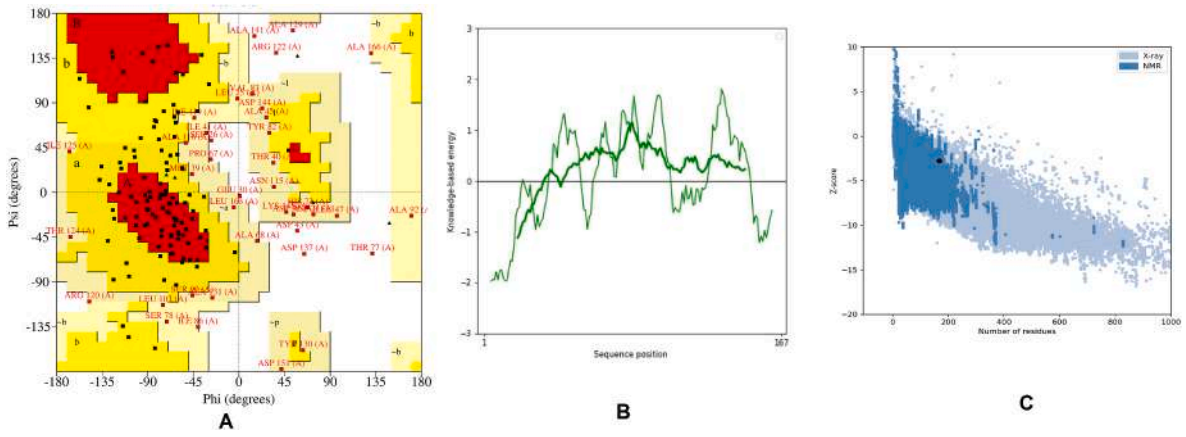


Fig. 6. Structural validation of MPXV vaccine candidate (A) The structure was validated through a Ramachandran plot for all residues of vaccine construct, and it shows the result of the validation, (B) It shows the energy plot of amino acid of vaccine component, (C) The 'Z' score of vaccine construct (black dot) within the 'Z' score range of experimentally proved structure.

Table 3
Distribution of amino acid residues showing in Ramachandran plot of the multiepitopic peptide-based potent vaccine candidate.

Amino acid position	Residue number	Percentage (%)	Total
Favored region	67	42.7%	157
Additional allowed region	52	33.1%	
Generously allowed region	30	19.1%	
Disallowed region	8	5.1%	
End-residues			2
Glycine residues			7
Proline residues			1
Total residues			167

4. Discussion

Due to the sudden surge of MPXV infection cases within a short time, global public health is under significant threat. Till today no specific vaccine is available that could resist the infection of MPXV, even though the vaccines for smallpox virus are also used against MPXV. But the administration of these vaccines are quite painful, and the vaccine's efficacy is also reduced due to the longer stockpile. Therefore, the present work was strategically planned for designing a peptide-based multiepitope subunit vaccine candidate from the whole genome encoded protein by computational tools and techniques that can induce an immunological response against the MPXV. We have planned to screen out all the potent antigenic epitopes (common B cell and T cell) from the whole genome encoded protein sequences of the MPXV. Subsequently, ten common epitopes with highest antigenic score with suitable peptide linkers and adjuvants were employed for developing the final vaccine construct. This proposed novel vaccine should operate on TLR4/MD2 complex (cell surface protein) as a target and be safe with the required functions. This peptide-based vaccine should be developed by synthesizing antigenic B cell and T cell epitopic sequences in a chemical process that stimulates specific antibody production [34,35].

The distribution and expression profile of HLA (MHC-I and MHC-II) alleles may differ worldwide between ethnicities and regions. Subsequently, for developing an effective vaccine, it is essential to conduct the baseline assessment of HLA allelic distribution within the whole world population [36,37]. Our study has shown the worldwide population coverage of the MHC-I and MHC-II alleles in respect to ten highest antigenic epitopes from the whole genome encoded protein of MPXV. The cumulative percentage (%) coverage for MHC I allele is 19.8% of ten epitopes for the entire world population. Whereas two epitopes have 70% coverage, three epitopes have 28% coverage, one epitope has 21% coverage, four epitopes have 9%, and five epitopes have 2% coverage, respectively (Fig. 2).

The B cell-derived T cell epitopes (9-mer) and common antigenic epitopes were characterized for the formulation of a multi-epitope peptide-based vaccine. The antigenic epitopes can provide a robust immunological response against MPXV in the host body. Our vaccine construct contains ten epitopes (9-mer) with 147 amino acids, PADRE sequence, CTxB adjuvant, and required peptide linker. Within the N-terminal end of the vaccine, the adjuvant is joined with the EAAAK peptide linker. It was merged for the dynamic separation of the bifunctional fusion protein [38]. In addition, the EAAAK linker was also joined by the PADRE sequence, which acts as a helper T cell epitope for uprising the CTL response in diverse antigens.

The outcomes of different bio-computational tools and servers also established that the final vaccine construct was slightly acidic and stable in the human physiological pH range. Furthermore, the assessed aliphatic index score specified that the construct is also thermostable [39]. Consequently, the positive value of GRAVY indicated its hydrophobic nature, with no such viable interaction with the water or non-soluble in water [40].

In the present study, the I-TASSER server was used for the development of the 3D structure of this MPXV whole genome encoded protein-derived peptide vaccine candidate. The 3D structure was also critically validated by ProSA and PROCHECK web servers, where the local energy plot and 'Z' score were accessed to understand the proper folding of the protein. The ProSA server predicted a 'Z' score negative value (−2.78), signifying an upright signal for the structural quality assessment [41]. Along with the analysis of the local quality model, this model has been considered as a reliable model.

The subsequent analysis also showed that the peptide vaccine constructs strongly interact with TLR4. It was also noted that the TLRs played an important role in recognizing pathogens and stimulating the proper innate immune response against infectious agents [11]. In the current study, we have executed the molecular docking between the TLR4/MD2 complex and the peptide vaccine construct using the HwkDock and HDOCK web servers. The negative docking score demonstrated that the vaccine construct contains high binding interaction characteristics. It supports that the potential vaccine candidate can trigger the TLR4 activation and boosts the immune response against the MPXV. The TLR4/MD2 and vaccine candidate docked complex bear the free binding energy of −98.37 kcal/mol. This value specifies a reliable binding affinity between the immune sensing (TLR4) cell receptor protein and our potential vaccine candidate.

The normal mode analysis (NMA) study was carried out to analyze the comparative deformability, molecular motion, and the B factor from the protein PDB. Our analysis has shown the eigenvalue of the studied

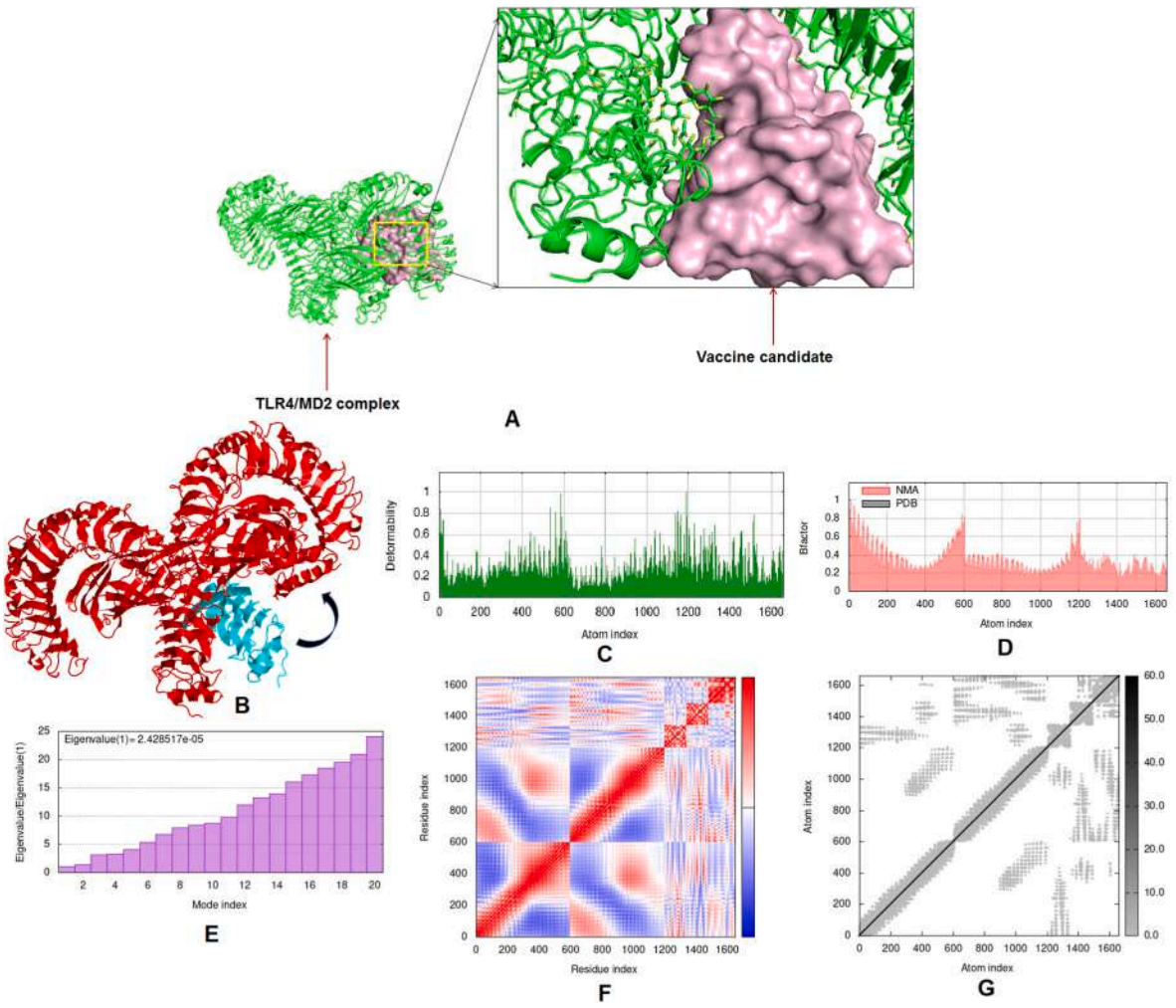


Fig. 7. The structure shows the docked complex and the outcome of our vaccine construct's NMA (Normal Mode Analysis). our vaccine construct was docked with TLR4/MD2 (A) docked complex (vaccine construct docked with TLR4/MD2) illustrates the ribbon model showing TLR4/MD2, and surface model shown vaccine construct, (B) Mobility of the docking complex of our vaccine and TLR4/MD2 indicated with arrows, (C) Deformability plot of the docking complex, (D) Calculated B-factor of NMA and PDB B-factor of our vaccine construct, (E) The eigenvalue for the docking complex, (F) Covariance matrix map of atomic pair of amino acid residues of docked complex, (G) Connection spring map of the elastic network model of the docked complex.

peptide sequence as 2.428517e-^{05}. This calculated eigenvalue designated the superior flexibility of the final construct of the vaccine.

The immune simulation profiling of the vaccine construct confirmed that the vaccine candidate could trigger both types (humoral and cellular) of immune responses against MPXV infection. Moreover, the secondary immune response produced by the designed vaccine was expressively higher than the primary immune response.

Finally, our vaccine construct showed a superior result from its structural and functional properties as evaluated during *in silico* studies. The structural validation shows a 'Z' score in negative value (−2.78), indicating a reliable protein model, docked complex of TLR4/MD2, and a vaccine candidate with a free binding energy of −98.37 kcal/mol specifies a reliable binding affinity for immune sensing by molecular interactions. Furthermore, in NMA analysis, the eigenvalue of the studied docked peptide sequence (2.428517e-^{05}) offered greater flexibility, good molecular motion, and lesser protein deformability. The immune simulation profiling of vaccine candidates also supports the production of abundant diversity in the T cell population and the rise of HTL cells in the mammalian immune system. Therefore, all the observed properties indicate that our vaccine construct is an ideal next-generation vaccine against MPXV.

5. Conclusion

The present work aimed to design a novel, peptide-based multi-epitopic next-generation potential vaccine candidate against the MPXV by employing various bioinformatics approaches. The vaccine was developed from the whole-genome encoded proteins, and it was found that the MPXV epitopic vaccine construct showed superior characterized properties in terms of antigenicity, non-allergenicity, and physico-chemical properties. Therefore, it might be concluded that our designed vaccine construct is not only ideal but also effective and safe to be used against MPXV.

This designed vaccine candidate might serve as a significant milestone and aid in developing an antiviral vaccine against MPXV. It might be administered to the host body (human) of MPXV next to the further outcomes of the fruitful results from the pre-clinical and clinical trials. In addition, the structural validation of our vaccine candidate and the binding affinity to immune receptor were performed, followed by the protein-protein molecular docking and normal mode analysis. Even though immune system simulation through the immunoinformatics approaches confirmed that the vaccine construct could elicit a good immune response, this vaccine candidate can deliver an extensive range of protection against MPXV. However, successive validation (*in vivo* and *in vitro*) is required to assure the vaccine candidate's effectiveness before

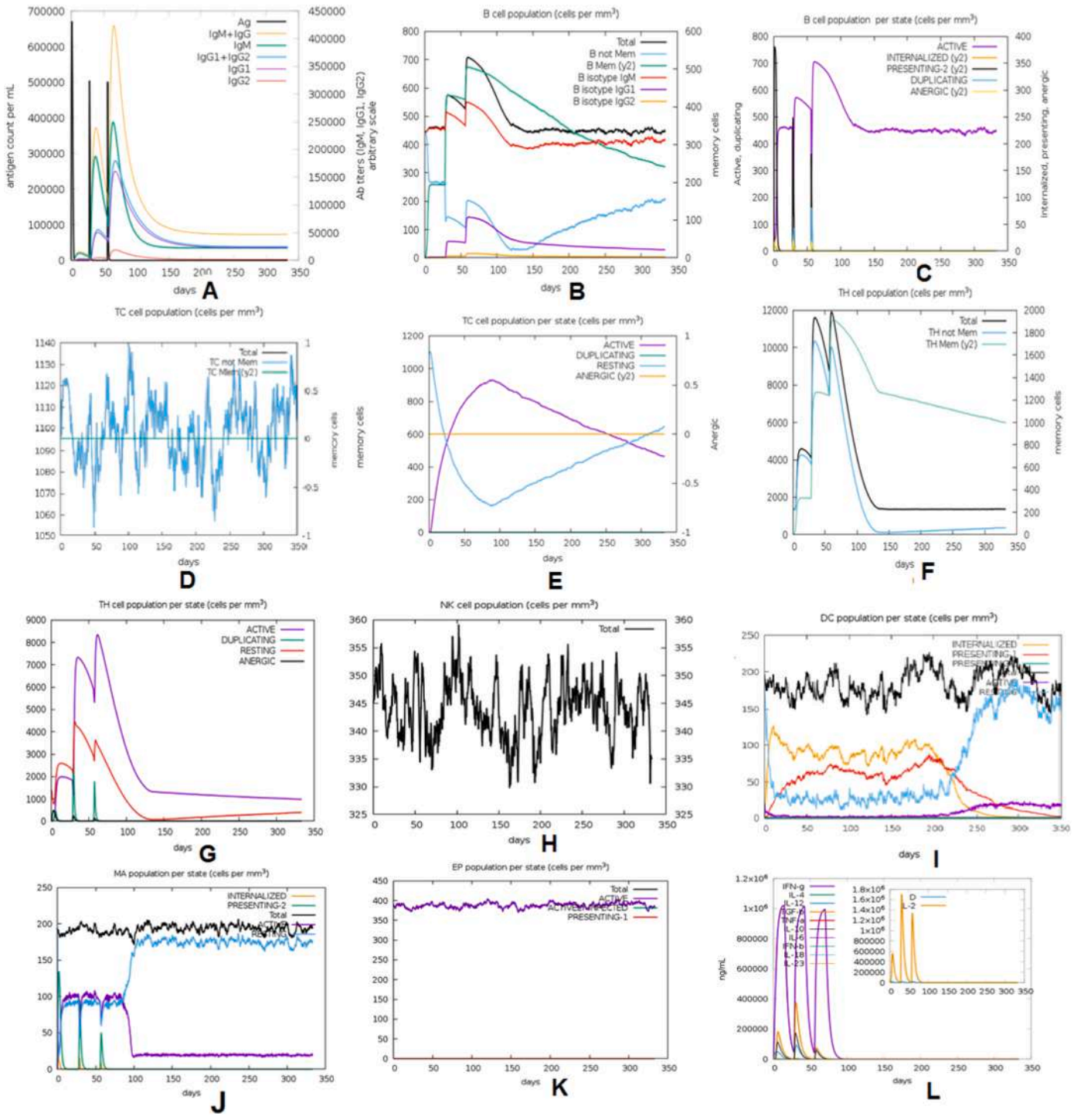


Fig. 8. The immune response of humans after injection of our MPXV vaccine construct. The study noted different results from the *In silico* simulation and machine learning approaches. (A) The simulation shows the contract can elevate immunoglobulins. The noted elevation of immunoglobulins at different concentrations of antigen, (B) The study indicates the population of B lymphocytes (IgM, IgG1, and IgG2) after three injections of our vaccine construct (C) The figure depicts the analysis outcome of the population per entity-state (i.e., showing counts for active, presenting on class-II, internalized the Ag, duplicating and antigenic by the different color variant, (D) The figure informs the cytotoxic T lymphocyte population in the time (days) after injection of our MPXV vaccine construct, (E) The figure illustrates the Cytotoxic T lymphocytes population in different states; resting and active, in the time (days) after injection of our MPXV vaccine construct (F) It shows the total count of TH cell population along with memory cells and sub-divided in isotypes IgM, IgG1 and IgG2 after injection of our MPXV vaccine, (G) It shows the population per entity-state of Helper T cell count in the resting and active states after injection of our vaccine, (H) The behaviour of the Population of Natural Killer cells. (I) The behaviour of the population of Dendritic cells in the active and resting states after injection of our vaccine, (J) Its shows the population of macrophages after MPXV vaccination, (K) Concentration of cytokines and interleukins with Simpson index [D], (L) Total count of EC cells that is broken down to active, virus-infected and presenting on class-I MHC molecule. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

its administration in the human body.

Ethical approval and consent to participate

Not applicable.

Source of funding

No funding received.

Data availability statement

All data included within the manuscript.

CRediT authorship contribution statement

Manojit Bhattacharya: Writing – original draft, Methodology, Investigation. **Srijan Chatterjee:** Data curation, validation, Formal analysis. **Sagnik Nag:** Data curation, validation, Formal analysis. **Kuldeep Dhama:** Writing – review & editing, Visualization. **Chiranjib Chakraborty:** Conceptualization, Writing – review & editing, Supervision, Project administration, All authors have read and agreed to the published version of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Sarwar S, et al. Re-Emergence of monkeypox amidst delta variant concerns: a point of contention for public health virology? J Med Virol 2022;94(3):805–6.
[2] Kumar N, et al. The 2022 outbreak and the pathobiology of the monkeypox virus. J Autoimmun 2022:102855.
[3] Kampf G. Efficacy of biocidal agents and disinfectants against the monkeypox virus and other orthopoxviruses. J Hosp Infect 2022;127:101–10.
[4] Cunha BE. Monkeypox in the United States: an occupational health look at the first cases. AAOHN J 2004;52(4):164–8.
[5] Centers for Disease Control and Prevention CDC. Monkeypox outbreak global map. Centers Dis Control Prev; 2022. 2022, https://www.cdc.gov/poxvirus/monkeypox/response/2022/world-map.html. [Accessed 11 September 2022].
[6] Bhattacharya M, Dhama K, Chakraborty C. Recently spreading human monkeypox virus infection and its transmission during COVID-19 pandemic period: a travelers' prospective. Trav Med Infect Dis 2022:102398.
[7] Kozlov M. Monkeypox goes global: why scientists are on alert. Nature 2022;606(7912):15–6.
[8] Monto AS. Vaccines and antiviral drugs in pandemic preparedness. Emerg Infect Dis 2006;12(1):55.
[9] Chakraborty C, et al. Appearance and re-appearance of zoonotic disease during the pandemic period: long-term monitoring and analysis of zoonosis is crucial to confirm the animal origin of SARS-CoV-2 and monkeypox virus. Vet Q 2022;42(1):1–11. https://doi.org/10.1080/01652176.2022.2086718.
[10] Simpson K, et al. Human monkeypox–After 40 years, an unintended consequence of smallpox eradication. Vaccine 2020;38(33):5077–81.
[11] Bhattacharya M, et al. Immunoinformatics approach to understand molecular interaction between multi-epitopic regions of SARS-CoV-2 spike-protein with TLR4/MD-2 complex. Infect Genet Evol 2020;85:104587.
[12] Ghosh P, et al. A novel multi-epitopic peptide vaccine candidate against Helicobacter pylori: in-silico identification, design, cloning and validation through molecular dynamics. Int J Pept Res Therapeut 2021;27(2):1149–66.
[13] Jespersen MC, et al. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res 2017;45(W1):W24–9.
[14] Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics 2016;32(4):511–7.
[15] Bui H-H, et al. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinf 2006;7(1):1–5.
[16] Kavoosi M, et al. Strategy for selecting and characterizing linker peptides for CBM9-tagged fusion proteins expressed in Escherichia coli. Biotechnol Bioeng 2007;98(3):599–610.
[17] Sadraeian M, et al. Cloning and expression of CtxB-StxB in Esherichia coli: a challenge for improvement of immune response against StxB. Iran J Pharm Sci 2011;7:185–90.
[18] Solanki V, Tiwari M, Tiwari V. Prioritization of potential vaccine targets using comparative proteomics and designing of the chimeric multi-epitope vaccine against Pseudomonas aeruginosa. Sci Rep 2019;9(1):1–19.
[19] Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinf 2007;8(1):1–7.
[20] Dimitrov I, et al. AllerTOP v. 2—a server for in silico prediction of allergens. J Mol Model 2014;20(6):1–6.
[21] Gasteiger E, et al. Protein identification and analysis tools on the ExPASy server. In: The proteomics protocols handbook; 2005. p. 571–607.
[22] McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics 2000;16(4):404–5.
[23] Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Bioinformatics 1995;11(6):681–4.
[24] Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 2010;5(4):725–38.
[25] Ko J, et al. GalaxyWEB server for protein structure prediction and refinement. Nucleic Acids Res 2012;40(W1):W294–7.
[26] Laskowski RA, et al. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26(2):283–91.
[27] Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 2007;35(suppl_2):W407–10.
[28] Schrodinger L. The PyMOL molecular graphics system. Versiones 2010;1(5). 0.
[29] Weng G, et al. HawkDock: a web server to predict and analyze the protein–protein complex based on computational docking and MM/GBSA. Nucleic Acids Res 2019;47(W1):W322–30.
[30] Yan Y, et al. The HDOCK server for integrated protein–protein docking. Nat Protoc 2020;15(5):1829–52.
[31] López-Blanco JR, et al. iMODS: internal coordinates normal mode analysis server. Nucleic Acids Res 2014;42(W1):W271–6.
[32] Rapin N, et al. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. PLoS One 2010;5(4):e9862.
[33] Castiglione F, et al. How the interval between prime and boost injection affects the immune response in a computational model of the immune system. Comput Math Methods Med 2012. 2012.
[34] Dermime S, et al. Vaccine and antibody-directed T cell tumour immunotherapy. Biochim Biophys Acta, Rev Cancer 2004;1704(1):11–35.
[35] Meloen RH, et al. Synthetic peptide vaccines: unexpected fulfillment of discarded hope? Biologicals 2001;29(3–4):233–6.
[36] Bui Huynh-Hoa, et al. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinf 2006;7(1):1–5.
[37] Oyarzun Patricio, Kobe B. Computer-aided design of T-cell epitope-based vaccines: addressing population coverage. Int J Immunogenet 2015;42(5):313–21.
[38] Kim HJ, et al. Intranasal vaccination with peptides and cholera toxin subunit B as adjuvant to enhance mucosal and systemic immunity to respiratory syncytial virus. Arch Pharm Res (Seoul) 2007;30(3):366–71.
[39] Ikai A. Thermostability and aliphatic index of globular proteins. J Biochem 1980;88(6):1895–8.
[40] Ali M, et al. Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. Sci Rep 2017;7(1):1–13.
[41] Bhattacharya M, et al. TN strain proteome mediated therapeutic target mapping and multi-epitopic peptide-based vaccine development for Mycobacterium leprae. Infect Genet Evol 2022;99:105245.