scientific reports

Check for updates

OPEN

Computationally designed multi-epitope vaccine construct targeting the SARS-CoV-2 spike protein elicits robust immune responses in silico

Varughese Deepthi[1,3], Aswathy Sasikumar[1,2,3], Kochupurackal P Mohanakumar[1,2] & Usha Rajamma[1]✉

Our research is driven by the need to design an advanced multi-epitope vaccine construct (MEVC) using the S-protein of SARS-CoV-2 to combat the emergence of new variants. Through rigorous computational screening, we have identified linear and discontinuous B-cell epitopes, CD8 + and CD4 + T-cell epitopes, ensuring extensive MEVC coverage across 90.03% of the global population. The MEVC, featuring four CD4 + and four CD8 + T-cell epitopes connected linearly with two adjuvant proteins on both ends, has been carefully designed to elicit robust immune response. Our in-silico analysis has confirmed the construct's antigenicity, non-allergenicity, and non-toxicity with optimized codon sequences for enhanced expression in E. coli K12. Furthermore, molecular docking and dynamics analyses have demonstrated its strong binding affinity with TLR-3 and TLR 4, and in-silico immune simulation yielded promising results on heightened B-cell and T-cell-mediated immunity. However, wet lab experiments are essential to validate computational findings to revolutionize the development of vaccines against SARS-CoV-2.

COVID-19, emerged in late 2019 has had a devastating global impact, leading to more than four million casualties. The virus responsible for the pandemic, known as severe acute respiratory syndrome-corona virus 2 (SARS-CoV-2), shares similarities with previous coronaviruses that have affected human health[1]. SARS-CoV-2 is a highly infectious virus that is transmitted rapidly from animals to humans and then within the human population[2]. While some coronaviruses cause respiratory and gastrointestinal illnesses, more dangerous ones like SARS-CoV, MERS-CoV, and SARS-CoV-2 can attack cells in the respiratory tract, resulting in severe breathing difficulties and lung damage[3]. Unfortunately, there isn't any treatment currently available to prevent or treat these infections.

The emergence of the SARS-CoV-2 virus has sparked economic uncertainty and social instability. As an RNA virus, it has undergone frequent mutations, resulting in new variants[4,5]. Scientists worldwide are dedicated to developing a vaccine to combat and manage its widespread infection. Once the virus's proteome is identified, the epitope selection phase becomes critical in optimizing the production of peptide-based vaccines. In silico epitope identification is a widely recognized method in vaccine development, providing guidance on immune targets for use in vaccine design[6,7].

The spike glycoprotein (S-protein), of SARS-CoV-2 has a receptor-binding domain (RBD) that is crucial for binding to the host angiotensin-converting enzyme 2 (ACE2) and fusing with the host mucosal cells in the oro-nasopharyngeal, gut, and lung areas. The S-protein is a key target in vaccine and therapeutic development due to its significance in SARS-CoV-2[8,9]. Our recent study used molecular modeling and docking studies to pinpoint the essential residues and domains in the S-protein responsible for effective binding with host ACE2[10]. To ensure the effectiveness of vaccines and therapies, it is essential to produce neutralizing antibodies that target specific

[1]Centre for Development and Aging Research, Inter University Centre for Biomedical Research & Super Speciality Hospital, Mahatma Gandhi University Campus at Thalappady, Rubber Board P.O, Kottayam 686009, Kerala, India. [2]Virus Research and Diagnostic Centre, Inter University Centre for Biomedical Research & Super Speciality Hospital, Mahatma Gandhi University Campus at Thalappady, Rubber Board P.O, Kottayam 686009, Kerala, India. [3]Varughese Deepthi and Aswathy Sasikumar contributed equally. ✉email: usharajamma1962@gmail.com; usharajamma@mgu.ac.in

residues or functional domains of the S-protein[11]. Although COVID-19 patients have antibodies against SARS-CoV-2 in their systems, it is unclear which ones precisely combat the virus and which ones may exacerbate the disease. Therefore, it is crucial to optimize vaccine efficacy and safety by increasing the production of neutralizing antibodies and reducing the risk of pulmonary inflammation or antibody-dependent enhancement (ADE)[12,13]. Researchers identify both B-cell and T-cell epitopes to create epitope-based vaccines, therapeutic antibodies, and diagnostic tools[14]. Antibodies recognize specific antigenic determinants or B-cell epitopes at distinct regions, while T-cells are activated by SARS-CoV-2 peptides presented on infected cells through human leukocyte antigen (HLA) molecules encoded by the major histocompatibility complex (MHC) genes. T-cell epitope prediction aims to identify peptides within the viral antigen that can stimulate CD4 or CD8 T-cells[15]. Assessments require synthetic peptides derived from the antigens to confirm the immunogenicity or the ability to stimulate T-cells[16]. Predicting epitopes using immuno-informatics tools such as B-cell epitope prediction and T-cell epitope prediction helps reduce the number of mandatory in vitro experiments and serves as a crucial tool in peptide vaccine design and testing the therapeutic effect[17].

Epitope-based vaccines have several advantages over traditional ones. They are non-self-reactive, resulting in enhanced immunogenicity, stability, and solubility. By using immunoinformatics tools vaccines can be designed by identifying effective epitopes, saving time and cost compared to traditional vaccine development. While studies have been conducted on both structural and non-structural proteins of the virus[18,19], our focus is on the S protein because of its crucial role and prominent presence on the virus surface. The S-protein is a component of SARS-CoV-2 envelope proteins, essential for attaching to and entering target cells through ACE2. In COVID-19 patients, the immune system response mainly targets the virus's structural proteins, particularly the S-protein[20]. Therefore, we are concentrating on the S protein as a potential target for developing an epitope-based vaccine against SARS-CoV2.

Our research focused on creating an innovative vaccine structure by examining the S-protein to predict B-cell and T-cell epitopes. We employed a comprehensive approach to pinpoint promising epitope components based on antigenicity, non-allergenicity, non-toxicity, and literature support. To the best of our knowledge, this is the first effort of its kind. We integrated immunogenic regions of the S-protein into the vaccine structure to generate neutralizing antibodies capable of obstructing virus binding and host cell entry. We aimed to design a potent multi-epitope vaccine construct (MEVC) that is highly antigenic, non-allergic, non-toxic, and capable of eliciting both humoral and cellular immune responses.

Results

Sequence analysis of Spike protein from all SARS-CoV-2 variants

We have obtained the amino acid sequences of the spike protein from the original Wuhan_hCoV-19 strain and seven major SARS-CoV-2 variants from the protein database maintained by the National Center for Biotechnology Information (NCBI). Utilizing Clustal Omega, a multiple sequence alignment (MSA) was conducted to identify non-conserved regions. The results of the MSA are provided in Supplementary Fig. 1. This MSA file enabled the selection of predicted B-cell and T-cell epitopes within the conserved region among all SARS-CoV-2 variants.

Prediction of linear B-Cell epitopes

We employed the IEDB Analysis Resource to anticipate B-Cell epitopes through the evaluation of various parameters. Furthermore, we made use of the ABCpred and Bpred servers for the same purpose. Subsequently, we curated the epitopes by consolidating predictions from multiple servers to guarantee precision (Supplementary Table 1). Following a comprehensive analysis, we successfully determined 18 linear B-cell epitopes that were universally predicted by all the tools. Please refer to Table 1 for the detailed list.

Discontinuous epitope prediction

We employed DiscoTope, Ellipro, and SEPPA prediction tools to detect discontinuous epitopes in the SARS-CoV-2 S-protein. The Discotope score, based on the predictive capability of the algorithm, revealed a positive prediction for values exceeding −3.7. The Ellipro method evaluates the 3D shape of proteins as a series of ellipsoids and computes a Protrusion Index (PI) score for each ellipsoid, identifying the highest scoring ellipsoid as the region of maximum surface accessibility. The SEPPA 3.0 prediction tool analyzes the 3D structure of the protein and assigns scores to individual residues based on the likelihood of being part of the epitope. Through these tools, we discovered potential epitope residues and predicted six discontinuous epitopes. Detailed information is available as supplementary data (Supplementary Fig. 2, Tables 2, 3 and 4). A consensus approach was employed to select common residues from these predictions. The most promising B-cell epitopes labelled, as B-Ep1 to B-Ep5 were designated based on their PI scores, indicating significant promise, especially for B-Ep1 to B-Ep4. Supplementary Fig. 3 illustrates the predicted discontinuous epitopes within the SARS-CoV2 S-protein structure (PDB: 6xe1).

Upon comparison of the anticipated linear and discontinuous epitopes, it was found that 53 residues from Table 2 are present in both prediction methods, implying a high likelihood of being part of the B-cell epitope. Except B-Ep2 and B-Ep3 residues, these epitopes were conserved in the spike protein among SARS-CoV-2 variants and are located closely to each other as depicted in Supplementary Fig. 3. This region plays a significant role in eliciting an immune response by binding to antibodies. The identified properties and thresholds can be leveraged to elicit a humoral response against SARS-CoV-2 via B-cell epitopes. However, it is important to note that B-cell epitope prediction is generally considered to be less reliable than its T-cell counterpart; hence the study also focused on predicting T-cell epitopes.

Predicted linear epitopes	Length	Start position	End position
SQCVNLTTRTQLPPAYTNSFTRG	23	13	35
HKNNKSWME	9	146	154
QPFLMDLEGKQGNFKNL	17	173	189
RSYLTPGDSSS	11	246	256
FTVEKGIYQTSNFRVQP	17	306	322
EVRQIAPGQTGKIADYNYKLP	21	406	426
NLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTN	62	440	501
ELLHAPATVCGPKKSTNLVKN	21	516	536
NCTEVPVAIHADQLTPTWRVYSTGSNVFQ	29	616	644
ASYQTQTNSPRRARSVASQ	19	672	690
YTMSLGAENSVAYSNN	16	694	710
LPDPSKPSKR	10	806	815
LADAGFIKQYGDCLGD	16	828	843
GQSKRVDFC	9	1035	1043
RNFYEPQIITTD	12	1107	1118
VNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGI	40	1133	1172
LIDLQELGKY	10	1197	1206
SCCKFDEDDSEPVLKG	16	1252	1267

Table 1. The anticipated common linear epitopes derived from four or more methods provided by IEDB and ABCpred.

Conformational epitopes	Epitope region	Length	PI score
B-Ep1	I326, V327, K529, S530, T531, N532	6	0.91
B-Ep2	L455, K458, S459, Q474, A475, G476, S477, T478, P479, C480, N481, G482, V483, E484, G485, F486, N487, C488, Y489, F490, P491	21	0.70
B-Ep3	R403, G404, D405, N437, N439, N440, L441, S443, K444, V445, G446, G447, N448, Y449, N450, Y451, G496, F497, Q498, P499, T500, N501, G502, V503, G504, Y505, Q506, P507	28	0.67
B-Ep4	P330, T333, V362, A363, V382, L390, C391, A522, T523, C525, G526, P527	12	0.65
B-Ep5	A411, P412, G413, Q414, T415, P426, D427, D428	8	0.54

Table 2. Predicted discontinuous epitopes of B-cells as determined by three different analytical tools. B-Ep: B-cell epitope.

T-Cell epitope prediction

In the pursuit of identifying highly immunogenic peptides originating from the SARS-CoV-2 S-protein capable of eliciting $CD4^+$ or $CD8^+$ T-cells upon binding and presentation by MHC molecules, T-cell epitopes were predicted. It is noted that HLA variants are expressed at varying frequencies across different ethnic groups. Interestingly, diverse HLA variants can bind to similar sets of peptides and have been categorized into HLA supertypes accordingly. The correlation between HLA-A*02:01, HLA-A*02:06, HLA-A*29:02, and HLA-A*01:01 was notable in the cluster analysis of the MHC-I alleles. The IEDB recommended NetMHCpan 4.1 EL and the consensus prediction method identified MHC class-I Cytotoxic T-lymphocyte (CTL) epitopes with a size 9–10 for the HLA set including HLA class I alleles HLA-A*01:01, HLA-A*02:02, HLA-A*03:01, HLA-B*35, HLA-C*04:01, HLA-C*12:03, and HLA-DRB1*15:01. The T-cell epitopes and their corresponding amino acid positions, predicted by at least three tools, with a peptide length of 9, with a high score indicating high affinity to MHCI molecule. The top 9 epitopes were selected based on a threshold score of 0.65 from the IEDB server. Table 3 describes the selected T-cell epitopes with scores exceeding 0.65, which were subjected to further analysis. In addition, ToxinPred predicted these peptides to be non-toxin (Table 3). For more in-depth information on the IEDB NetMHCpan 4.1 EL prediction outcomes, refer to Supplementary Table 5.

An analysis utilizing a data-driven and consensus-based approach has successfully predicted nine CTL epitopes from the S-protein of SARS-COV-2. Notably, these predicted epitopes, with the exception of T-Ep7, were situated within the conserved region of the virus, a finding supported by the examination of the multiple sequence alignment (MSA) of SARS-CoV-2 variants (Supplementary Fig. 1). Employing the NetMHCII 2.3 algorithm of IEDB server revealed a total of 34 Helper T lymphocyte (HTL) epitopes specific to HLA-DQA1*01:02/DQB1*06:02. These epitopes have undergone comprehensive assessments for antigenicity, toxicity, and stability. In Tables 4 and 19 epitopes with high rank and low IC50 values are listed. Additionally, the majority of these epitopes have the potential to induce cytokines such as IL-4, IL-10, or IFN-gamma, as verified by the SVM score from the IL4pred, IL-10Pred, and IFN epitope servers. For further details, please refer to Supplementary Table 6.

T-Cell Epitopes	Position	Peptide	Length	Score	Toxicity
T-Ep1	865	LTDEMIAQY	9	0.99	Non-toxin
T-Ep2	896	IPFAMQMAY	9	0.99	Non-toxin
T-Ep3	687	VASQSIIAY	9	0.97	Non-toxin
T-Ep4	604	TSNQVAVLY	9	0.91	Non-toxin
T-Ep5	258	WTAGAAAYY	9	0.91	Non-toxin
T-Ep6	440	NLDSKVGGNY	10	0.69	Non-toxin
T-Ep7	652	GAEHVNNSY	9	0.69	Non-toxin
T-Ep8	361	CVADYSVLY	9	0.68	Non-toxin
T-Ep9	30	NSFTRGVYY	9	0.67	Non-toxin

Table 3. A comprehensive list of shared CTL epitopes across HLA-A, HLA-B and HLA-C alleles. *T-Ep: T-cell Epitope.

Allele	Start	End	Core-peptide	Peptide	IC50	Rank
HLA- DQA1*01:02/DQB1*06:02	1009	1023	IRAAEIRAS	TQQLIRAAEIRASAN	10.8	0.03
HLA-DQA1*01:02/DQB1*06:02	602	616	NQVAVLYQD	TNTSNQVAVLYQDVN	55.5	1.2
HLA-DQA1*01:02/DQB1*06:02	39	53	FRSSVLHST	PDKVFRSSVLHSTQD	62.6	1.4
HLA-DQA1*01:02/DQB1*06:02	886	900	GAGAALQIP	WTFGAGAALQIPFAM	66.8	1.5
HLA-DQA1*01:02/DQB1*06:02	256	270	TAGAAAYYV	SGWTAGAAAYYVGYL	67.2	1.5
HLA-DQA1*01:02/DQB1*06:02	885	899	GAGAALQIP	GWTFGAGAALQIPFA	73.9	1.8
HLA-DQA1*01:02/DQB1*06:02	633	647	STGSNVFQT	WRVYSTGSNVFQTRA	106.3	3.1
HLA-DQA1*01:02/DQB1*06:02	868	882	YTSALLAGT	EMIAQYTSALLAGTI	128.7	4
HLA-DQA1*01:02/DQB1*06:02	684	698	ASQSIIAYT	ARSVASQSIIAYTMS	135.2	4.2
HLA-DQA1*01:02/DQB1*06:02	1169	1183	INASVVNIQ	ISGINASVVNIQKEI	135.9	4.3
HLA-DQA1*01:02/DQB1*06:02	635	649	STGSNVFQT	VYSTGSNVFQTRAGC	139.9	4.5
HLA-DQA1*01:02/DQB1*06:02	754	768	SFCTQLNRA	LQYGSFCTQLNRALT	140.5	4.5
HLA-DQA1*01:02/DQB1*06:02	604	618	NQVAVLYQD	TSNQVAVLYQDVNCT	154.3	5.1
HLA-DQA1*01:02/DQB1*06:02	883	897	TFGAGAALQ	TSGWTFGAGAALQIP	155.8	5.2
HLA-DQA1*01:02/DQB1*06:02	683	697	ASQSIIAYT	RARSVASQSIIAYTM	171.3	6
HLA-DQA1*01:02/DQB1*06:02	688	702	IAYTMSLGA	ASQSIIAYTMSLGAE	175.1	6.1
HLA-DQA1*01:02/DQB1*06:02	867	881	AQYTSALLA	DEMIAQYTSALLAGT	183.6	6.4
HLA-DQA1*01:02/DQB1*06:02	686	700	SQSIIAYTM	SVASQSIIAYTMSLG	199.5	7.1
HLA-DQA1*01:02/DQB1*06:02	464	478	DISTEIYQA	FERDISTEIYQAGST	207.5	7.3

Table 4. List of predicted HTL epitopes using NetMHCII 2.3.

Modeling and evaluation of the 3D structure of T-cell epitopes

The PEPFOLD 3.0 webserver was utilized to predict the three-dimensional structure of T-cell epitopes T-Ep1 to T-Ep5, which exhibited high CTL epitope prediction scores. Furthermore, we analyzed the Ramachandran plot (Supplementary Fig. 4) for each modeled peptide. Our observations revealed that epitopes demonstrating regular secondary structures displayed all amino acids with φ and ψ angles in the highly allowed conformation, positioned in the third quadrant. The models representing T-Ep1 to T-Ep5 were identified as conformationally stable 3D structures and were subsequently employed for binding analysis with the MHC I molecule.

Protein-protein Docking analyses suggested that the epitopes exhibit a strong binding affinity for MHC I

Molecular docking analyses were conducted using the ZDOCK tool of Discovery Studio to predict binding of five T-cell epitope peptides with MHC I molecule, PDB:7RTD. The refinement of the selected docked poses was performed using the CHARMM energy minimization algorithm of the RDOCK program to eliminate non-bond clashes and determine the best-docked pose based on RDOCK score. The binding energy (BE), ΔG, and dissociation constant (Kd) of the best-docked poses of the five epitopes were estimated. Upon detailed examinations of the docking scores and binding energy values, it was revealed that T-Ep3 and T-Ep4 exhibited the most favorable properties, demonstrating the lowest BE and Kd values for the docked complex, alongside the highest docking score and highest number of intermolecular interactions (Fig. 1; Table 5). During the docking process, T-Ep3 exhibited 20 non-bond interactions with MHC I with ZDock score and RDock scores calculated as 11.86 and −9.99, respectively. The binding energy of the docked T-Ep3-MHC I complex was determined to be −10.3 Kcal/mol, and the dissociation constant (Kd) was found to be 2.9e-08 M. T-Ep4 also demonstrated comparable values for ZDock score (11.2), RDock score (−9.5), binding energy (−10.3 Kcal/mol),

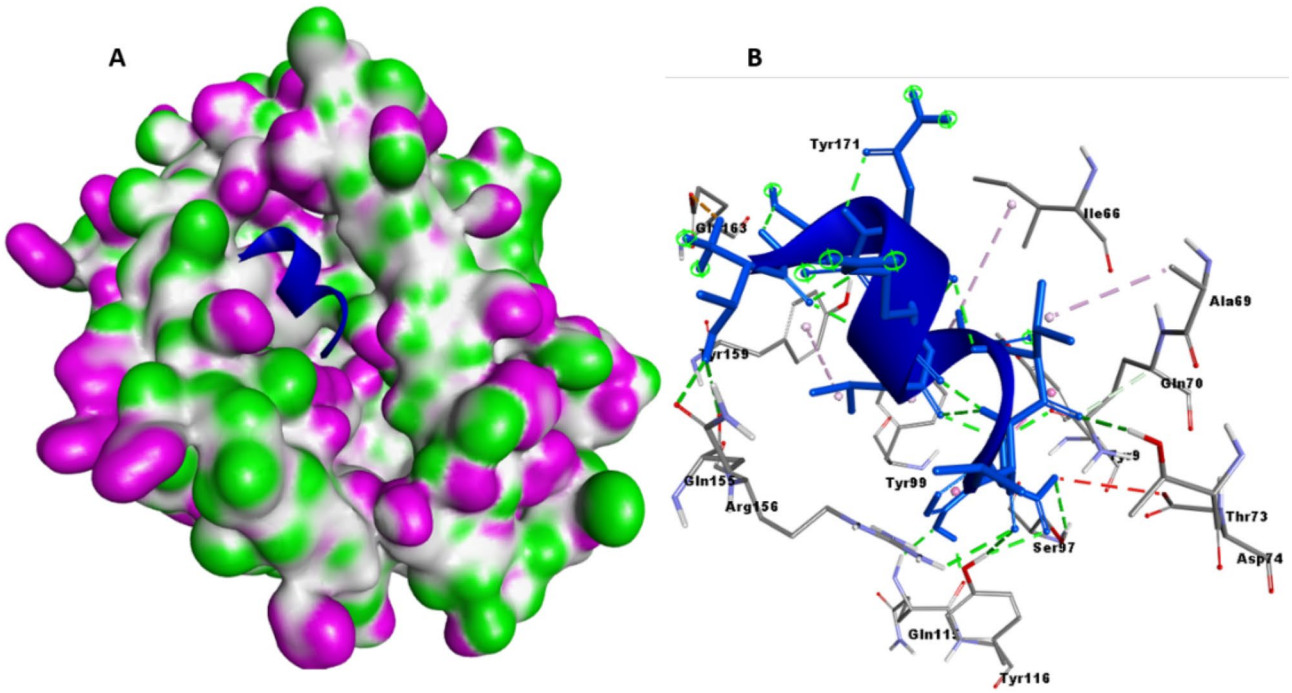


Fig. 1. Docking of T-Ep4 with MHC I molecule. (**A**) T-Ep4 bound to the MHC I major groove. (**B**) Intermolecular interactions of T-Ep4 and MHC I molecule.

T-cell Epitope	Zdock Score	Rdock	Binding Energy (ΔG) Kcal/mol	Kd (M)	No. of Inter-molecular contacts	Types of Interactions
T-Ep1	10.1	−4.24	−9.7	7.3e-08	15	SB(1),HB(6 + 1)HI(6)
T-Ep2	12.36	−3.552	−10.4	2.5e-08	13	HB(6),EI(1),HI(6)
T-Ep3	11.86	−9.99	−10.3	2.9e-08	20	EI(8),HB(9 + 1),HI(8)
T-Ep4	11.2	−9.529	−10.3	2.7e-08	17	SB(1),HB(10 + 1),HI(5)
T-Ep5	9.96	−4.630	−8.3	7.5e-07	13	EI(1),HB(5),HI(7)

Table 5. Summarized findings from the Docking studies illustrating the interaction between the epitopes and MHC I. *T-Ep: T-cell Epitope, Kd: Dissociation constant, EI: Electrostatic interaction, SB: Salt bridge, HB: Hydrogen bond, HI: Hydrophobic interaction.

and dissociation constant (Kd = 2.7e-08). It formed 17 intermolecular contacts with the MHC I molecule (Fig. 1; Table 5). The remaining epitopes also showed significant interaction with the MHC I molecule. Moreover, it was observed that the region of T-Ep3 spanning from V687 to Y695 is situated immediately after the S1-S2 cleavage site, R685-S686 of the S protein. The SARS-CoV-2 S protein is distinct from other coronaviruses due to an insertion at the cleavage site, especially in the sequence ^{680}SPRRAR↓SV^{687}, resulting in a cleavage motif RxxR recognized by proteases such as furin or TMPRSS2. This efficient cleavage at the S1/S2 site is essential for the virus to enter the host target cells[21,22].

Molecular docking studies using a structure-based approach revealed the efficacy of predicted T-cell epitopes in binding with MHC I molecules. Two predicted T-cell epitopes showed maximum binding capacity with the MHC I molecule, with one epitope located close to the superantigen region of the S protein of SARS-CoV-2, specifically at the site where S1-S2 cleavage occurs[23]. These findings emphasize the potential of the predicted T-cell epitopes in designing effective immunotherapeutic strategies.

Molecular dynamics simulation studies identified a stable conformation in the peptide-MHC I complex

A 100 nanosecond Molecular Dynamics (MD) simulation was conducted to evaluate the stability and flexibility of the T-Ep4-MHC I complex. Trajectory files derived from the simulation were analyzed using the Root Mean Square Deviation (RMSD) to assess the deviation between all C-alpha atoms of each conformation and the reference conformation. The RMSD analysis indicated that the complex achieved a steady state within the initial 3 nanoseconds and maintained stability until 58 nanoseconds, which some minor fluctuations. The average RMSD value during this period was 5.1 Å, which subsequently increased to 7.1 Å until the completion of the 100 nanoseconds simulation (Fig. 2).

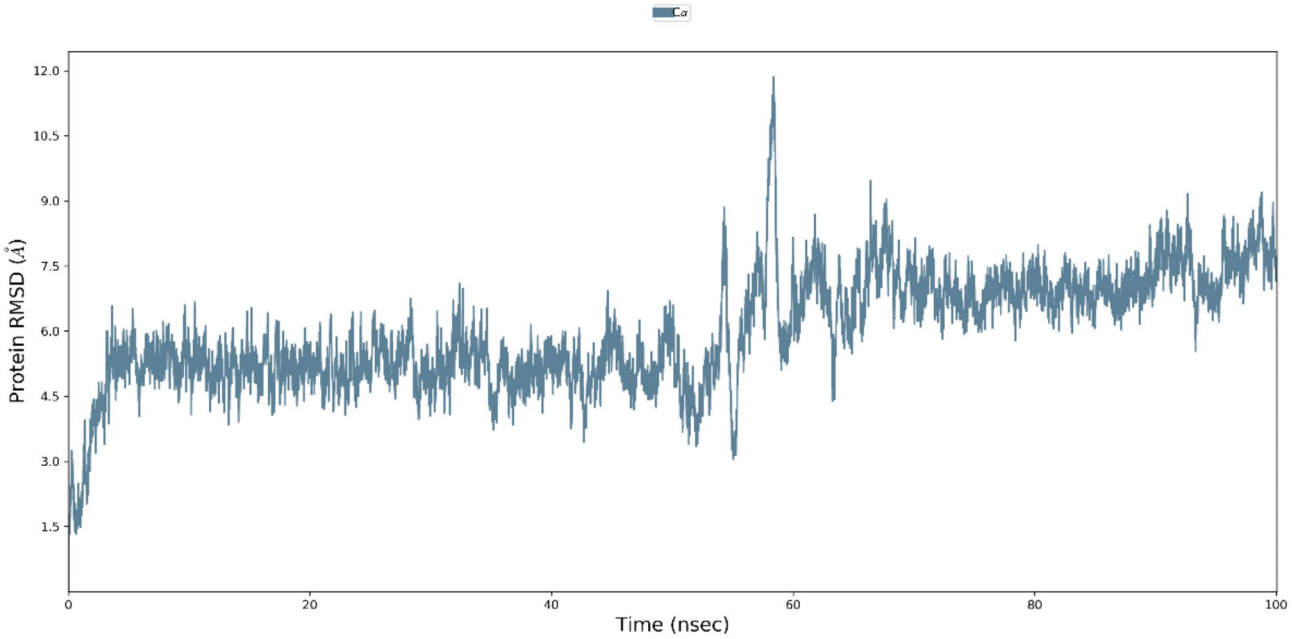


Fig. 2. Creation of RMSD plot to analyze the stability of the T-Ep4-MHC I complex after molecular dynamics (MD) simulation. The RMSD values, in Å, of the C-atoms of each MD-generated conformation in comparison to the reference conformation is plotted over a period of 100 ns.

An analysis of the conformations generated of each peptide-MHC I complex revealed that the molecules remained bound throughout the simulation under physiological mimicking conditions, suggesting potential effectiveness in stimulating T-cells by sustaining MHC I binding. Examination of the docked complexes indicated no significant loss of conformation, with only minor fluctuations observed in loop regions. This discovery holds significant promise and may offer valuable insights for design and development of epitope-based vaccines.

Selection of vaccine subunits to construct a multi-epitope vaccine sequence

The vaccine subunit's antigenic components were chosen based on their non-allergenic and non-toxic properties, their capacity to elicit immune responses, and their proximity to known epitopes.

It is noted that T-Ep1 reveals a strong MHC class I binding and CTL prediction score of 0.99, coupled with a low IC50 value of 2.78nM for the HLA-C*12:03 allele. Moreover, it demonstrates a significant binding affinity with HLA-B*35:01 and HLA-A*01:01 within the ^{865}LTDEMIAQY^{873} region. Adjacent to T-Ep1, the highly ranked CTL epitope, T-Ep2 ^{896}IPFAMQMAY^{904}, also demonstrated a score of 0.99. In the specified sequential region, an MHC class II binding HTL epitope, ^{886}WTFGAGAALQIPFAM^{900}, has been predicted with a favorable percentile rank of 1.5 and a lower IC50 value of 66.8nM. The close proximity of HTL epitopes suggests their potential as candidates for inclusion in a multiepitope vaccine. The proposed vaccine subunit covers T-Ep1, T-Ep2, and the HTL epitope, denoted as 'S1' and represented by the amino acid sequence LLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM.

In our analysis, T-Ep4 has demonstrated significant potential, achieving a high score of 0.97 in CTL prediction and a low IC50 value of 40.59nM for HLA-C*12:03. Notably, it exhibits strong binding affinity for HLA-B*35:01 and HLA-A*01:01 in the ^{604}TSNQVAVLY^{612} region. Furthermore, we have identified an HTL ^{602}TNTSNQVAVLYQDVN^{616} with a commendable percentile rank (0.64) and a superior prediction score (0.20) in the same region, The amino acid sequences 604–612 display overlapping epitopes for MHC class I & II alleles, indicative of potential antigen presentation via both pathways. Our analysis also revealed a predicted linear B-cell epitope in the overlapping region ^{616}NCTEVPVAIHADQLTPTWRVYSTGSNVFQ^{644}. As a result of these findings, we propose these CTL, HTL, and B-cell epitopes identified as strong candidates for inclusion as multiepitope vaccine subunits. Upon integrating these findings, we suggest that the peptide ^{604}TSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQ^{644}, referred to as 'S2', has the potential to elicit T-cell and a B-cell mediated responses or both, making it a promising vaccine candidate.

The peptide, T-Ep6 ^{440}NLDSKVGGNYNYLY^{453} exhibited a high percentile rank in both the Artificial Neural Network (ANN) and Stabilized Matrix Method (SMM) analyses, with values of 0.05 and a low IC50 value of 22.56 nM to HLA-A*01:01 independently (Table 3). The predicted peptide for HTL response is situated in the nearby region ^{464}FERDISTEIYQAGST^{478} with a percentile rank of 1.8 and prediction score of 0.12 (Table 4). Furthermore, linear and conformational B-cell epitopes, B-Ep3 have been identified within the region covering the amino acid sequence, ^{440}NLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTN^{502} (Table 1). The B-Ep3 has a favorable PI score of 0.67 (Table 2). Of note, the Uniport database confirms the presence of an immunodominant HLA epitope region ^{448}NYNYLYRLF^{456} recognized by CD8^{+} cells, consistent with our initial prediction. Based on these findings, it is hypothesized that the proposed peptide

[440]NLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTN[502], referred to as 'S3' has the potential to elicit a T-cell mediated or B-cell mediated immune response, or both.

The T-Ep3, the CTL epitope anticipated within the [687]VASQSIIAY[695] region demonstrated a notably high prediction score of 0.97 with a percentile rank of 0.01 (Table 3). It demonstrated a strong affinity with an IC50 value of 10.63 to HLA-B*35:01 and HLA-A*01:01. Additionally, an HTL epitope, [684]ARSVASQSIIAYTMS[697] was identified in the same region using the consensus 2.22 IBDB prediction method (Table 4). This epitope presented a favorable percentile rank of 4.2 and an IC50 value of 135nM. Furthermore, a linear B-cell epitope [672]ASYQTQTNSPRRARSVASQ[690] has been forecasted in the aforementioned region (Table 1). Our study highlights the significance of T-Ep3 due to its spatial proximity to the previously recognized superantigen region, [681]PRRA[684]of the S-protein of SARS-CoV-2[23]. Found at the S1-S2 cleavage site, this region plays a pivotal role in initiating viral entry processes. Consequently, the peptide region [687]VASQSIIAYTMS[697] designated as 'S4' shows promise as a subunit for a promising vaccine construct.

Consequently, there are four proposed subunits for the vaccine construct, which are as follows:

S1: LLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM

S2: TSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQ

S3: NLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTN

S4: VASQSIIAYTMS

Population coverage analysis

Considering the high global prevalence of SARS-CoV-2 infection, it is essential for vaccine epitopes to cover a wide range of allele populations. Our examination of population coverage for the CTL, utilizing the IEDB population coverage tool, revealed that each CTL exhibited coverage of 71%, 81%, 89%, and 90% for the Indian, Chinese, North American, and global populations, respectively, based on a specific set of HLA-A, HLA-B, and HLA-C alleles. Detailed information can be found in Supplementary Tables 7 and Table 8.

Design and characterization of multi-epitope vaccine construct

The vaccine construct was created utilizing subunits S1-S4, which were devised based on the overlapping regions within CTL, HTL, and B-cell epitopes, as well as existing literature. Human β-defensins were employed as the vaccine adjuvant at the N and C terminals, with the subunits interconnected by GPGPG linkers. Specifically, human Beta-defensin 2 (hBD-2) (PDB ID: 1FD3 sequence:

GIGDPVTCLKSGAICHPVFCPRRYKQIGTCGLPGTKCCKKP)

and the human Beta-defensin 3 (hBD-3) (PDB ID: 1KJ6, sequence:

GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRKK) were utilized in this context as adjuvants at N and C terminals respectively using the linker EAAAK as illustrated in Fig. 3.

The vaccine sequence exhibits a theoretical isoelectric point (PI) of 9.19, a molecular weight of 28.2 kDa, and an estimated half-life of 30 h in vitro for mammalian reticulocytes. With an aliphatic index of 69.39, it demonstrates significant thermostability, while the instability index (II) of 32.53 categorizes the sequence as stable. Additionally, the vaccine sequence showcases a negative grand average of hydropathicity (GRAVY) of −0.22, thereby fulfilling the essential physicochemical criteria for a promising vaccine candidate.

Antigenicity, allergenicity, and toxicity of the designed vaccine construct

Upon analysis, the vaccine construct demonstrated antigenicity with a prediction score of 0.44 when assessed using the VaxiJen v2.0 tool, indicating its potential to provoke an immune response. Subsequent evaluations utilizing the AllerTop v2.0 and AllergenFPv.1.0 servers confirmed the non-allergenic properties of the vaccine construct. Furthermore, the toxin prediction tool, ToxinPred2, substantiated the non-toxic nature of the vaccine construct, yielding a machine-learning prediction score of 0.70, surpassing the established threshold of 0.60.

Cytokine-inducing ability of the MEVC

The PIP-EL server identified the construct as proinflammatory, yielding a score of 0.73. Conversely, the Pre-AIP server classified it as anti-inflammatory with a high confidence AIP score of 0.55. Moreover, the IL-10Pred server projected the construct to be IL-10-inducing with an SVM score of 0.84. Additionally, the IL4pred server indicated that nearly the entire construct region possesses IL-4-inducing ability. Also, the IFNepitope2 server predicted most of the MEVC sequence as IFN-gamma-inducing. A detailed graphical representation of IL4pred server and IFNepitope2 result is provided in Supplementary Fig. 5A-B. Given the accuracy of the three-dimensional structure in estimating vaccine construct design efficiency, we have modeled the construct for further studies.

Structure prediction and validation of designed vaccine construct

The 3-D structure of the designed construct was modeled using the trRosetta webserver. Among the refined models, the model with the highest TM score (0.335) was chosen and further refined using ModRefiner and GalaxyWeb. The final model displayed a favorable 95.4% region in the Ramachandran plot, representing the best

Fig. 3. The in silico designed vaccine construct MEVC: The S1-S4 subunits are linked by GPGPG linkers and the adjuvants Human Beta-defensin 2 and 3 attached at the N and C terminals using the EAAAK linkers.

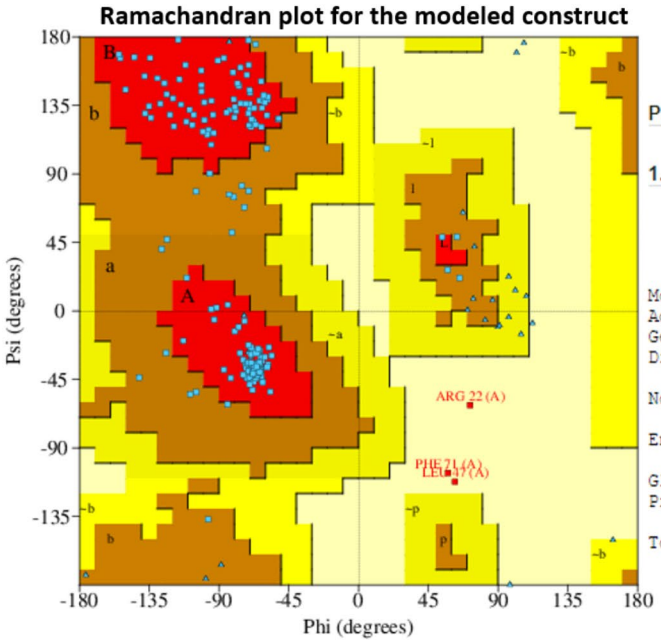
score among all refined models (Supplementary Fig. 6). Additionally, the TM-score was measured at 0.9261, RMSD at 0.27, MolProbity at 1.8, with a minimal clash score of 11.1 and no poor rotamers (Supplementary Fig. 6). The stereochemical quality of the refined structure was validated using PROCHECK in the PDBSum. Notably in Ramachandran plot statistics, 90% of the residues were found within the 'Most Favored regions,' 8.5% within the 'Additionally allowed regions,' and only 1.4% (3 residues) fell in the 'Disallowed regions' (Fig. 4). These results affirm the decent quality of the model, and the conformationally stable models were employed for subsequent binding analysis with the TLR molecules[24].

Molecular Docking of vaccine construct with host immune receptor

The binding efficiency of the designed construct with TLR-3 and TLR-4 was assessed using the ClusPro server. Analysis revealed that the highest number of docking poses clustered with the lowest energy-weighted score of −1303.2for TLR-3 and −1133.6 for TLR-4, surpassing previous studies in a similar domain[25]. The docked complex model of MEVC-TLR-3 was examined in Discovery Studio's molecular viewer to observe intermolecular interactions (Fig. 5A) and further analyzed in a closer view (Fig. 5B). Intermolecular interactions analysis using the Prot-Prot analysis of PDBSum estimated 11 salt bridges, 13 hydrogen bonds, and 196 non-bond contacts between MEVC and TLR-3, indicating robust binding efficiency (Fig. 5C). The same analysis was performed to examine MEVC-TLR-4 interactions, and the server computed 6 salt bridges, 11 hydrogen bonds, and 190 non-bond contacts between MEVC and TLR-4 (Fig. 5D-F). The PRODIGY server (https://bianca.science.uu.nl/prodigy/)[26] predicted a Kd constant of 1.7e-12 M and 1.3e-11 for MEVC-TLR-3 and MEVC-TLR4 complex, respectively. The binding free energy (ΔG) of −16.0 Kcal/mol and − 14.8 Kcal/mol were computed for the MEVC-TLR-3 and MEVC-TLR4 complexes, respectively, highlighting the strong binding between the two molecules in both complexes. We also performed docking of MEVC with TLR4-dimer, and the docked complex (lowest energy-weighted score of −1060.6) was bound with an estimated Kd value of 2.3e-13 M and BE value of −17.3Kcal/mol. Detailed examination showed that MEVC formed many contacts with TLR-4 dimer Chain A (1 salt bridge, 2 hydrogen bonds, and 64 non-bonded) and Chain B (6 salt bridges, 13 hydrogen bonds, and 148 non-bonded contacts), indicating strong binding efficiency of MEVC with TLR-4 also. Results are provided as supplementary details (Supplementary Fig. 7). The calculated binding energy value, docking score, and the number of intermolecular contacts collectively suggest that the engineered peptide construct, MEVC, exhibits a strong binding affinity with the TLR-3 and TLR-4 molecules. This finding underscores the efficacy of the construct in its intended function.

Dynamic simulation of MEVC-TLR-3 complex

Dynamic simulation study of the MEVC-TLR-3 complex was performed using the iMODS online server to assess and measure the proteins' flexibility. From the analysis reports provided as supplementary information (Supplementary Fig. 8), the deformability peaks in the graph of the vaccine-TLR3 complex indicate flexible regions in the protein (Supplementary Fig. 8a). The eigenvalue of the vaccine-receptor complex is 5.399225e-06, representing the protein's stiffness (Supplementary Fig. 8b). Variance and covariance graphs show associated



PROCHECK statistics

1. Ramachandran Plot statistics

		No. of residues	%-tage
Most favoured regions	[A,B,L]	191	90.1%
Additional allowed regions	[a,b,l,p]	18	8.5%
Generously allowed regions	[~a,~b,~l,~p]	0	0.0%
Disallowed regions	[XX]	3	1.4%*
Non-glycine and non-proline residues		212	100.0%
End-residues (excl. Gly and Pro)		1	
Glycine residues		31	
Proline residues		19	
Total number of residues		263	

Fig. 4. Ramachandran plot statistics for the modeled and refined MEVC: 90% of the total residues are in the most favored regions and 8.5% in the additionally allowed regions. Only 3 residues fall in the disallowed region.

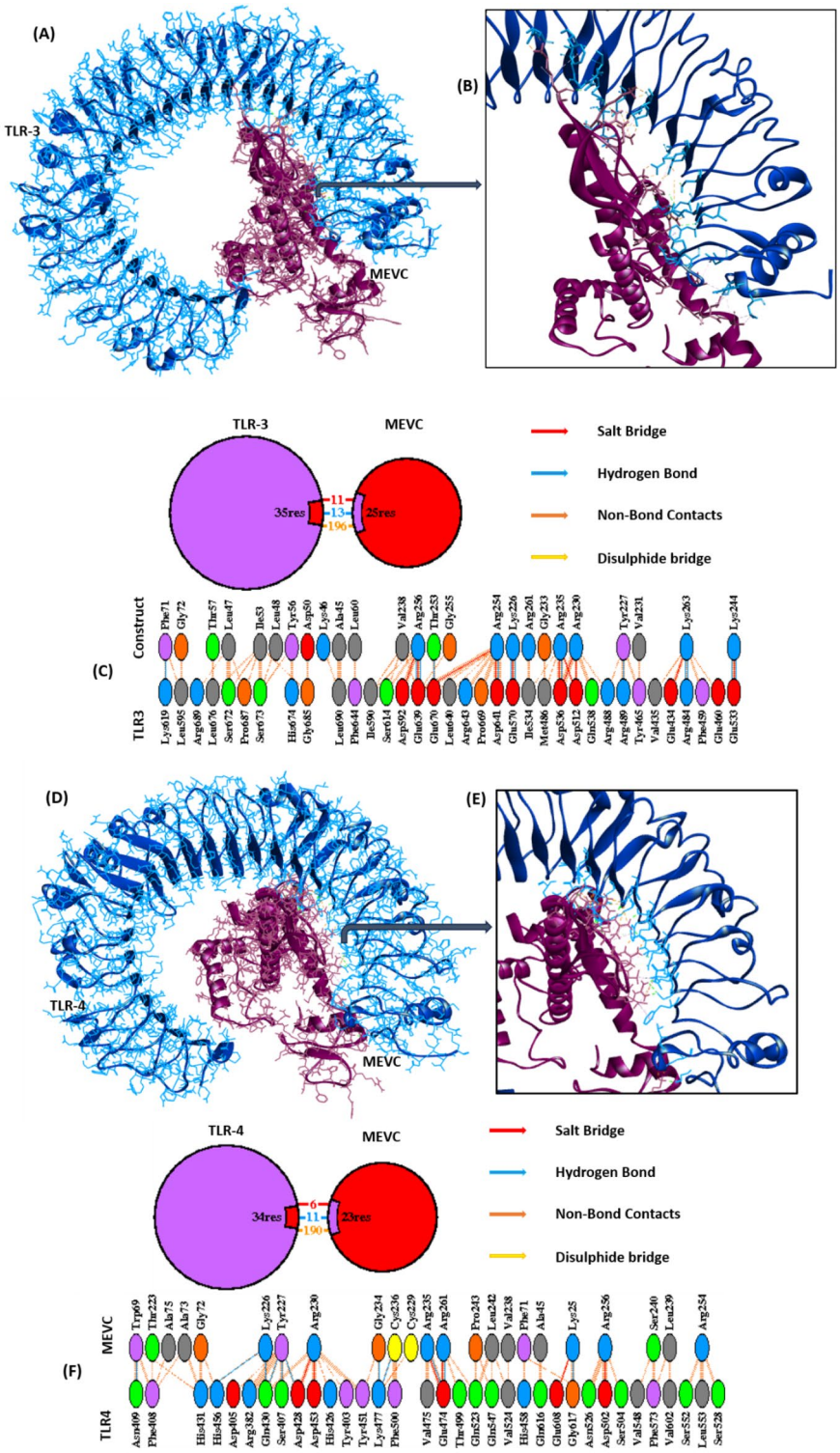


Fig. 5. The docked complex model of the MEVC-TLR-3 and MEVC-TLR-4. (**A** and **D**) The docked complex MEVC-TLR-3 and MEVC-TLR-4-model as examined in the molecular viewer of Discovery Studio, (**B** and **E**) A closer view of the MEVC-TLR-3and MEVC-TLR-4complex, (**C** and **F**) Intermolecular interactions between MEVC and TLR-3 and MEVC and TLR-4 in the docked complex structure of Prot-Prot Analysis.

amino acids and rigid regions in the complex (Supplementary Fig. 8c). The elastic map illustrates the connections between atoms and highlights rigid regions in the complex (Supplementary Fig. 8d).

An all-atom Molecular Dynamics (MD) simulation was conducted for 50 ns to evaluate the stability and flexibility of the MEVC-TLR-3 complex. Trajectory files derived from the simulation were analyzed for the

Root Mean Square Deviation (RMSD) to assess the deviation between all C-alpha atoms of each conformation and reference starting conformation. The C-alpha RMSD was computed for MEVC-TLR-3complex to examine the deviation in conformation during the simulation period. The RMSD analysis indicated that the complex gradually achieved a steady state trend in the initial 15 nanoseconds and maintained a stabilized nature until the 50 nanoseconds, with very minor fluctuations. The average RMSD value during this period was 7.5 Å, subsequently increasing to 8.5 Å until the completion of the 50 ns simulation (Fig. 6A), which indicates that the binding of MEVC-TLR-3 is strong enough and the complex remains in the stable bound state. The number of hydrogen bonds formed between TLR-3 and MEVC is computed throughout the simulation period and plotted (Fig. 6B). It was found that an average of 20 intermolecular H bonds were formed in the MEVC-TLR3 complex which indicates how strongly the molecules are bound to each other throughout the simulation period. The binding free energy (ΔG) for the last generated MEVC-TLR-3 conformation was computed as −15.8 Kcal/mol and the dissociation constant, Kd as 2.6 e-12 M which is almost the same as that of the initial conformation (BE= −16.0 Kcal/mol and Kd = 1.7e-12 M) indicating that molecules remained strongly bound even after the 50 nano

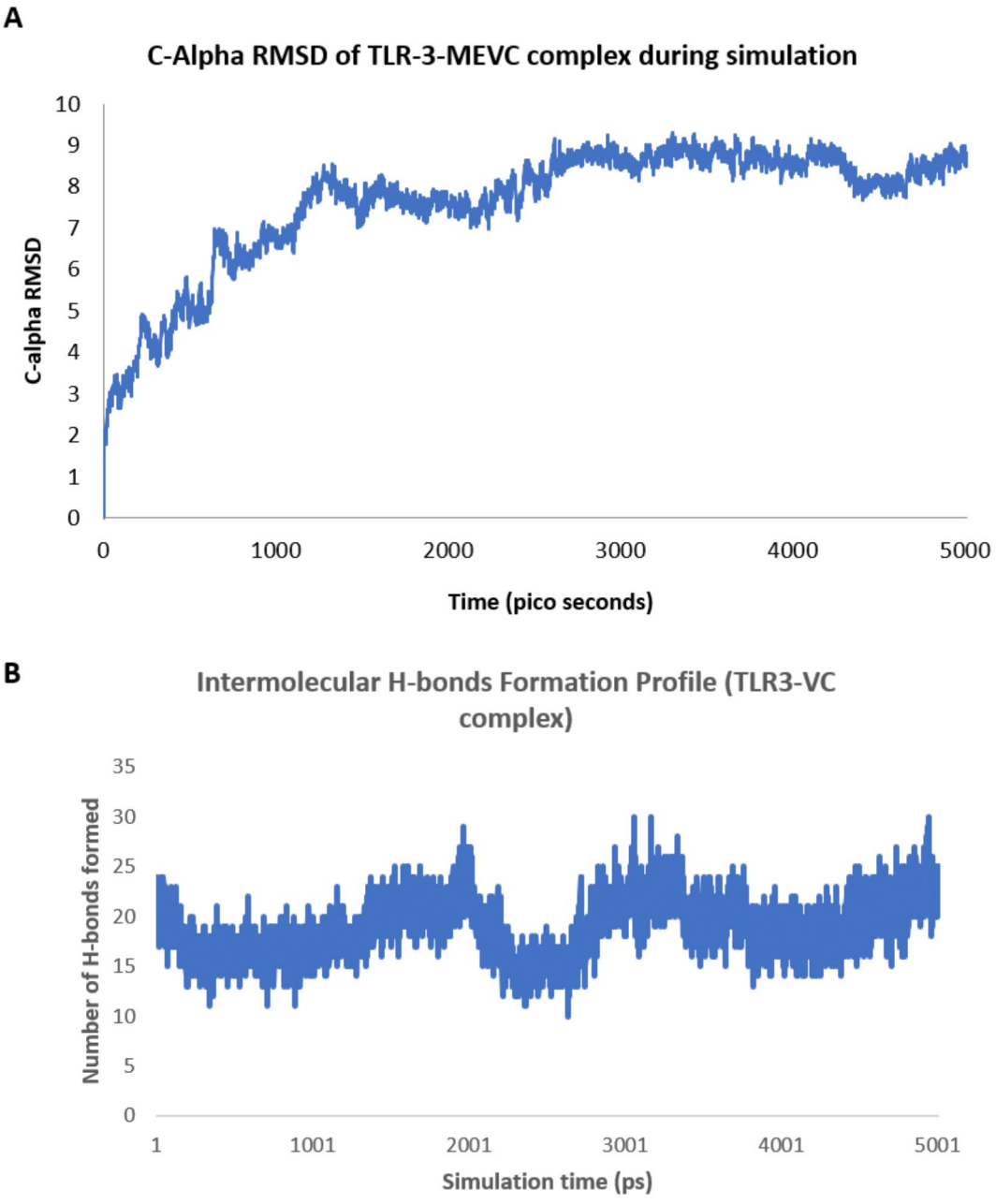


Fig. 6. C-alpha RMSD plot and Intermolecular H-bond profile of MEVC-TLR3 complex. (**A**) C-alpha RMSD plot for 50 ns simulation period to analyze the stability of the MEVC-TLR-3 complex. The RMSD values, in Å, of the C-atoms of each MD-generated conformation in comparison to the reference conformation is plotted over50 ns. (**B**) Hydrogen bond formation between the TLR-3 and MEVC during the simulation period.

seconds of simulation time. Molecular Dynamics (MD) simulation of MEVC-TLR-3 complex revealed that the molecules remained strongly bound during the simulation, indicating the potential effectiveness of our vaccine construct in TLR-3-mediated immune response.

Codon adaptation and in Silico cloning of the vaccine construct

The Java Codon Adaptation Tool was employed to optimize the codon usage of the vaccine, resulting in a 789-nucleotide sequence with a Codon Adaptation Index (CAI) of 1.0 and a GC content of 53.73%. These findings suggest the potential for expression of the vaccine construct in the host cell, E. coli. The CAI value indicates the likelihood of cDNA expression in a specific expression system. To facilitate cloning, XhoI and NotI restriction sequences were incorporated at the N and C termini of the optimized codon sequence. Furthermore, the optimized sequence was integrated into the pET-28(+) vector using Serial Cloner 2.6 software for in silico cloning (Fig. 7). This operation successfully inserted the optimized codon sequence of the vaccine construct into the plasmid following cleavage at the restriction enzyme sites (Fig. 7).

In Silico immune simulation studies in response to our vaccine injection

The designed vaccine construct was subjected to in silico testing to assess its impact on the immune system's response. The simulation conducted using the C-IMMSIM server indicated that upon administration of the

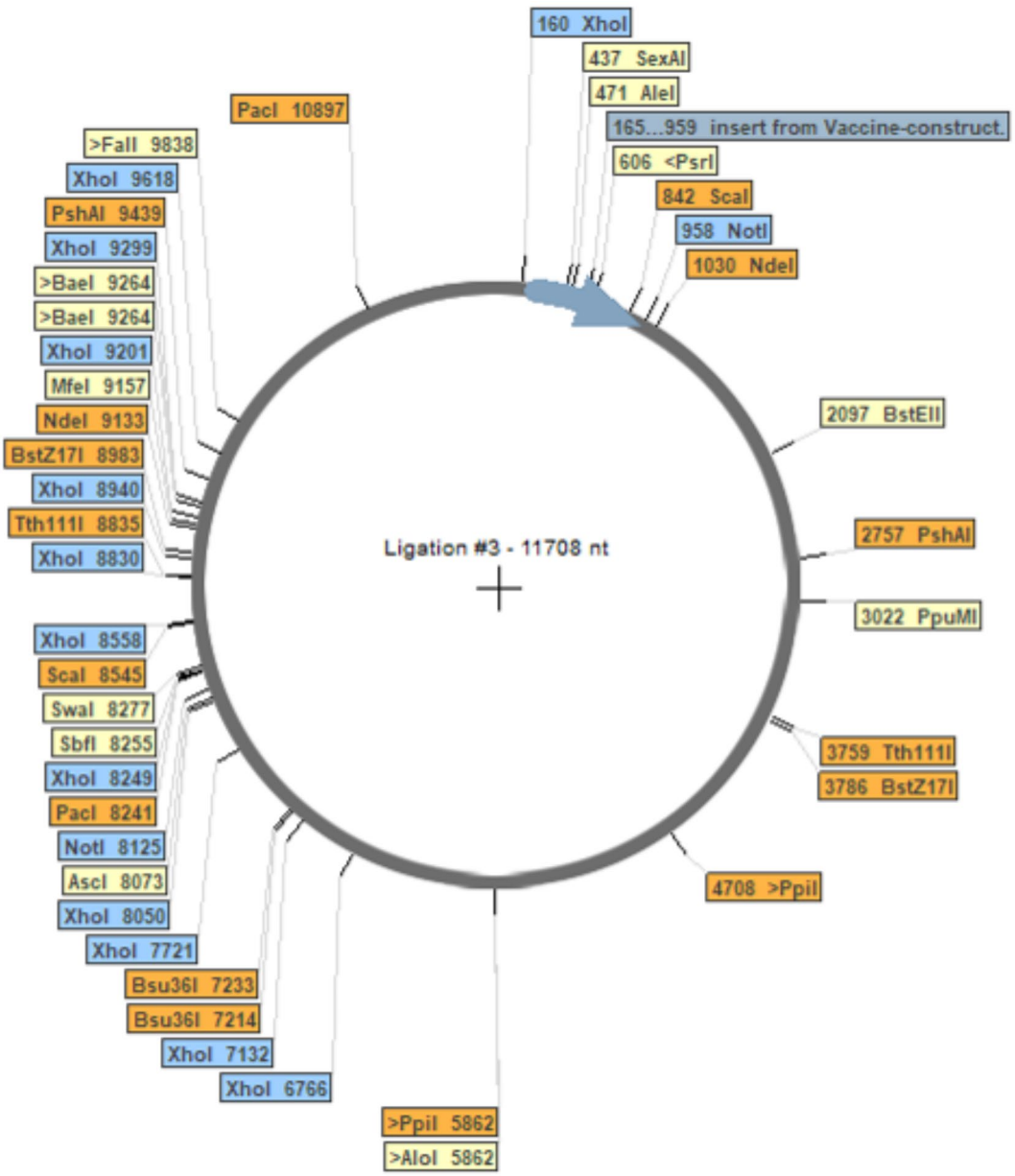


Fig. 7. In silico cloned vaccine construct: The 789 nucleotide length codon sequence of the insert has been ligated to the pET-28(+) vector at the 165–959 position following XhoI and NotI restriction enzyme digestion.

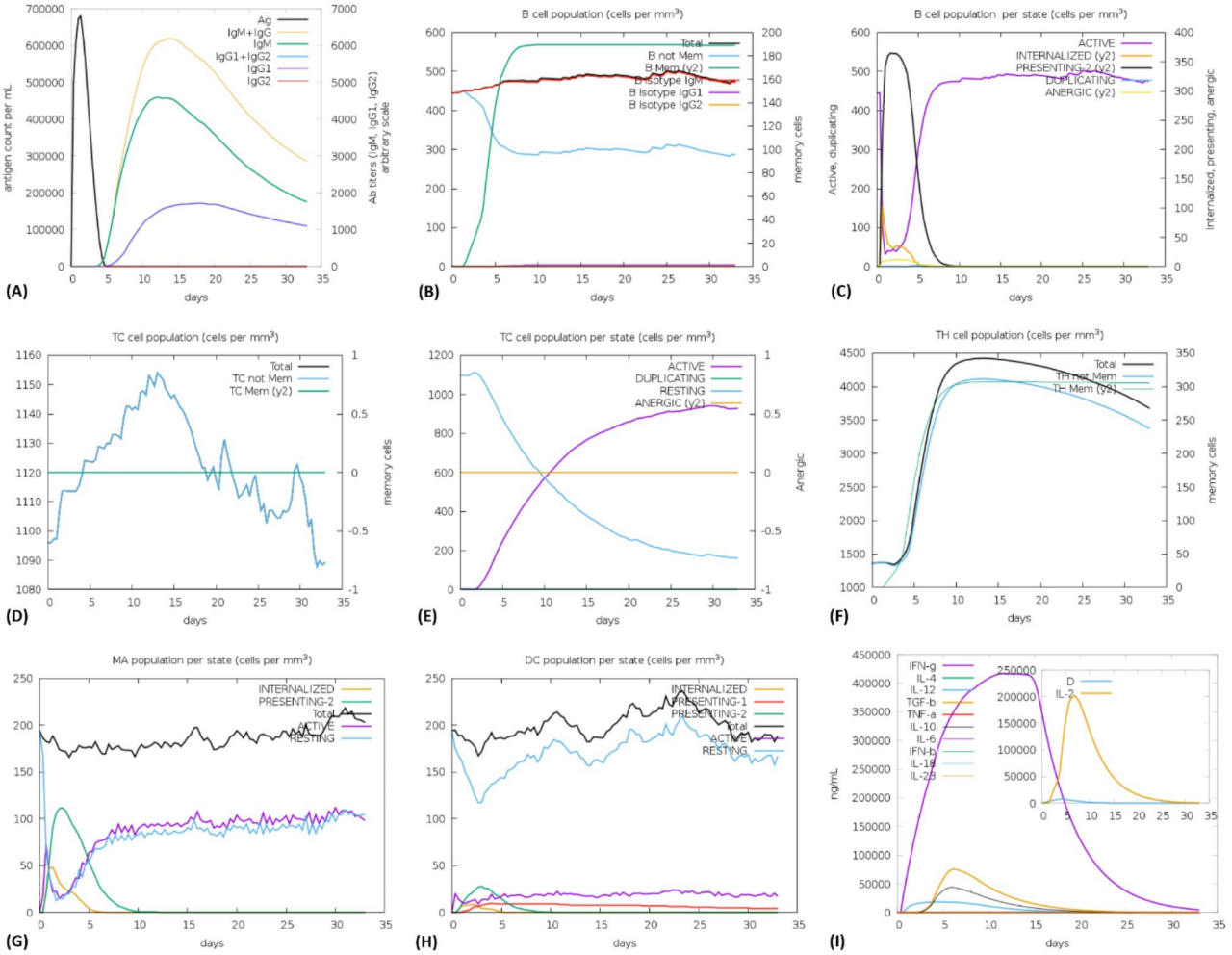


Fig. 8. In silico simulation of immune response after the administration of the vaccine as an antigen. (**A**) levels of antigen and immunoglobins. (**B**) density of B-cell population. (**C**) density of B-cell population per state. (**D**) density of cytotoxic T-cell population. (**E**) density of cytotoxic T-cell population per state. (**F**) density of helper T-cell population. (**G**) density of macrophages population per state. (**H**) density of dendritic cell population per state. (**I**) level of cytokine production.

vaccine construct, the initial response showed elevated levels of IgM + IgG. Subsequently, the secondary response displayed increased levels of IgM, IgG1, and B-cell populations within the 10–15-day period (Fig. 8A-C). It is noteworthy that the antigen level (680000 counts per mL) decreased significantly by the 3rd day following the injection. Moreover, the B-cell response demonstrated a logarithmic growth phase from day 3 to day 7, reaching a peak level of approximately 600 cells per mm3. This underscores the vaccine construct's ability to stimulate substantial levels of B cells that produce IgM and IgG subclass antibodies. Both CTL and HTL elicited a robust immune response after a brief exposure, indicating the immunogenic nature of the T cell epitopes integrated into the vaccine construct (Fig. 8D-F). Additionally, the populations of macrophages and dendritic cells were observed to increase post-exposure (Fig. 8G-H). Our research also revealed a sharp increase in the production of pro-inflammatory cytokines such as IFN-γ, IL-2, and IL-12 post-vaccination (Fig. 8I). Concurrently, peaks in certain anti-inflammatory cytokines such as TGF-β and IL-10 were noted, indicating a reliable and crucial immune-triggering response upon the administration of the vaccine construct (Fig. 8H-I).

Discussion

The current global health emergency induced by the COVID-19 pandemic has underscored the critical necessity for widespread vaccination and the development of effective antiviral treatments. Significant concerns exist regarding the efficacy of antiviral medications, particularly in individuals with compromised immune systems, and the potential risks associated with specific vaccination approaches. The progress in immuno-informatics presents a hopeful prospect for adeptly managing and mitigating the transmission of newly identified viruses. Further research is indispensable for the assessment of both existing and innovative SARS-CoV-2 vaccines against emerging variants, as well as for bolstering the formulation of resilient strategies to counter the global health crisis.

In this study, we have identified potential immunogenic targets within the Spike (S) protein of SARS-CoV-2 and devised a strategy for creating an effective vaccine to combat its variants. Utilizing immunoinformatics tools, we have identified B-cell and T-cell epitopes within the spike protein, with the objective of developing a vaccine that is effective against all variants of SARS-CoV-2. Our analysis has revealed the presence of 13 linear B-cell epitopes and 5 discontinuous or conformational epitopes, which have the capacity to elicit a humoral response against SARS-CoV-2. Notably, we have observed common residues with a high probability of being part of the B-cell epitopes.

Our research focused on the prediction of T-cell epitopes, with an emphasis on their superior reliability in comparison to B-cell epitopes. T-cell epitopes play a crucial role in the development of pathogen-specific vaccines, with antigenicity and binding affinity to MHC molecules being pivotal determinants. To ensure accuracy, we employed a consensus approach, identifying epitopes predicted by multiple servers. These epitopes were then prioritized based on various parameters, including antigenicity scores, conservation, affinity for multiple alleles, broad population coverage, and molecular interaction with HLA alleles. Utilizing a data-driven approach, we successfully identified nine CTL epitopes, each comprised of 9–10 residues. Computational modeling facilitated the prediction of the efficacy of two potent epitopes in binding to MHC I and stimulating T-cells. We rigorously assessed the binding energy to validate the effectiveness of the epitopes. Notably, one of the top two T-cell epitopes identified aligns with the previously reported super-antigenic region of the spike protein of SARS-CoV-2[22]. This significant convergence between our predicted epitopes and existing reports enhances the precision of our epitope prediction. Molecular interaction studies indicated that the top two epitopes, T-Ep3 and T-Ep4, exhibit high binding affinity. Furthermore, simulation studies confirmed the stability of the peptide-MHC I complex. We also utilized toxicity prediction tools to verify that these predicted T-cell epitopes are non-toxic, non-allergenic, and highly antigenic. Our findings hold significant promise for advancing the development of novel T-cell-based vaccines and immunotherapies against SARS-CoV-2.

In the second phase of the investigation, a new epitope-based vaccine construct named MEVC was proposed through prediction from the Spike protein. While prior research has predominantly focused on structural and non-structural proteins[17,18], a limited number of studies have emphasized the spike protein due to its crucial role and prominent exposure on the virus surface. The construct's design leveraged prospective T-cell and B-cell prediction data to formulate four subunits characterized by high immunogenicity, with the objective of eliciting both innate and adaptive immunity. These subunits were formulated by integrating overlapping high-scoring B-cell, CTL, and HTL epitopes, supplemented by existing literature data. This represents the inaugural endeavor in this direction. Evaluation of population coverage unveiled global potential, with extensive coverage in North America, India, China, and various other countries. The resulting vaccine encompasses 263 amino acid residues, with a molecular weight of 28.2 kDa. The vaccine demonstrates stability, hydrophilicity, and basic nature, with an estimated half-life of 30 h. These findings correspond to previously documented vaccine constructs for SARS-CoV-2[25]. Notably, the designed vaccine construct exhibits characteristics indicative of antigenicity, non-allergenicity, and non-toxicity based on the analysis. Moreover, optimization and validation of the 3D model of the vaccine construct confirm its high quality, with over 90% of residues falling within the favored region on the Ramachandran plot. Consequently, the designed vaccine construct fulfills all requisite criteria and stands as a promising multiepitope vaccine candidate[25].

Our study focused on Toll-like receptor 3 (TLR3) due to its potential to trigger a robust and immediate immune response, which is essential in the context of rapidly evolving SARS-CoV-2 variants, and was complemented by a related investigation using Calu-3/MRC-5 multicellular spheroids (MTCSs), which showed significant upregulation of TLR3-associated interferon regulatory factor 3 (IRF3) expression within 24 h post-infection, along with increased levels of pro-inflammatory cytokines such as IL-1α, IL-1β, IL-4, IL-6, and type I interferons (IFN-α and IFN-β), highlighting the critical role of TLR3 in the early immune response to SARS-CoV-2 infection[27]. The present study opens avenues for a more comprehensive and extensive investigation of the synergistic effects of various other receptors, like TLR7 and 8, on the MEVC-mediated immune response. The results obtained from the molecular docking analysis reveal a high binding affinity between the MEVC and the TLR-3 and TLR-4 molecule. This suggests that the construct has the potential to function as a marker for identifying molecular patterns of pathogens and subsequently triggering an immune response. The computed lowest energy score of the MEVC-TLR-3 complex, which is −1303.2, exceeds the previously reported value[25] indicating the stability of the MEVC-TLR-3 complex. Furthermore, the deformability and Eigenvalue assessment was also performed as higher eigenvalues designate the rigidity of protein-protein complexes. The eigenvalue predicted for the MEVC-TLR-3 complex in this study was 5.399225e-06 which supports the verity of integrity of the MEVC-TLR-3 complex. All-atom MD simulation of MEVC-TLR-3 also revealed that the molecules remained bound throughout the simulation under physiological mimicking conditions, suggesting the potential effectiveness of our vaccine construct in TLR-3-medicated immune response against SARS-CoV-2 variants. These findings have great potential and may provide important insights for designing and developing vaccine constructs. We need further wet lab experiments to confirm the vaccine's binding affinity with TLR-3 and its interaction with CTL epitopes and HLA molecules.

Additionally, codon adaptation has significantly improved the expression of the vaccine candidate in the E. coli strain K12, as demonstrated by a substantial codon adaptation index and increased GC content, thus ensuring high levels of expression within the host cell. The in silico administration of the vaccine construct has led to an increase in memory related to B cells and T cells, alongside consistently elevated levels of IFN-gamma and various pro-inflammatory and anti-inflammatory cytokines, demonstrating the potential of our vaccine candidate to trigger an immune response effectively. Furthermore, the study has confirmed the absence of adverse effects such as cytokine storms, thus affirming the positive impact of the vaccine on the immune system. This systematic approach underscores our comprehensive efforts to design a safe and effective vaccine construct that addresses all essential requirements[27].

Our research presents a new approach to developing a multi-epitope vaccine against SARS-CoV-2 using immunoinformatic tools. Comprehensive safety and effectiveness assessments have demonstrated the vaccine's potential in eliciting an immune response. The vaccine is specifically designed to activate both humoral and cell-mediated immunity and offers strong support for accelerating vaccine development[28], subject to experimental validation.

Methods

SARS-CoV-2 Spike protein sequence retrieval

The S-protein sequences for the Indian isolates of SARS-CoV-2 were obtained from the NCBI protein database in FASTA format. The CoV server mutation app from the global data science initiative (GISAID) was used to detect and identify different variants of hCoV-19 documented worldwide[29]. Sequences corresponding to the Wuhan_hCOV19, Alpha (B.1.1.7), Beta (B.1.351), Gamma (B.1.1.28.1), Delta (B.1617.2), Omicron (B.1.1.529), Lambda (C.37), Mu (B.1.621) variants were sourced from the NCBI protein database. For the epitope prediction study, the latest variants of interest (VOI) of hCoV-19 until the start of 2024 were included. The sequences of BA2.86, XBB.1.16, XBB.1.5, EG-5.1, and JN.1.1 were unavailable in the NCBI protein database and were sourced from the GISAID.

Multiple sequence alignment of the SARS-CoV-2 Spike protein

In our study, we analyzed S protein sequences of the SARS-CoV-2 virus to identify conserved regions and used Clustal Omega and Mega 11 aligner software to execute this task. The raw data of the multiple sequence alignment were saved in the Mega format, which is listed in Supplementary Table 1. Identifying these conserved regions helped us to search for predicting B-cell and T-cell epitopes that trigger innate and adaptive immune responses using various immune-informatics software.

Prediction of linear B-cell epitopes

Understanding B-cell epitopes is crucial for comprehending how the immune system recognizes antigens. These precise regions trigger an immune response and bind to antibodies, leading to the activation of B-lymphocytes and the production of antibodies[30]. B-cell epitopes are classified into linear, discontinuous, and conformational epitopes based on their spatial orientation, and accurate prediction requires a thorough analysis of peptide sequence and structural information, which is vital for the development of vaccines and immunotherapies.

The NIAID, NIH, USA maintains the Immune Epitope Database (IEDB) (https://www.iedb.org/), which predicts linear B-cell epitopes based on factors like surface accessibility, hydrophilicity, and antigenicity. IEDB offers various B-cell prediction methods using different scales or parameters such as the Emini surface accessibility scale[31], Karplus and Schulz's flexibility prediction[32], the hydrophilicity prediction method[33], and Kolaskar and Tongaonkar's antigenicity scale[34]. These methods establish a relationship between physicochemical parameters and the location of continuous epitopes, including hydrophilicity, flexibility, accessibility, turns, exposed surface, polarity, and antigenic propensity of polypeptide chains. Propensity scales have been developed for each of the 20 amino acids to predict the location of continuous epitopes more accurately.

Prediction of discontinuous B-cell epitopes

This approach is grounded on the concept that protein surface residues are more likely to bind with antibodies. The SEPPA 3.0[35] model has been enhanced with logistic regression, proving effective in improving B-cell epitope prediction in glycoproteins. Utilizing the SEPPA 3.0 tool, we identified potential B-cell epitope residues on the SARS-CoV-2 S-protein. We conducted a comparison of results obtained from different tools to ascertain the most probable discontinuous B-cell epitopes.

In our investigation, we utilized DiscoTope 2.0[36], Ellipro[37], and SEPPA[38]to anticipate discontinuous B-cell epitopes from the SARS-CoV-2 spike protein. These tools discern epitopes based on the 3D coordinates of the protein's atoms. We acquired the 3D structure of the spike protein (PDB ID: 6xe1) from the Protein Data Bank, selecting it due to being the lengthiest protein solved using the X-ray technique. DiscoTope 2.0 incorporates amino acid statistics, spatial information, and surface exposure in its algorithm[36]. The propensity score is computed from epitope log-odds ratios and then summed according to their proximity in the antigen's three-dimensional structure. The DiscoTope algorithm explores a 10Å radial sphere around each residue to identify intramolecular contact residues, and the contact number predicted represents the number of C-atoms in the antigen within 10Å of the respective residue's C-α. The DiscoTope score is derived by subtracting the total number of residues within the sphere from the sum of the propensity scores of those contact residues. Higher scores compared to the threshold indicate a positive prediction. Ellipro, included in the IEDB Analysis Resource, differs from other B-cell epitope prediction methods in that it does not entail training[37]. Ellipro utilizes the Thornton method and a residue clustering algorithm to accurately forecast B-cell epitopes. This approach is based on the notion that protein surface residues have a higher likelihood of binding with antibodies. The SEPPA 3.0 model has undergone enhancement through logistic regression, which has been proven effective in enhancing B-cell epitope prediction in glycoproteins[35,38]. Using the SEPPA 3.0 tool, our analysis revealed potential B-cell epitope residues on the S-protein. The comparative evaluation of outcomes obtained from various tools ascertained the most likely discontinuous B-cell epitopes.

T-cell epitope prediction

An effective multi-epitope subunit vaccine should contain both CTL and HTL epitopes. T cell epitopes are recognized by T-cell receptor (TCRs) when bound to MHC molecules on APCs. Prediction methods employ data-driven techniques and analyze sequences from library databases to predict T cell epitopes through peptide-MHC binding.

CTL epitope prediction: To predict the MHC class-I or CTL epitopes, the IEDB recommended NetMHCpan 4.1 EL epitope prediction[39] and consensus method of the IEDB server (http://tools.immuneepitope.org/mhci/)[40]. The HLA system, a highly variable genetic system contributes to the varying severity of COVID-19 cases[41]. In our study, we have predicted CTLs that are recognized by the following HLA class I alleles: HLA-A*01:01, HLA-A*02:02, HLA-A*03:01, HLA-B*35, HLA-C*04:01, HLA-C*12:03, and HLA-DRB1*15:01. These alleles have been reported to be positively associated with either the severity or mortality of COVID-19 cases[42–47]. The toxicity of the CTL epitopes was examined by ToxinPred2 tool[48], which employs the Hybrid (RF + BLAST + MERCI) algorithms. All these predictions were made based on the primary sequence input.

HTL epitope prediction: Fifteen long MHC class II binding peptides were predicted from the S-protein sequence using the IEDB server (http://tools.immuneepitope.org/mhcii/). Studies have shown that individuals carrying HLA-DQB1*06 alleles produce higher antibody responses against SARS-CoV-2 S-protein and the RBD post-vaccination[49]. We used the positive association of HLA-DQB1*06 with infectivity severity to predict Helper T cell response. The IBDB recommended consensus 2.22 method and NetMHCIIpan 2.3 were used to predict HTL epitopes and select peptides based on their binding affinity. We identified good binders of MHC II with IC50 values < 250nM and percentile rank < 10. These methods accurately predicted epitopes that are essential in effective vaccine development.

Antigenicity, allergenicity, toxicity prediction of HTL epitopes

We used VaxiJen v2.0[50]to identify whether T cell epitopes can trigger an immune response inside the host and to predict the protective antigens and vaccine subunits from the S-protein sequence. Allergenicity determination with AllerTop v2.0[51]assessed allergic reaction potential. The ToxinPred2 tool[48]was used to check toxicity using Hybrid (RF + BLAST + MERCI) algorithms. Peptide stability was evaluated with the Protpram online server[52].

Cytokine-inducing ability prediction of the HTL epitopes

Prediction of the cytokine-inducing capability of HTL or MHC class-II epitopes is important for vaccine design. We used the IFN epitope server[53]to determine the IFN-gamma-inducing ability of predicted HTL epitopes[54] and combined Motif and SVM predictions to ensure accuracy. Furthermore, we assessed the IL-4 and IL-10 inducing properties of the HTL epitopes by using the IL4pred and IL10pred servers, respectively, employing the default thresholds for prediction, and both employed the SVM method.

Structure modeling and validation of the T-cell epitopes

The PEPFOLD 3.0 web server was used to employ a de novo modeling approach to determine the binding of T-cell epitopes with the MHC I molecule[55,56]. The stereochemical quality of the modeled structure was evaluated from the Ramachandran plot using the PROCHECK tool of PDBsum (https://www.ebi.ac.uk/thornton-srv/databases/pdbsum/)[57].

Interaction analysis of the predicted CTL with the MHC I molecule

We then performed an interaction-based docking study to analyze the interaction between our modeled T-cell epitopes and MHC I using the Zdock protein-protein docking module from the Biovia Discovery Studio 2022 (DS) software suite[58]. In consideration of the narrower binding pockets for MHCII, which are less demanding and less accurate, our study focused on MHC I. Specifically we downloaded the most prevalent HLA allele, HLA-A* 02:01 (PDB ID 7RTD) for analysis. Utilizing the essential contact residues of the MHC I protein, we conducted docking with modeled 5 CTL epitopes, chosen based on antigenicity scoring. Our selection of the most significant docked poses for each peptide was based on intermolecular contacts and ZDOCK score. Following this stage, we refined the poses employing RDOCK's CHARMM energy minimization and PRODIGY webserver for predicting binding energy (ΔG) and dissociation constant. Subsequently, we assessed and selected the best-fit docked pose based on molecular interactions, ZDOCK score, RDOCK score, and binding energy.

Molecular dynamics study on the modeled Epitope-MHC I complex

The stability and flexibility of the epitope-MHC I complex in physiological environments were analyzed through Molecular Dynamics Studies utilizing the Schrödinger Desmond tool (Academic version)[59] with NVIDIA. The system was subjected to default initialization and equilibration using Brownian dynamics simulation in the NVT ensemble for 10 picoseconds, followed by equilibration in the NPT ensemble at 10 K and 1 bar for 36 picoseconds. MD production runs were conducted over a duration of 100 nanoseconds, employing the Nose-Hoover thermostat and Martyna-Tobias-Klein barostat at 300 K and 1.01 bar. The stability was evaluated using the C-alpha atom Root Mean Square Deviation (C-RMSD) and C-alpha atom Root Mean Square Fluctuation (C-RMSF) in Angstroms. Additionally, the intermolecular binding strength was analyzed by examining the hydrogen-bound formation profile between the epitope and the MHC I throughout the entire MD simulation run.

Selection of overlapping epitopes and construction of the multiepitope vaccine construct Immunoinformatics have been used to identify specific epitope regions to elicit responses from cytotoxic T cells (CTL), helper T cells (HTL), and B-cells. These epitopes can be integrated to formulate a multiepitope subunit vaccine construct, utilizing in silico methods for design. This innovative approach has recently emerged as a promising strategy for the development of vaccines targeting comparable infectious diseases[19,60–64]. In those studies, adjuvants like APPHALS and β-defensin, along with linkers such as AYY, KK, EAAAK, and GPGPG, were utilized to improve protein stability by creating an extended conformation, separating functional domains, and promoting protein folding.

Our modified approach to designing a multiepitope vaccine construct involved selecting candidate subunits based on their CTL, HTL, and linear B-cell epitopes. We specifically chose subunits that were in proximity, had overlapping regions, or were linked through published reports. These subunits were connected using the GPGPG linker in various combinations, and adjuvants such as human Beta-Defensin 2 and 3 were linked using the EAAAK linker at the N and C terminals, respectively, to produce multiple subunit constructs. The ultimate selection of the final construct was based on a highly confident 3D structure that exhibited strong binding affinity with the receptor protein during subsequent assessments. Consequently, the selected subunits were combined to produce the final multiepitope vaccine construct, referred to as 'MEVC'.

Population coverage analysis of the epitope subunits

The IEDB population coverage analysis tool was employed to project the immune response of the global, Indian, Chinese, and North American populations to specific overlapping CTL epitopes. This projection was based on the corresponding HLA genotype frequencies. The analysis utilized the HLA-A, HLA-B, and HLA-C allele sets to encompass the entire population. This approach aimed to assess the potential immune reactivity to these epitopes in populations affected by COVID-19.

Characterization of designed vaccine construct

Physicochemical property analysis of the construct The physicochemical properties of the vaccine were analyzed by submitting the final vaccine sequence to the Expasy Protpram online server[52]. The server provides estimates for several parameters, including the molecular weight, estimated half-life, theoretical isoelectric point (pl), aliphatic index, stability index, and grand average of hydropathicity (GRAVY).

Anitgenicity allergenicity toxicity analysis of vaccine construct The proposed vaccine's antigenicity was predicted using the VaxiJen v2.0 server[50]. Allergenicity prediction was conducted using the Allertop v.2.0 tool[51]. ToxinPred2 tool employing the Hybrid (RF + BLAST + MERCI) algorithms was utilized to evaluate the toxicity of the final construct[48]. These predictions were based on the primary sequence input.

Proinflammatory and anti-inflammatory response prediction

A successful vaccine should elicit the early release of protective cytokines, followed by the release of anti-inflammatory cytokines to mitigate potential immunopathology. Our vaccine construct was subjected to evaluation for its anti-inflammatory and pro-inflammatory properties using PIP-EL and Pre-AIP servers. PIP-EL tool (http://211.239.150.230/PIP-EL/)[65]was employed to ascertain whether the peptide can provoke pro-in flammatory cytokines, such as IL-8, IL-12, IL-18, IFN-γ, and TNF-α. A peptide is deemed pro-inflammatory if it has the capability to elicit these cytokines. An activation of proinflammatory cytokine is indicated if the threshold value is 0.45 or higher. This investigation offers valuable insights into the levels of proinflammatory cytokines, such as interleukin-6 (1 L-6), interleukin-1 (IL-1), interleukin-17 (IL-17), and tumor necrosis factor-alpha (TNF-α), which can trigger cytokine storm and are linked to cases of acute respiratory distress syndrome (ARDS) in individuals with COVID-19[66].

The Pre-AIP server (http://kurata14.bio.kyutech.ac.jp/PreAIP/)[67]employs a random forest classifier to assess peptides for their anti-inflammatory properties based on primary sequence, evolutionary, and structural data. Peptides that elicit anti-inflammatory cytokines like IL-10, IL-4, IL-13, IL-22, and TGF-β are classified as anti-inflammatory. A score of ≥ 0.46 is indicative of high-confidence anti-inflammatory potential. Also, IL4-inducing regions of the vaccine construct were scanned using IL4Pred, and the IFN-gamma-inducing regions of the construct were evaluated using the IFNepitope2 server[53], which provides details on the host-specific IFN-gamma-inducing region of the given protein sequence.

Molecular modeling refinement and validation of MEVC

The 3-D structure of the designed construct was modeled using various tools, and the trRosetta webserver (https://yanglab.qd.sdu.edu.cn/trRosetta/), which employs deep neural network technology for generating protein structures[68] yielded a high evaluation quality 3D conformation. The model with the highest TM score (0.335) from trRosetta webserver underwent refinement through the ModRefiner (https://zhanggroup.org/ModRefiner/)[69] and GalaxyWeb (https://galaxy.seoklab.org/index.html)[70]. Subsequently, the refined model was subjected to validation using the PROCHECK tool of PDBsum (https://www.ebi.ac.uk/thornton-srv/databases/pdbsum/)[57] to examine the Ramachandran Plot, ensuring the structure adheres to stringent stereochemical quality standards.

Molecular docking of vaccine construct with toll-like receptor (TLR) receptor

The 3D structure of human TLR-3 and TLR-4, obtained from the protein data bank (PDB ID: 2A0Z and 3FXI respectively) was preprocessed using the molecular viewer in Discovery Studio. Molecular docking analysis of the interaction of the vaccine with TLR-3 and TLR-4 was conducted using the Cluspro v.2 protein-protein docking server (https://cluspro.org/help.php)[71]. The server provides cluster scores based on rigid docking by sampling billions of conformations and performing pairwise RMSD energy minimization. The final vaccine-TLR-3 and vaccine-TLR-4 complex models were selected based on the lowest energy weight score. The intermolecular interactions in the complex structure were assessed using the Prot-Prot analysis in PDBSum[57] and the PRODIGY server (https://bianca.science.uu.nl/prodigy/)[26] was employed to predict the contact-based binding affinity of the complex. This involved computing the binding energy (ΔG) and dissociation constant (Kd) for the docked MEVC-TLR-3 and MEVC-TLR-4 complex structure.

Molecular dynamic simulation of MEVC-TLR-3 complex

Dynamic simulation study of the MEVC-TLR-3 complex was carried out using the online server iMODS (http://imods.chaconlab.org/)[72]to assess and measure the flexibility of the proteins. This tool efficiently predicts various dynamic parameters including deformability, B-factor, mobility profiles, individual values, variance, covariance map, and protein complex elastic network. All-atom MD simulation was performed for 50 ns chemical time using the Schrödinger Desmond tool[59] with NVIDIA GPU support. The system was equilibrated using Brownian dynamics simulation in the NVT ensemble for 10 picoseconds, followed by equilibration in the NPT ensemble at 10 K and 1 bar for 36 picoseconds. MD production runs were performed for 50 ns, employing the Nose-Hoover thermostat and Martyna-Tobias-Klein barostat at 300 K and 1.01 bar. The stability was evaluated using the C-alpha atom Root Mean Square Deviation (C-RMSD) and intermolecular H-bond formation profile. Binding free energy (ΔG) and dissociation constant (Kd) of the MEVC-TLR-3 complex conformation after the 50 ns simulation were also computed using the PRODIGY webserver (https://bianca.science.uu.nl/prodigy/) and compared with that of the initial conformation to evaluate the binding affinity changes of the complex.

In-silico cloning and vaccine optimization To ensure the efficiency of a multi-epitope vaccine construct, it is essential to employ careful design measures. The construct was optimized using the Java Codon Adaptation Tool (JCat) (https://www.jcat.de/Result.jsp)[73] to achieve this objective. The optimization process included cod on adaptation for expression in E. coli K12 strain, specifically carried out in E. coli strain K12. The selection of specific parameters, such as avoidance of transcription terminators, ribosome binding sites, and cleavage sites of restriction enzymes, was undertaken to optimize the sequence. The JCat tool has been utilized to provide the CAI value and GC content of the optimized sequence. For optimal translation, stability, and transcription efficiency, it is recommended that the GC content of the vaccine should fall within the range of 30–70%. Additionally, the vaccine's CAI value should exceed 0.8 or be closer to 1.0. A higher CAI value acts as an indicator of the high expression level of the gene. To clone the construct in silico, the Serial Cloner software 2.6.1 version (Serial Cloner. Vers. 2.6.1 2021) was used, and the restriction enzymes Xho1 and NotI were employed to insert it into the E. coli pET–28(+) vector.

Immune simulations of vaccine construct In order to assess the immunogenic profiles and immune response of the multi-epitope vaccine on the human immune system, we utilized the C-ImmSim server (https://kraken.iiban.rm.cnr.it/C-IMMSIM/index.php)[74]. The C-IMMSIM tool leverages position-specific scoring matrix (PSSM) and machine learning techniques to evaluate the production of cytokines, interferon, and antibodies. Furthermore, it provides predictions for T helper cell 1 and T helper cell 2 (Th1 and Th2). Default parameters like random seed, simulation volume, and simulation step were used. As part of safety measures, we also assessed the potential for cytokine storm upon the introduction of the vaccine. This involved evaluating proinflammatory cytokine levels based on the 3D structure of the vaccine construct using the C-IMMSIM tool.

Data availability

Additional data is provided as a supplementary file.

Received: 6 August 2024; Accepted: 4 March 2025
Published online: 20 March 2025

References

1. Zhang, T., Wu, Q. & Zhang, Z. Probable Pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Curr. Biol.* **30** (7), 1346–1351. https://doi.org/10.1016/j.cub.2020.03.022 (2020).
2. Muralidar, S., Ambi, S. V., Sekaran, S. & Krishnan, U. M. The emergence of COVID-19 as a global pandemic: Understanding the epidemiology, immune response and potential therapeutic targets of SARS-CoV-2. *Biochimie* **179**, 85–100. https://doi.org/10.1016/j.biochi.2020.09.018 (2020).
3. Singh, D. On the origin and evolution of SARS-CoV-2. *Exp. Mol. Med.* https://doi.org/10.1038/s12276-021-00604-z (2021).
4. Rahman, M. S. et al. October., Gene Reports Mutational insights into the envelope protein of SARS-CoV-2, *Gene Reports*, vol. 22, no. p. 100997, 2021, (2020). https://doi.org/10.1016/j.genrep.2020.100997
5. Singh, P. K., Kulsum, U., Rufai, S. B., Mudliar, S. R. & Singh, S. Mutations in SARS-CoV-2 Leading to Antigenic Variations in Spike Protein: A Challenge in Vaccine Development, *J. Lab. Physicians*, vol. 12, no. 02, pp. 154–160, Aug. (2020). https://doi.org/10.1055/s-0040-1715790
6. Lee, C. H. & Koohy, H. In silico identification of vaccine targets for 2019-nCoV, *F1000Research*, vol. 9, p. 145, Apr. (2020). https://doi.org/10.12688/f1000research.22507.2
7. Choudhury, A. & Mukherjee, S. In silico studies on the comparative characterization of the interactions of SARS-CoV-2 spike glycoprotein with ACE-2 receptor homologs and human TLRs, *J. Med. Virol.*, vol. 92, no. 10, pp. 2105–2113, Oct. (2020). https://doi.org/10.1002/jmv.25987
8. Du, L. et al. The spike protein of SARS-CoV — a target for vaccine and therapeutic development, *Nat. Rev. Microbiol.*, vol. 7, no. 3, pp. 226–236, Mar. (2009). https://doi.org/10.1038/nrmicro2090
9. Agarwal, V., Tiwari, A. & Varadwaj, P. K. Designing of Epitope-Based Vaccine from the Conserved Region of the Spike Glycoprotein of SARS-CoV-2 (prepint), *bioRxiv*, no. August, p. 2020.08.27.269456, (2020). https://doi.org/10.1101/2020.08.27.269456
10. Deepthi, V., Mohanakumar, K. P. & Rajamma, U. Efficacy of defensins as neutralizing agents against the deadly SARS-CoV-2. *J. Biomol. Struct. Dyn.* **41** (7), 2911–2925. https://doi.org/10.1080/07391102.2022.2041487 (2023).
11. Chi, X. et al. Aug., A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2, *Science (80-.).*, vol. 369, no. 6504, pp. 650–655, (2020). https://doi.org/10.1126/science.abc6952
12. Arvin, A. M. et al. A perspective on potential antibody-dependent enhancement of SARS-CoV-2, *Nature*, vol. 584, no. 7821. Nature Research, pp. 353–363, Aug. 20, (2020). https://doi.org/10.1038/s41586-020-2538-8
13. de Alwis, R., Chen, S., Gan, E. S. & Ooi, E. E. Impact of immune enhancement on Covid-19 polyclonal hyperimmune globulin therapy and vaccine development, *EBioMedicine*, vol. 55. Elsevier B.V., May 01, (2020). https://doi.org/10.1016/j.ebiom.2020.102768
14. Testa, J. S. & Philip, R. Role of T-cell epitope-based vaccine in prophylactic and therapeutic applications, *Future Virol.*, vol. 7, no. 11, pp. 1077–1088, Nov. (2012). https://doi.org/10.2217/fvl.12.108

15. Sanchez-Trincado, J. L., Gomez-Perosanz, M. & Reche, P. A. Fundamentals and Methods for T- and B-Cell Epitope Prediction, *Journal of Immunology Research*, vol. Hindawi Limited, 2017. (2017). https://doi.org/10.1155/2017/2680160
16. Khailany, R. A., Safdar, M. & Ozaslan, M. Genomic characterization of a novel SARS-CoV-2. *Gene Rep.* **19**, 100682. https://doi.org/10.1016/j.genrep.2020.100682 (2020).
17. Abraham Peele, K., Srihansa, T., Krupanidhi, S., Ayyagari, V. S. & Venkateswarulu, T. C. Design of multi-epitope vaccine candidate against SARS-CoV-2: a in-silico study. *J. Biomol. Struct. Dyn.* **0** (0), 1–9. https://doi.org/10.1080/07391102.2020.1770127 (2020).
18. Medha, P., Bhatt, Priyanka, M., Sharma & Sharma, S. Prediction and identification of T cell epitopes of COVID-19 with balanced cytokine response for the development of peptide based vaccines. *Silico Pharmacol.* **9** (1), 1–17. https://doi.org/10.1007/s40203-021-00098-7 (2021).
19. Singh, A., Thakur, M., Sharma, L. K. & Chandra, K. Designing a multi – epitope peptide based vaccine against SARS – CoV – 2. *Sci. Rep.* **no. 0123456789**, 1–12. https://doi.org/10.1038/s41598-020-73371-y (2020).
20. Shin, H. S. et al. Mar., Immune Responses to Middle East Respiratory Syndrome Coronavirus During the Acute and Convalescent Phases of Human Infection, *Clin. Infect. Dis.*, vol. 68, no. 6, pp. 984–992, (2019). https://doi.org/10.1093/cid/ciy595
21. Shang, J. et al. Cell entry mechanisms of SARS-CoV-2, *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 21, (2020). https://doi.org/10.1073/pnas.2003138117
22. Hoffmann, M. et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181** (2), 271–280. https://doi.org/10.1016/j.cell.2020.02.052 (2020).
23. Cheng, M. H. et al. Superantigenic character of an insert unique to SARS-CoV-2 spike supported by skewed TCR repertoire in patients with hyperinflammation, *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 41, pp. 25254–25262, (2020). https://doi.org/10.1073/pnas.2010722117
24. Wlodawer, A. Stereochemistry and Validation of Macromolecular Structures, in *Methods Mol Biol.* 2017, pp. 595–610. (2017). https://doi.org/10.1007/978-1-4939-7000-1_24
25. Magazine, N. et al. Immune epitopes of SARS-CoV-2 Spike protein and considerations for universal vaccine development. *Oct* https://doi.org/10.1101/2023.10.26.564184 (2023).
26. Rodrigues, P., Kastritis, P. L., Mjj, A., Xue, L. C. & Vangone, A. Structural bioinformatics PRODIGY: a web server for predicting the binding affinity of protein – protein complexes, vol. 32, no. August, pp. 3676–3678, (2016). https://doi.org/10.1093/bioinformatics/btw514
27. Bortolotti, D. et al. TLR3 and TLR7 RNA sensor activation during SARS-CoV-2 infection. *Microorganisms* **9** (9), 1820. https://doi.org/10.3390/microorganisms9091820 (Aug. 2021).
28. Bhatta, T. K. & Nimesh, S. *The Design and Development of Novel Drugs and Vaccines: Principles and Protocols*. (2021). https://doi.org/10.1016/B978-0-12-821471-8.09995-5
29. Khare, S. et al. GISAID's role in pandemic response. *China CDC Wkly.* **3** (49), 1049–1051. https://doi.org/10.46234/ccdcw2021.255 (2021).
30. Nair, D. T. et al. Epitope recognition by diverse antibodies suggests conformational convergence in an antibody response. *J. Immunol.* **168** (5), 2371–2382. https://doi.org/10.4049/jimmunol.168.5.2371 (2002).
31. Emini, E. A., Hughes, J. V., Perlow, D. S. & Boger, J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J. Virol.* **55** (3), 836–839. https://doi.org/10.1128/jvi.55.3.836-839.1985 (Sep. 1985).
32. Karplus, P. A. & Schulz, G. E. Prediction of chain flexibility in proteins. *Naturwissenschaften* **72** (4), 212–213. https://doi.org/10.1007/bf01195768 (1985).
33. Parker, J. M. R., Guo, D. & Hodges, R. S. New Hydrophilicity Scale Derived from High-Performance Liquid Chromatography Peptide Retention Data: Correlation of Predicted Surface Residues with Antigenicity and X-ray-Derived Accessible Sites,*Biochemistry*, **25**, no. 5425–5432, doi: https://doi.org/10.1021/bi00367a013. (1986).
34. Kolaskar, A. S. & Tongaonkar, P. C. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett.* **276**, 1–2. https://doi.org/10.1016/0014-5793(90)80535-Q (Dec. 1990).
35. Zhou, C. et al. SEPPA 3.0 - enhanced spatial epitope prediction enabling glycoprotein antigens, *Nucleic Acids Res.*, vol. 47, no. W1, pp. W388–W394, (2019). https://doi.org/10.1093/nar/gkz413
36. Haste Andersen, P., Nielsen, M. & Lund, O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci.* **15** (11), 2558–2567. https://doi.org/10.1110/ps.062405906 (2006).
37. Ponomarenko, J. et al. ElliPro: A new structure-based tool for the prediction of antibody epitopes. *BMC Bioinform.* **9**, 1–8. https://doi.org/10.1186/1471-2105-9-514 (2008).
38. Sun, J. et al. SEPPA: A computational server for spatial epitope prediction of protein antigens, *Nucleic Acids Res.*, vol. 37, no. SUPPL. 2, pp. 612–616, (2009). https://doi.org/10.1093/nar/gkp417
39. Reynisson, B., Alvarez, B., Paul, S., Peters, B. & Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif Deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **48**, W449–W454. https://doi.org/10.1093/NAR/GKAA379 (2021).
40. Moutaftsi, M. et al. A consensus epitope prediction approach identifies the breadth of murine TCD8+-cell responses to vaccinia virus. *Nat. Biotechnol.* **24** (7), 817–819. https://doi.org/10.1038/nbt1215 (2006).
41. Pretti, M. A. M. et al. Class I HLA Allele Predicted Restricted Antigenic Coverages for Spike and Nucleocapsid Proteins Are Associated With Deaths Related to COVID-19, *Front. Immunol.*, vol. 11, no. December, pp. 1–14, (2020). https://doi.org/10.3389/fimmu.2020.565730
42. La Porta, C. A. M. & Zapperi, S. Estimating the binding of Sars-CoV-2 peptides to HLA class I in human subpopulations using artificial neural networks. *Cell. Syst.* **11** (4), 412–417. https://doi.org/10.1016/j.cels.2020.08.011 (2020).
43. De Moura, R. R. et al. Immunoinformatic approach to assess SARS-CoV-2 protein S epitopes recognised by the most frequent MHC-I alleles in the Brazilian population. *J. Clin. Pathol.* **74** (8), 528–532. https://doi.org/10.1136/jclinpath-2020-206946 (2021).
44. Alnaqbi, H. et al. HLA repertoire of 115 UAE nationals infected with SARS-CoV-2, *Hum. Immunol.*, vol. 83, no. 1, pp. 1–9, (2022). https://doi.org/10.1016/j.humimm.2021.08.012
45. Shkurnikov, M. et al. Association of HLA class I genotypes with severity of coronavirus Disease-19. *Front. Immunol.* **12, no.** https://doi.org/10.3389/fimmu.2021.641900 (February, 2021).
46. Tay, G. K. et al. HLA class I associations with the severity of COVID-19 disease in the united Arab Emirates. *PLoS One*. **18**, 1–14. https://doi.org/10.1371/journal.pone.0285712 (2023). 9 September.
47. Nguyen, A. et al. Human leukocyte antigen susceptibility map for SARS-CoV-2. *J. Virol.* **94**, 1–12. https://doi.org/10.1128/JVI.00510-20 (2020).
48. Neelam, S., Devi, N. L., Shipra, J. & S, R. G. P. ToxinPred2: an improved method for predicting toxicity. *Brief Bioinform.* **23**(5), 1–12 (2022).
49. Mentzer, A. J. et al. Human leukocyte antigen alleles associate with COVID-19 vaccine immunogenicity and risk of breakthrough infection. *Nat. Med.* **29** (1), 147–157. https://doi.org/10.1038/s41591-022-02078-6 (2023).
50. Doytchinova, I. A. & Flower, D. R. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinform.* **8**, 1–7. https://doi.org/10.1186/1471-2105-8-4 (2007).
51. Dimitrov, I., Bangov, I., Flower, D. R. & Doytchinova, I. AllerTOP v.2 - A server for in Silico prediction of allergens. *J. Mol. Model.* https://doi.org/10.1007/s00894-014-2278-5 (2014).
52. Gasteiger, E. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31** (13), 3784–3788. https://doi.org/10.1093/nar/gkg563 (Jul. 2003).

53. Dhall, A., Patiyal, S. & Raghava, G. P. S. A hybrid method for discovering interferon-gamma inducing peptides in human and mouse. *BioRxiv* **302**, 2022–2023 (2023).
54. Dhanda, S. K., Vir, P. & Raghava, G. P. S. Designing of interferon-gamma inducing MHC class-II binders. *Biol. Direct*. **8** (1), 1–15. https://doi.org/10.1186/1745-6150-8-30 (2013).
55. Shen, Y., Maupetit, J., Derreumaux, P. & Tufféry, P. Improved PEP-FOLD approach for peptide and miniprotein structure prediction. *J. Chem. Theory Comput.* **10** (10), 4745–4758. https://doi.org/10.1021/ct500592m (2014).
56. Thévenet, P. et al. PEP-FOLD: an updated de Novo structure prediction server for both linear and disulfide bonded Cyclic peptides. *Nucleic Acids Res.* **40**, 288–293. https://doi.org/10.1093/nar/gks419 (2012).
57. Laskowski, R. A. PDBsum new things, vol. 37, no. November pp. 355–359, 2009, (2008). https://doi.org/10.1093/nar/gkn860
58. Systemes, D. Biovia Discovery Studio ® 2016 Comprehensive Modeling and Simulations. *3D Exp.*, p. 4, (2022).
59. Schrödinger Release 2023-1, System, D. M. D., Shaw Research, D. E., York, N. & York, N. NY, Maestro-Desmond Interoperability Tools, Schrödinger, NY, 2024, *Desmond Mol. Dyn. Syst. D. E. Shaw Res. New York, NY, 2024. Maest. Interoperability Tools, Schrödinger, New York, NY, 2024*, 2023. (2024).
60. Ali, M., Pandey, R. K., Khatoon, N., Narula, A. & Mishra, A. Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection, *Sci. Rep.*, no. April, pp. 1–13, (2017). https://doi.org/10.1038/s41598-017-09199-w
61. Chauhan, V., Rungta, T., Goyal, K. & Singh, M. P. Designing a multi-epitope based vaccine to combat Kaposi sarcoma utilizing immunoinformatics approach. *Sci. Rep.* **9** (1), 2517. https://doi.org/10.1038/s41598-019-39299-8 (Feb. 2019).
62. Oladipo, E. K. et al. Bioinformatics analysis of structural protein to approach a vaccine candidate against Vibrio cholerae infection. *Immunogenetics* **75** (2), 99–114. https://doi.org/10.1007/s00251-022-01282-5 (2023).
63. Joshi, A., Chandra, B., Mannan, M. A. & Kaushik, V. Epitope based vaccine prediction for SARS-COV-2 by deploying immuno-informatics approach informatics in medicine unlocked epitope based vaccine prediction for SARS-COV-2 by deploying immuno-informatics approach. *Inf. Med. Unlocked*. **19**, 100338. https://doi.org/10.1016/j.imu.2020.100338 (2020).
64. Dong, R., Chu, Z., Yu, F. & Zha, Y. Contriving Multi-Epitope Subunit of Vaccine for COVID-19: Immunoinformatics Approaches, vol. 11, no. December 2020, (2019). https://doi.org/10.3389/fimmu.2020.01784
65. Manavalan, B., Shin, T. H., Kim, M. O. & Lee, G. A new ensemble learning method for improved Proinflammatory peptide predictions. *Front. Immunol.* **9** https://doi.org/10.3389/fimmu.2018.01783 (Jul. 2018).
66. Montazersaheb, S., Mahdi, S., Khatibi, H., Hejazi, M. S. & Tarhriz, V. COVID – 19 infection: an overview on cytokine storm and related interventions. *Virol. J.* https://doi.org/10.1186/s12985-022-01814-1 (2022).
67. Khatun, M. S., Hasan, M. M. & Kurata, H. PreAIP: Computational Prediction of Anti-inflammatory Peptides by Integrating Multiple Complementary Features, *Front. Genet.*, vol. 10, Mar. (2019). https://doi.org/10.3389/fgene.2019.00129
68. Du, Z. et al. The trrosetta server for fast and accurate protein structure prediction. *No Dec.* https://doi.org/10.1038/s41596-021-00628-9 (2021).
69. Xu, D. & Zhang, Y. Improving the physical realism and structural accuracy of protein models by a Two-Step Atomic-Level energy minimization. *Biophysj* **101** (10), 2525–2534. https://doi.org/10.1016/j.bpj.2011.10.024 (2011).
70. Ko, J., Park, H., Heo, L. & Seok, C. GalaxyWEB server for protein structure prediction and refinement, vol. 40, no. May, pp. 294–297, (2012). https://doi.org/10.1093/nar/gks493
71. Kozakov, D. et al. The cluspro web server for protein-protein Docking, **12**, 2, pp. 255–278, (2018). https://doi.org/10.1038/nprot.2016.169.The
72. López-Blanco, J. R., Aliaga, J. I., Quintana-Ortí, E. S. & Chacón, P. iMODS: internal coordinates normal mode analysis server, *Nucleic Acids Res.*, vol. 42, no. W1, pp. W271–W276, Jul. (2014). https://doi.org/10.1093/nar/gku339
73. Grote, A. et al. A novel tool to adapt codon usage of a target gene to its potential expression host. *JCat* **33**, 526–531. https://doi.org/10.1093/nar/gki376 (2005).
74. Rapin, N., Lund, O., Bernaschi, M. & Castiglione, F. Computational immunology Meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system, **5**, 4, (2010). https://doi.org/10.1371/journal.pone.0009862

Acknowledgements

The work was funded by the Department of Higher Education and the Department of Health and Family Welfare, Govt. of Kerala. VD is a recipient of the Indian Council for Medical Research, Govt. of India, Research Associateship (No.3/1/13/Neuro/139/2020-NCD-I).

Author contributions

V.D. is one of the team members who conceived the idea, executed the computational work, and drafted and reviewed the MS. A.S. performed the initial computational analysis and drafted a portion of the manuscript. K.P.M. is another team member who was involved in generating the concept, discussing the results, and editing the MS. U.R. is the Coordinator of the team that conceived the project idea, monitored the work, discussed the results and finalized the MS.

Declarations

Competing interests

The authors declare no Competing interests in publishing this manuscript in the Journal - Nature Computational Science.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-92956-z.

Correspondence and requests for materials should be addressed to U.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025