

RESEARCH

Open Access



Designing a novel and combinatorial multi-antigenic epitope-based vaccine “MarVax” against Marburg virus—a reverse vaccinology and immunoinformatics approach

Bishal Debroy^{1†} , Sribas Chowdhury^{2†}  and Kuntal Pal^{3,4*} 

Abstract

Context Marburg virus (MARV) is a member of the Filoviridae family and causes Marburg virus disease (MVD) among humans and primates. With fatality rates going up to 88%, there is currently no commercialized cure or vaccine to combat the infection. The National Institute of Allergy and Infectious Diseases (NIAID) classified MARV as priority pathogen A, which presages the need for a vaccine candidate which can provide stable, long-term adaptive immunity. The surface glycoprotein (GP) and fusion protein (FP) mediate the adherence, fusion, and entry of the virus into the host cell via the TIM-1 receptor. Being important antigenic determinants, studies reveal that GP and FP are prone to evolutionary mutations, underscoring the requirement of a vaccine construct capable of eliciting a robust and sustained immune response. In this computational study, a reverse vaccinology approach was employed to design a combinatorial vaccine from conserved and antigenic epitopes of essential viral proteins of MARV, namely GP, VP24, VP30, VP35, and VP40 along with an endogenous protein large polymerase (L).

Methods Epitopes for T-cell and B-cell were predicted using TepiTool and ElliPro, respectively. The surface-exposed TLRs like TLR2, TLR4, and TLR5 were used to screen high-binding affinity epitopes using the protein-peptide docking platform MdockPeP. The best binding epitopes were selected and assembled with linkers to design a recombinant multi-epitope vaccine construct which was then modeled in Robetta. The *in silico* biophysical and biochemical analyses of the recombinant vaccine were performed. The docking and MD simulation of the vaccine using WebGro and CABS-Flex against TLRs support the stable binding of vaccine candidates. A virtual immune simulation to check the immediate and long-term immunogenicity was carried out using the C-ImmSim server.

Results The biochemical characteristics and docking studies with MD simulation establish the recombinant protein vaccine construct MarVax as a stable, antigenic, and potent vaccine molecule. Immune simulation studies reveal 1-year passive immunity which needs to be validated by *in vivo* studies.

Keywords Marburg virus disease, Immunoinformatics, Epitope, Multi-epitope vaccine, Molecular docking, Molecular dynamics

[†]Bishal Debroy and Sribas Chowdhury contributed equally.

*Correspondence:

Kuntal Pal

kuntal.pal@vit.ac.in

Full list of author information is available at the end of the article

Background

The severe, acute, and recurring Marburg virus disease (MVD) is caused by the Marburg virus (MARV) and has been linked to many devastating outbreaks with fatality rates going as high as 83–88%. Reports of the first known outbreak were described in 1967 in the cities of Marburg in Germany and Belgrade in Serbia, leading to the detection of the disease. This was followed by several outbreaks of the fatal disease that recurred over subsequent periods with the most recent outbreak reported in Ghana (2022), Tanzania, and Equatorial Guinea (2023). The most fatal endemic was the MVD outbreak in Angola in 2004–2005 which had 252 cases reported and 227 confirmed deaths inferring a 90% fatality rate [1]. MARV is related to the hemorrhagic fever-causing Ebola virus (EBOV). Due to its non-segmented, negative-strand RNA genome serving as a genetic template for reproduction, the virus is categorized in the order of mononegavirales [2, 3]. The Egyptian fruit bat, *Rousettus aegyptiacus*, is the reservoir species for MARV, and its rare outbreak is mostly related to the geographical range of these bats. Prior research revealed that the human population is in significant danger from the MARV infection of the *Rousettus* bats, which may climb up to 10% in young bats

during seasonal surges [4]. According to some reports, the virus may also exist in pigs, African green monkeys, and other susceptible reservoirs [1, 5, 6].

The most virulent version of this pleomorphic virus measures 80 nm in diameter and 790 nm in length [7–9]. As shown in Fig. 1, this mononegavirales has a 19.1 Kb length RNA that codes for seven structural proteins, including nucleoprotein (NP), glycoprotein (GP), viral protein 24 (VP24), viral protein 30 (VP30), viral protein 35 (VP35), viral protein (VP40), and large protein (L) [9–11]. The nucleocapsid complex is made up of the proteins NP, VP30, VP35, and L, where L functions as RNA-dependent RNA polymerase (RdRp) and VP35 is a co-factor of polymerase as well as an IFN antagonist [9, 12]. GP is required for the virus to adhere to its host cell [13]. VP40 interferes with the JAK-STAT pathway and is linked to the virus’s budding, whereas VP24 is responsible for the progeny virion’s release from the host cell [9, 14, 15]. Despite some findings showing the presence of viral particles in the rectal, mouth, and urine samples of the MARV-infected bats, transmission of the virus from the reservoir to humans is yet to be known [9, 16]. Further research revealed that the virus is present in an infected bat’s intestine, lung, kidney, salivary gland, and

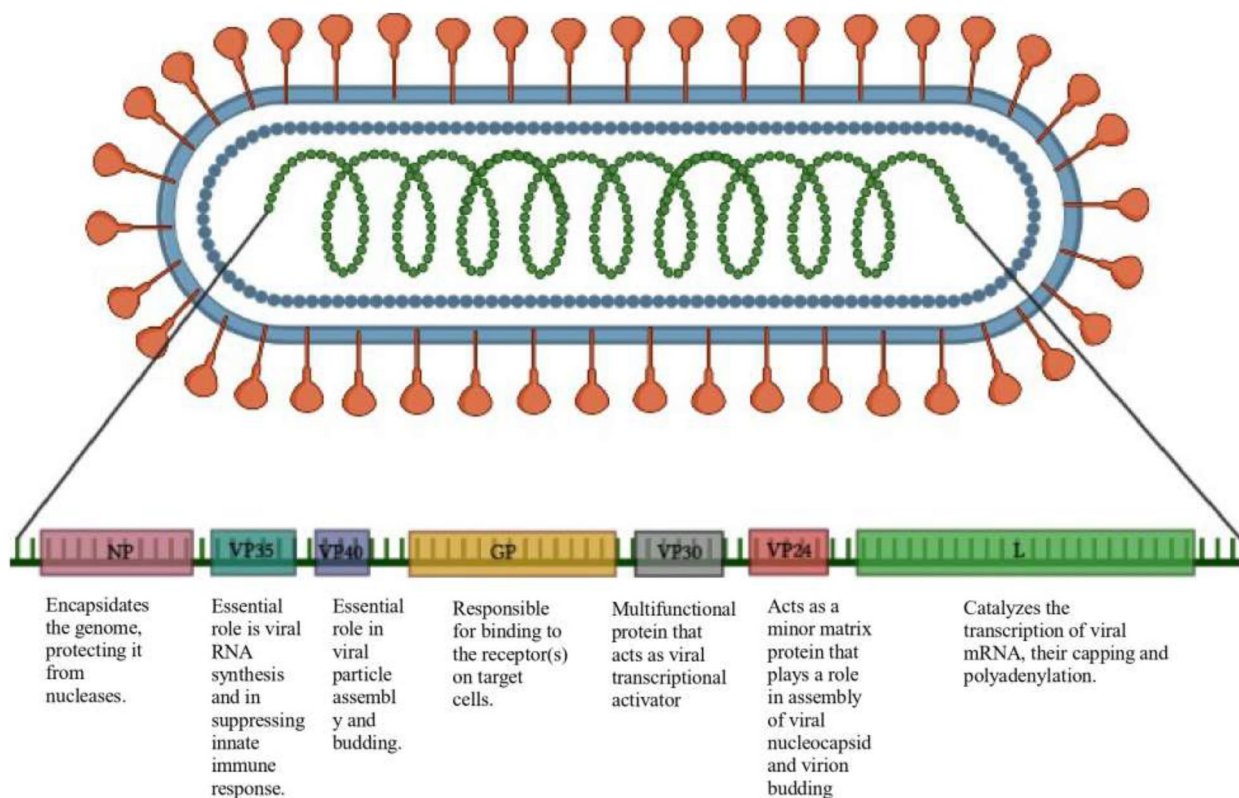


Fig. 1 Genomic organization of Marburg virus representing the function of structural and non-structural proteins, created with BioRender (<https://www.biorender.com/>)

reproductive systems, which raises the possibility of both vertical and horizontal transmission of the virus [17]. The blood, body fluids like saliva and breast milk, and sexual contact are all possible routes for human-to-human transmission [9].

The virus can enter the host in several different ways, and after cellular attachment to the TIM-1 receptor, endocytosis, and fusion, it releases its viral RNA into the host cell [2, 9]. The MARV VP40 interacts with the viral nucleocapsid complex and serves as an interface for both filopodia and sub-viral MARV particles. Filopodia are in intimate contact with the adjacent cells, which promotes the spreading of MARV and raises the viral titer in the blood of infected humans [2, 18]. The liver assumes a pivotal role in MARV replication, leading to hepatocyte degradation, reticuloendothelial system impairment, and hepatocyte injury via a measured inflammatory cascade, resulting in edema and significant damage to the infected host's system. Primary infection-mediated immunomodulation decreases the proliferation of immune cells that cause secondary infections [2, 9]. The liver damage in MARV-infected patients is more severe than in the case of EBOV infection, with the lymph nodes, spleen, testes, ovaries, gastrointestinal system, and endocardium suffering from severe necrotic lesions [2].

The virus has three stages of infection and an incubation period of 2 to 21 days [1, 9]. Phase 1, also known as the generalized phase, lasts for 5 days and is characterized by a high fever (39 to 40 °C) and influenza-like illness. Fatigue, dysphagia, pharyngitis, leukopenia, and thrombocytopenia are other manifestations [6, 9, 19]. Phase 2 continues with a high fever and then progresses to liver, pancreatic, and renal dysfunction. About 75% of the patients experience hemorrhagic signs, along with neurological symptoms, dyspnea, and abnormal vascular permeability [9, 20]. Phase 3, the final stage, might result in either of two outcomes. The patient may enter the phase of recovery or the infection could become lethal. Fatality is characterized by shock and multi-organ failure which is the primary cause of death [21]. A severe metabolic imbalance takes place, leaving a negative impact on a patient's health. During this stage, it is common to experience exhaustion, partial amnesia, sweating, peeling skin in rash-affected areas, and secondary infections [9].

Classification of MARV as priority pathogen A by The National Institute of Allergy and Infectious Diseases (NIAID) and its categorization as category A bio-terrorism agent by the Centers for Disease Control and Prevention (CDC) necessitates the need of a vaccine candidate with long-term adaptive immunity [22]. There is currently no commercialized vaccine or known cure targeted towards combating MVD. The previously reported vaccines against MARV are mostly based on

certain exogenous proteins, which were found to be susceptible to significant rates of mutation [2, 23]. In light of the COVID-19 pandemic, it is clear that a high number of hypervariable regions in surface-exposed viral proteins is related to increased viral pathogenicity and decreased efficiency of vaccines in conferring long-term immunity against viral infection [24]. Hence, developing a vaccine which is not only antigenic and inclusive of epitopes of all antigenic viral proteins but also addresses the long-term stability of the vaccine construct is the need of the hour.

In this study, we have used computational biology and immunoinformatics to design a multi-epitope vaccine that is stable and non-allergic to humans while yet having the capacity to trigger the required immune response against the Marburg virus. We have used B-cell and T-cell epitopes of GP, VP24, VP35, VP40, and RdRp to design a combinatorial vaccine with the ability to generate a strong immune response while also conferring stable, long-term immunity. The viral protein VP24 is in charge of the release of new viral offspring, and the surface glycoprotein, or GP, facilitates the viral entry in the host cell. By altering the JAK-STAT pathway, immune gene suppression, and IFN-driven cascade, VP35 and VP40 play a part in the host system's immune evasion. RdRp, out of all of them, is crucial for viral replication since it replicates the viral genome. All these factors establish the epitopes of these proteins as suitable antigenic targets for the development of an efficient vaccine construct.

Based on our analysis of the immunogenicity and conserved regions of epitopes of the five crucial viral proteins, we designed a combinatorial multi-epitope protein as a potential vaccine construct. The binding stability and molecular interactions were checked by docking with Toll-like receptors TLR2, TLR4, and TLR5 and molecular dynamic studies further confirmed the stability of vaccine construct-receptor complex. Finally, immune simulation results demonstrated a human immune response to the vaccine construct, which was found to confer reliable short-term and long-term immunity. As the outcome of this study, we hereby predict MarVax as a highly antigenic and stable multi-epitope vaccine against the Marburg Virus and can be further studied and validated using *in vitro* and *in vivo* models.

Methods

Retrieval of the primary data

The primary information about the five essential proteins encoded by the MARV genome, including the glycoprotein (GP) (UniProtKB ID: P35253), RNA-dependent RNA polymerase (RdRp) (UniProtKB ID: P31352), viral proteins VP24 (UniProtKB ID: P35256), VP35 (UniProtKB ID: P35259), and VP40 (UniProtKB ID: P35260), were retrieved from UniProtKB (<https://>

www.uniprot.org/). The Protein Data Bank from the RCSB server (<https://www.rcsb.org/>) was used to retrieve the protein structures for GP (PDB ID: 5UQY), VP25 (PDB ID: 4OR8), and VP40 (PDB ID: 5B0V). Robetta (<https://rosetta.bakerlab.org/>) [25–33] and Swiss Model (<https://swissmodel.expasy.org/>) [34] were used respectively, to model the protein structures of RdRp and VP35. Both the servers perform homology modeling of query protein using a suitable protein as a template which is having high sequence homology, and structure is available in RCSB-PDB. Lastly, the best-modeled structure of the RdRp and VP35 generated by the servers was subjected for structural validation by utilizing the web server called PROCHECK (<https://saves.mbi.ucla.edu/>) [35, 36]. The conformational stability of modeled structures was analyzed by evaluating the Ramachandran Plot for RdRp and VP35. Finally, each protein model was edited in PyMOL ver2.4 and prepared for further studies.

T-cell and B-cell epitope prediction, screening, and sorting

All the five proteins namely GP, RdRp, VP24, VP35, and VP40 were predicted for the presence of T-cell epitopes. The TepiTool server (<http://tools.iedb.org/tepitool/>) [37–42] of the IEDB analytical resource tool was used to predict MHC class I restricted epitopes for GP, RdRp, VP24, VP35, and VP40. The list of representative alleles from various HLA super-types was selected, and the percentile rank was set to the value of one for the prediction of MHC class I epitopes. The same server was also used for the prediction of MHC class II-restricted epitopes for all of these five proteins. For this prediction, a pre-selected panel of alleles covering the three human MHC class II isotypes which are HLA-DR, HLA-DQ, and HLA-DP was used, and percentile rank 10 was chosen. After screening the predicted epitopes based on percentile rank, peptide length, and conservancy levels, the top five epitopes for both MHC class I and class II were chosen for docking.

Similarly, the five MARV proteins were predicted for the presence of B-cell epitopes using the IEDB server ElliPro (<http://tools.iedb.org/ellipro/>) [43]. This program predicted both linear and discontinuous epitopes. For the development of a multi-epitope vaccine, a linear epitope is the best choice. Thus, the 3D structure of the protein was uploaded to this web server in PDB format. By averaging the protrusion index (PI) values over the residues of the anticipated epitopes, ElliPro assigns a score to each predicted epitope. The parameter was fixed by limiting the maximum amino acid length to ten residues and the minimum score of 0.5. The top five predicted

linear epitopes were chosen for docking after being filtered based on PI score, peptide length, and conservancy levels.

BLAST and multiple sequence alignment

Based on the conservation level of each protein sequence as assessed by BLAST in NCBI, five MARV strains were chosen. The five strains were the Musoke strain (1980) and Ravn strain (1987) from Kenya, Ozolin strain (1975) from South Africa, Popp strain (1967) from West Germany, and Angola Strain (2005). A multiple-sequence alignment tool ClustalOmega was used to align the five protein sequences of the chosen strains. The aligned sequences were rendered using ESPript3.0 (<https://esprpt.ibcp.fr/ESPript/ESPript/>), which highlighted the conserved regions of the protein.

Sorting of epitope and designing of the protein vaccine

The first step in designing a protein vaccine was to screen the predicted T-cell epitopes and B-cell epitopes present in the five MARV proteins. The T-cell epitopes predicted using the TepiTool web server were filtered based on the percentile rank and peptide length of the epitopes. In the same manner, the linear B-cell epitopes predicted using the Ellipro web server were filtered on the basis of the PI score and peptide length of the epitopes. Further, the filtered epitopes were sorted based on the alignment which was performed using ESPript3.0 to identify the conserved regions in the amino acid sequence of the proteins.

Based on the screened T-cell and B-cell epitopes, and the most conserved sequence found from the alignment, the best five T-cell and B-cell epitopes for all the five MARV proteins were sorted and a docked with TLRs to analyze the interacting property of these epitopes.

Docking analysis with Toll-like receptors (TLRs)

Toll-like receptors or TLRs are a family of proteins found on the surface of many cells, including the immune cells such as the macrophage and dendritic cells and they recognize antigens associated with the pathogen. In general, TLR1, TLR2, TLR4, TLR5, and TLR6 are the most commonly expressed TLRs in human cells. Among this cohort, TLR2, TLR4, and TLR5 are particularly significant.

TLR1 and TLR6 form heterodimers with TLR2 which is expressed on the surface of many cells, including monocytes, macrophages, dendritic cells, and endothelial cells. Similarly, TLR4 is also expressed on the surface of macrophages, dendritic cells, and epithelial cells. Both of them can recognize bacterial lipopolysaccharide (LPS). On the other hand, TLR5 is expressed on the surface of

cells such as intestinal epithelial cells, and it can recognize flagellin, a component of bacterial flagella.

Based on the above facts, three Toll-like receptors (TLRs), TLR2 (PDB ID: 3A79), TLR4 (PDB ID: 3FXI), and TLR5 (PDB ID: 3J0A) were obtained from PDB. Using the server MDockPeP (<https://zougrouptoolkit.missouri.edu/mdockpep/>), all of the chosen T-cell and B-cell epitopes of the five MARV proteins were initially docked with these three TLRs. This web server, for protein-peptide docking, models the peptide first, then globally and flexibly examines protein-peptide binding patterns, and lastly rates and scores the observed binding modes [44, 45].

Secondary structure prediction and molecular docking

Five of the best epitopes from both T-cells and B-cells were chosen based on the initial docking of all the selected peptides from the five proteins with the three TLRs. A GSAGSAGSA linker was then used to link these selected peptide sequences, and the synthetic protein was modeled in the Robetta server. The PROCHECK (<https://saves.mbi.ucla.edu/results?job=1247870&p=procheck>) and PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>) [46, 47] web servers were used to validate the predicted structure. Using the internet server Galaxy Refine (<https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>) [48, 49], the protein was optimized. Only one of the five models created by Galaxy Refine was chosen based on the structural validation by PROCHECK. After that, the chosen protein model with the best score was docked again with the three TLRs (TLR2, TLR4, and TLR5) using the internet server ClusPro2.0 (<https://cluspro.bu.edu/login.php>) [50, 51].

Physiological property prediction of the multi-epitope protein

Alongside protein-TLR docking, the physiological properties of the modeled protein were evaluated. Using ProtParam, Scratch Protein Predictor, and NetChop, the various vaccine parameters were assessed, including stability, molecular weight, isoelectric point (pI), antigenicity, solubility, and solvent accessibility. The ProtParam tool (<https://web.expasy.org/protparam/>) [52] provides theoretical information regarding the physicochemical properties of the protein, including stability, molecular weight, pI, Grand Average of Hydrophobicity (GRAVY) index, and others. The Scratch Protein Predictor site (<http://scratch.proteomics.ics.uci.edu/>) [53] contains a subset of tools, including ACCpro and ACCpro2.0 for solvent accessibility, SOLpro for solubility, and the ANTIGENpro for antigenicity evaluation. The Proteasomal Cleavage Prediction tool of IEDB analytical resources (<http://tools.iedb.org/netchop/>) [54, 55]

provides information about the proteasomal decay of the protein in graphical format. The data was cross-validated using DTU Healthtech's NetChop (<https://services.healthtech.dtu.dk/service.php?NetChop-3.1>), which generates neural network predictions for the cleavage sites of the human proteasome. To estimate the characteristics of this protein, we used the default parameters of these websites.

Molecular dynamics simulation and immune simulation

Protein in water molecular dynamics (MD) simulation of the protein-TLR5 complex was carried out using the internet server WebGro (<https://simlab.uams.edu/>) [56, 57] for the evaluation of binding stability, conformation changes, and interaction modes. For the simulation, the PDB file of the protein-TLR5 complex was uploaded and the simulation was performed by applying GROMOS96 43a1 force field. The complex structure was solvated in a triclinic box with a simple point charge (SPC) water model. Then, 0.15M NaCl was added as a neutralizer in the simulation system. The simulation was allowed to run for 50 ns in the NPT equilibrium type with a constant pressure (1.0 Bar) and temperature (300 k). One thousand frames per simulation was generated to determine the root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), solvent-accessible surface area (SASA), and average number of H-bonds in each frame.

Molecular dynamics (MD) simulation to determine the root mean square fluctuation (RMSF) of the vaccine construct-TLR5 complex was carried out using the internet server CABSflex2.0 (<http://biocomp.chem.uw.edu.pl/CABSflex2>). The cycle count was set to 50, whereas all other parameters were kept the same.

At last, the designed vaccine was also run through with an immune simulation using the C-ImmSim online web server (<https://kraken.iac.rm.cnr.it/C-IMMSIM/>) [58, 59] to evaluate how well this foreign antigenic protein would influence the immune response of the human host. This simulation was done targeting the A0101 and B0702 HLA alleles of MHC class I and the DRB1_0101 and DRB1_0701 HLA alleles of MHC class II. Ten microliters of simulation volume containing 1000 antigenic particles without lipopolysaccharide (LPS) was injected for analyzing the host's immune response for 350 days by setting simulation steps to 1050 (1 step = 8 h in real life).

Results

Conservation check and epitope selection

The highly conserved portions of a protein become a suitable target for the vaccine designed for a virus that is evolving. The conserved regions of the essential viral proteins were identified using sequence alignment in

ESPrpt3. The most conserved proteins were found to be VP35 and VP40, followed by VP24, RDRP, and GP which exhibited the least level of conservation.

Using the TepiTool web server, T-cell specific epitopes were predicted which gave the list of T-cell epitopes for each of the five essential MARV proteins with percentile ranks ranging from 1 to 0.01. Similarly, B-cell-specific epitopes predicted using the ElliPro web server generated a list of epitopes with scores based on the protrusion index (PI) value and located linear B-cell epitopes with PI values ranging from 1.0 to 0.5.

Five epitopes for each T-cell MHC class I and class II and B-cell were selected after the predicted epitopes were screened based on peptide length, sequence conservation, and prediction scores. The selected epitopes are listed in the Supplementary tables (1), (2), and (3). All the selected epitopes were then employed for docking analysis with three TLRs.

Molecular docking analysis

The three TLRs were used for initial docking with the chosen antigenic peptides, which are listed in Supplementary tables (1), (2), and (3). For each submission, the docking process generates ten docking models, with the best model selected based on the binding energy. A total of 2250 docked models were projected, of which the binding energies of the top 225 docking models for T-cell and B-cell combined, ranging from - 12.1 to - 6.2 kcal/mol and from - 13.5 to - 6.0 kcal/mol, respectively, were recorded. Based on the lowest binding energy, five docked models both from T-cell and B-cell including all of the MARV proteins and three TLRs were selected from a total of 225 peptide-TLR complexes. Table 1 lists the selected dock complexes alongside their peptide sequences and binding energies.

Multi-epitope vaccine construction and structure modeling

A multi-epitope vaccine MarVax was created by assembling the selected epitopes listed in Table 1 and additional regions on both N- and C-termini to maintain the local fold of the region. The GSAGSAGSA linker has been used to link the epitopes one after the other. The glycine-serine-alanine (GSA) is a preferred amino acid sequence to utilize as a linker to connect two peptides. In general, polar uncharged or charged residues are the preferred amino acids [60]. They should also be flexible to allow the adjoining protein domains to fold and move properly with respect to one another. According to studies by Agros [61] and George and Heringa [62], glycine, serine, and alanine had propensities of 1.25, 1.46, 1.05, and 0.835, 0.947 and 0.964, respectively. The result greater than 1 (> 1) denotes that these amino acids are found in high numbers in the linker. Three repeats of GSA serve to prevent steric hindrance between the peptides and also give flexibility together with glycine and serine. The generated protein sequence is:

MGSAGSAGSAGSALRIWSVQEDDLAAGLSWIPFF-GPGIEGLYTAVLIKNQNNLVCRLRRLANQTAK-SLELLLRVTTEERTFSLINRHAIDFLLTRWGSAGSAG-SATCTVDVANFLRAYSWSDVLKKGKRLIGAGSAG-SAGSANLSAKDLALLLFTHLPGNNTPFHILAQVL-SKIAYKSGKSGGSAGSAGSAEPFLALRILLGVALK-DQELQQSLIPGFRSIVHMLSEWLLLEVTSAGSAG-SAGSAKLTPQQYCELFSLQKHWGHPVLYGSAGSAG-SAMWDSSYMQQVSEGLMTGKGSAGSAGSSYT-ITQFTHNGQKFVRVNRGL

Robetta was used for *ab-initio* modeling of the multi-epitope vaccine construct MarVax, and GalaxyRefine was used to refine the modeled vaccine construct. Additionally, the modeled structure was validated using two servers namely PROCHECK and PSIPRED. Figure 2 shows

Table 1 Binding energy of the selected T-cell and B-cell epitopes

	Peptide-TLR complex	Peptide sequence	Binding energy (kcal/mol)	HLA allele
T-cell	VP24-TLR2	¹¹⁸ IVHMLSEWLLLEVTS ¹³²	- 9.6	HLA-DRB1*01:01
	GP-TLR5	⁵⁷⁹ EERTFSLINRHAIDF ⁵⁹³	- 9.1	HLA-DRB1*07:01
	RDRP-TLR4	³⁶⁶ QQYCELFSLQKHWGH ³⁸⁰	- 11.5	HLA-DRB1*11:01
	VP35-TLR5	¹ MWDSSYMQQVSEGLM ¹⁵	- 12.1	HLA-DQA1*05:01 HLA-DQB1*02:01
	VP40-TLR5	¹¹² QFTHNGQKFVRVNR ¹²⁶	- 9.3	HLA-DRB3*02:02
B-cell	GP-TLR5	⁵⁴⁸ LIKNQNNLVCRLRRLANQ ⁵⁶⁵	- 10.9	---
	RDRP-TLR4	¹¹¹⁵ FLRAYSWSDVLKKGKRL ¹¹³⁰	- 13.5	---
	RDRP-TLR5	¹¹¹⁵ FLRAYSWSDVLKKGKRL ¹¹³⁰	- 9.8	---
	VP35-TLR2	²⁰⁸ LSAKDLALLLFTHLPGNNT ²²⁶	- 10.9	---
	VP24-TLR2	¹⁰⁴ DQELQQSLIPGFRSIVHM ¹²¹	- 11.6	---

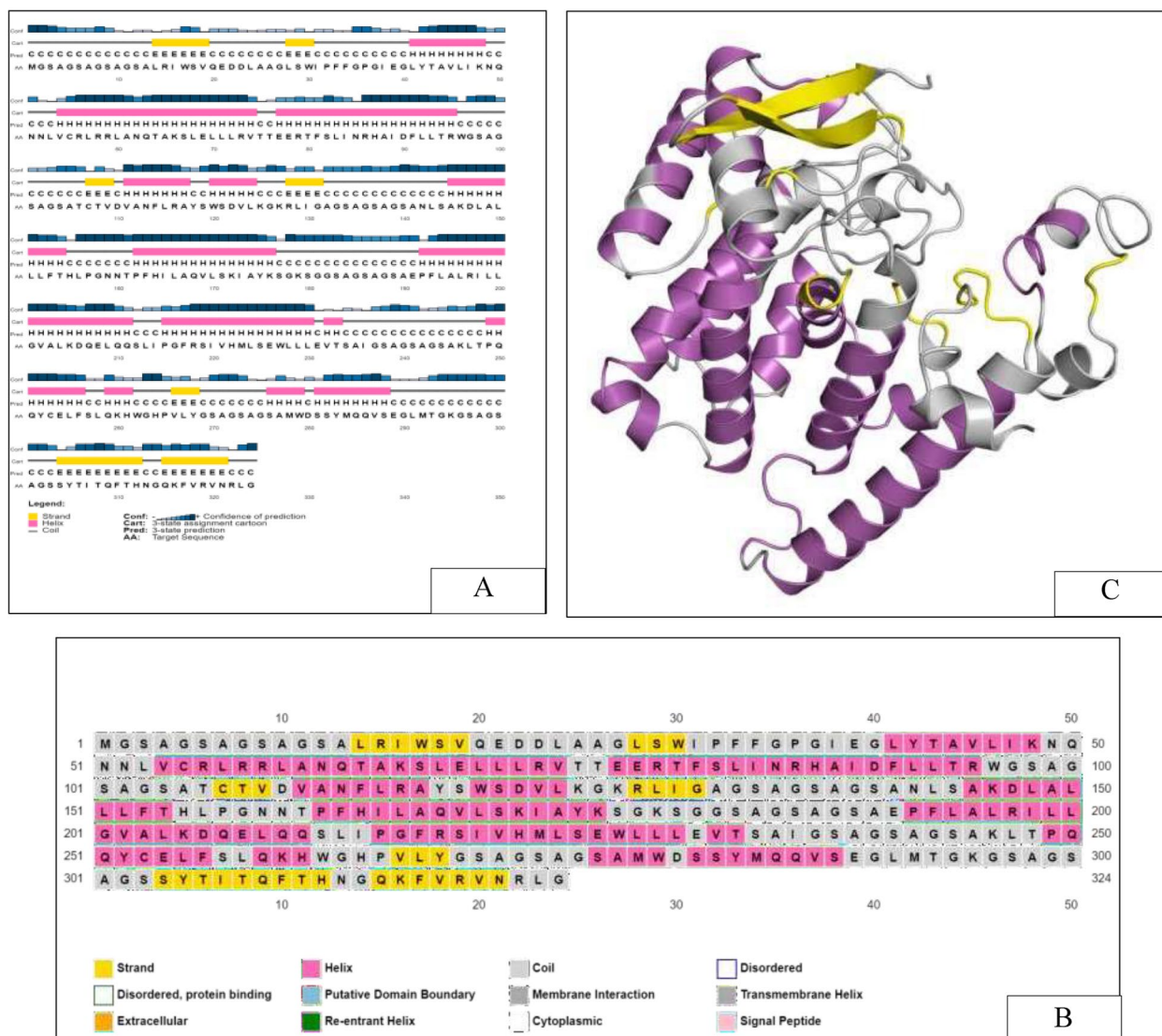


Fig. 2 Structural organization of multi-epitope vaccine molecule “MarVax”. (A) PSIPRED predicted secondary structure organization with confidence prediction. (B) Distribution of the secondary structure within the vaccine molecule. (C) Three-dimensional conformation of recombinant vaccine “MarVax”, modeled using Robetta web server

the secondary structure organization, as well as the 3D conformation of the designed protein vaccine, complementing the PSIPRED predicted data.

The Ramachandran plot of the modeled protein vaccine was generated using the PROCHECK web server and showed 97.8% residues residing in allowed regions, while 2.2% residues are in disallowed regions. The Ramachandran plot further exhibited 248 residues in the most favored region and six residues in disallowed regions, thereby supporting the correctness of the molecular conformation. The Z-score of the constructed multi-epitope protein was calculated using the PROSA server (<https://prosa.services.came.sbg.ac.at/prosa.php>) which came out

to be 7.98, indicating that the vaccine construct is structurally stable. The Ramachandran plot and Z-score plot of the modeled vaccine construct are given in Supplementary Figure 3A and 3B, respectively. The distribution of the B-cell and T-cell epitopes in the protein is illustrated in Supplementary Figure 1. The T-cell epitopes are colored blue, whereas the B-cell epitopes are colored red. The modeled protein has all of its epitopes on the surface, which makes it a suitable antigenic protein that immune cells can detect. The three TLRs were once more docked with this modeled protein vaccine to examine its binding properties.

Table 2 Binding energy scores of protein-TLR complex

Complex	Binding energy (kcal/mol)
Protein-TLR2	- 18.2
Protein-TLR4	- 19.8
Protein-TLR5	- 20.8

one from each TLR, were chosen after these models were screened based on their lowest weighted score. As shown in Table 2, this provided us with information indicating that the MarVax-TLR5 complex has the least negative binding score, followed by the MarVax-TLR4 complex and the MarVax-TLR2 complex.

MarVax-TLR5 complex is stabilized by eight potential hydrogen bonds between Asp:147 and Val:648, Asn:142 and Leu:652, Ala:141 and Phe:653, Ala:138

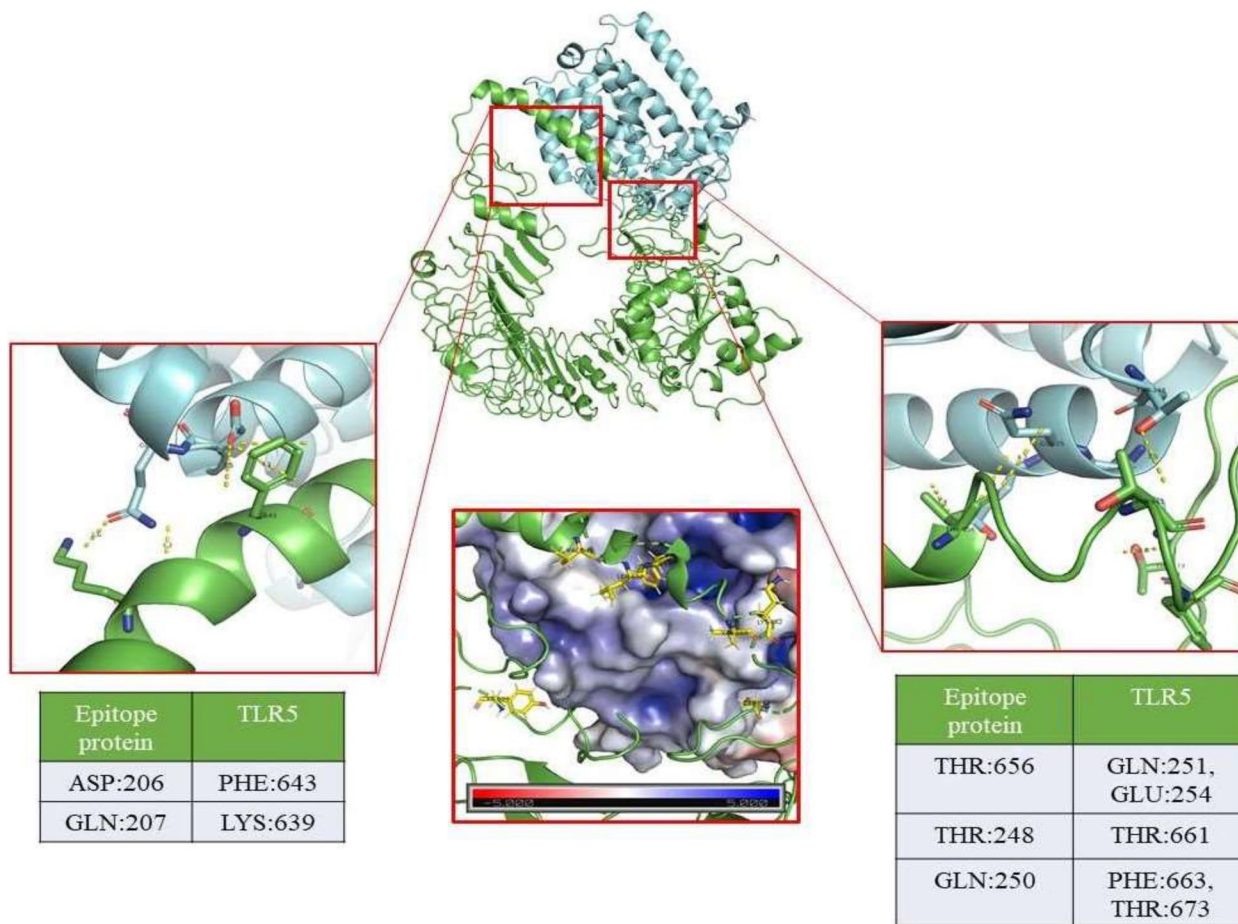


Fig. 3 Molecular interaction of recombinant vaccine MarVax-TLR5. TLR5 is highlighted in green color; the MarVax in cyan color. The electrostatic interaction between the vaccine construct and TLR5 was developed by the APBS wizard. Image generated using the PyMOL Molecular Graphics System, Version 2.4, Schrödinger, LLC

Efficacy analysis of recombinant vaccine MarVax by docking with TLRs

To evaluate the immunogenic ability of the vaccine molecule, the modeled multi-epitope protein was docked with the three TLRs. The web server Clus-Pro2.0 was used to conduct this protein-protein docking. The docking of this multi-epitope protein with the three TLRs produced a total of 30 models as each submission yielded 10 models. The top three models,

and Val:660, Glu:39 and Lys:662, Ile:38 and Lys:662, Ser:137 and Cys:670, and Tyr:175 and Tyr:609 as shown in Fig. 3. The complex’s electrostatic analysis revealed the presence of hydrophobic and charged residues close to the interacting region. This suggests the possibility of potential hydrophobic interactions, weak Van der Waals interactions, and Pi-interactions. Supplementary Figures 2A and 2B further show the interaction of the vaccine complex with TLR 2 and TLR 4.

Table 3 Physicochemical properties of the multi-epitope protein

Property	Value
Total amino acids	324
Molecular weight	34 kDa
GDT-HA	0.9877
Rama favored	94.1%
Clash score (C-score)	14.0
pI	9.49
Antigenicity	0.520688
GRAVY	0.09
Z-score	7.98
Proteasomal cleavage site	110
Estimated half-life (human reticulocytes)	30 h

Vaccine property analysis

The modeled recombinant vaccine consists of 324 amino acids with 34 kDa molecular weight. The Global distance test-high accuracy (GDT-HA) scores, which range from 0 to 1, are the overall indicators of how well a predicted model matches the experimental structure (with one corresponding to a maximum accuracy) [63, 64]. The GDT-HA value of this protein is 0.9877, which indicates the correctness of the model. The isoelectric point of the recombinant vaccine is 9.49, and its predicted antigenicity is 0.520688. The hydrophobicity of a protein is denoted by the grand average of hydropathicity index (GRAVY), which calculates the sum of the hydropathy values of all the amino acids divided by the length of the sequence. The greater protein solubility is associated with a lower GRAVY index. GRAVY index of MarVax was estimated to be 0.09. Proteins with a hydrophobicity score (arbitrary unit) less than 0 are more likely to be globular (hydrophilic), while proteins with a score greater than 0 are more likely to be membranous (hydrophobic) [65]. DeepTMHMM server (<https://dtu.biolib.com/DeepTMHMM>) was used to predict the transmembrane helices of the protein which showed that the protein had no transmembrane helices. Both these GRAVY index values and trans-membrane predictions validate the globular three-dimensional model of the vaccine. To check the nobility of the vaccine, its sequence was analyzed for structural homology with humans using pBLAST which showed 0% similarity with any human protein, thereby ruling out any possibility of autoimmune stimulation. Half-life of the vaccine was estimated to be 30 h with about 110 Proteasomal cleavage sites. Table 3 lists the physicochemical characteristics of the designed protein vaccine.

Molecular dynamics simulation and immune simulation study

The molecular dynamics simulation was performed for 50 ns to understand the dynamics stability associated with the complex formation of the vaccine molecule with TLR5. The simulation gave the RMSD of the complex, which when analyzed; it was found that the bound complex is stable with a very small deviation. The interactions were found to be stable between the vaccine and the TLR5 within this time frame. The RMSD graph of this molecular dynamics is illustrated in Fig. 4A. The average Radius of Gyration (R_g) value of the MarVax-TLR5 complex was 3.75 nm during this 50 ns simulation. The solvent-accessible surface area (SASA) of the protein-protein complex was estimated to be 525 nm².

To understand the stabilization of vaccine molecule upon complex formation in terms of RMSE, a MD simulation of the protein complex was performed. The comparable RMSE results were obtained by running the simulation of the apo-protein and the complex protein independently. The fluctuations in the amino acid residues are represented in the form of the graph shown in Fig. 4B. The vaccine molecule in the protein-TLR5 complex is found to be stable and exhibits little atomic fluctuation. Additionally, TLR5 is also structurally stable following its interaction with the multiple-epitope protein, validating the structural integrity of the MarVax-TLR5 complex.

The immune simulation of the recombinant vaccine molecule was performed, which demonstrated a high IgG, IgM, IFN- γ , and IL-2 production upon vaccination, as shown in Fig. 5. Prominently, IgG and IgM were present in blood serum with a peak in the first 30 days and then gradually decreased. IFN- γ remained at a high peak for 30 days, followed by a progressive decrease. This denotes that the vaccine molecule confers an active immunity lasting for almost 2 months. Following vaccination, there was a rise in the population of both cytotoxic T-cells and natural killer cells. Furthermore, B-cell, helper T-cell, cytotoxic T-cell, and natural killer (NK) cell production increased.

Figure 6 illustrates a graph demonstrating the proliferation of immune cells and their concentration in mm³. The graph demonstrates the consistency of IgM-releasing B-cells and memory B-cells which persisted at high peaks for almost 60 days and steadily declined to 100 cells/mm³ on day 300. Furthermore, helper T-cells peaked within 35 days and remained almost consistent for 350 days. This validates the vaccine construct's ability to confer an immediate immune response along with lasting long-term immunity. The levels of NK cells and cytotoxic T-cells also remained at considerably high levels, further

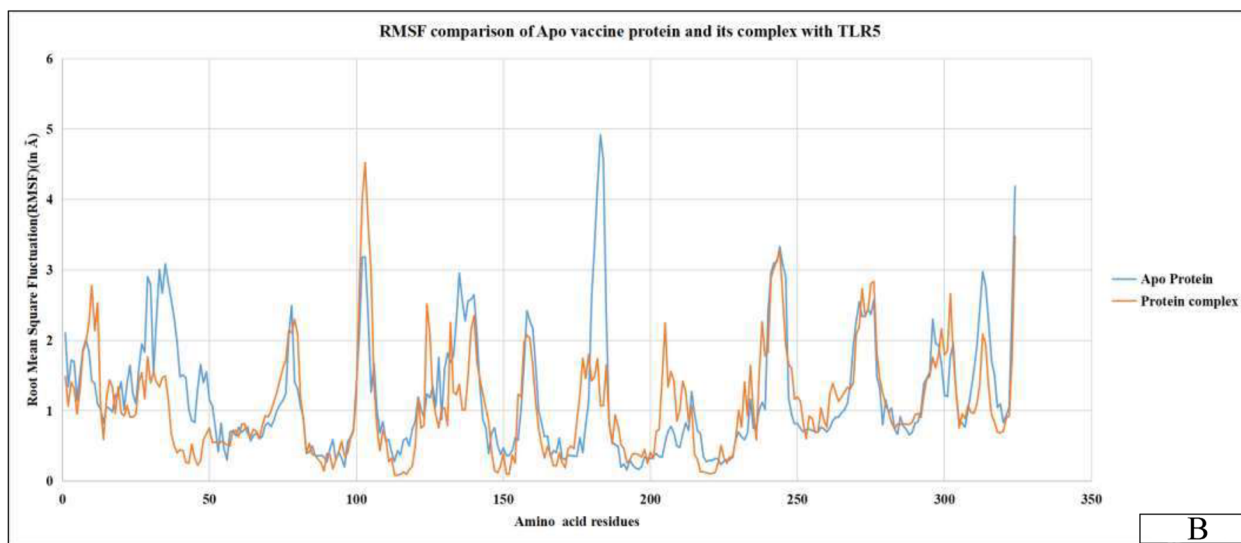
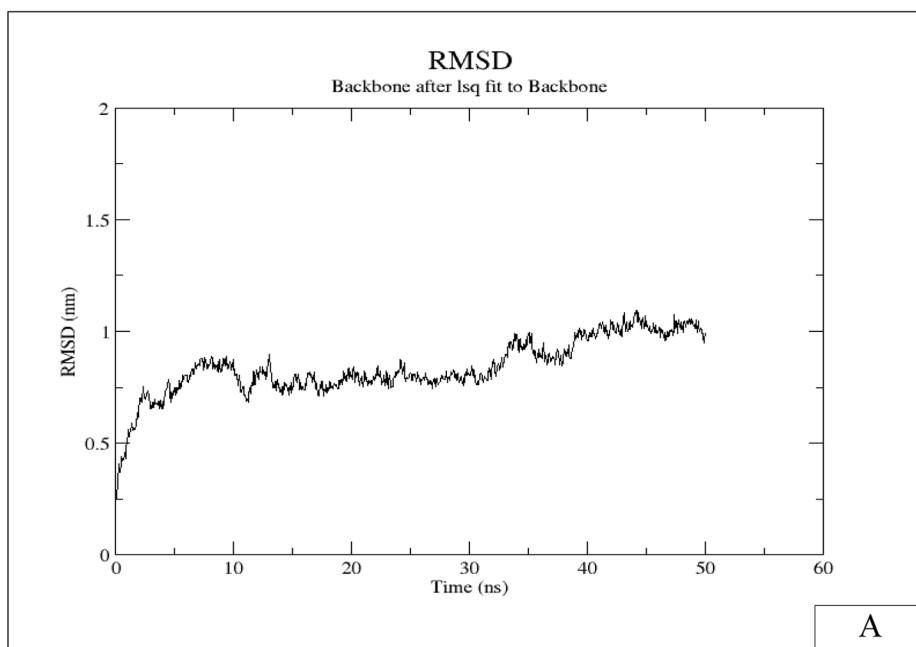


Fig. 4 Molecular dynamics analysis of vaccine molecule MarVax-TLR5 complex. **(A)** RMSD analysis of the MarVax-TLR5 complex for 50 ns **(B)** RMSF analysis of Apo vaccine (blue) and MarVax-TLR5 complex states (orange) showed significant stabilization in many regions

eliciting the vaccine construct’s efficacy in mediating immunity against MARV infection.

Discussion

A vaccine should induce an infection-induced natural immunity after a long-lasting adaptive immunity. Previously reported epitope-based vaccine constructs against MARV are mostly based on surface-exposed viral proteins. As delineated in studies conducted on the efficacy of vaccines against the SARS-Cov-2 virus [24], a conspicuous correlation becomes apparent between

elevated mutation rates and heightened pathogenicity of viral infection, accompanied by the concomitant attenuation of the long-term relevancy of vaccine constructs. Similarly, the high mutation rates reported in VP40 and NP-VP intergenic regions [2] might lead to a potential loss of antigenicity of vaccine constructs based on these proteins over a period of time. Hunegnaw R et al. [66] reported about clinical trial of a single-shot ChAd3-MARV vaccine on non-human primates resulting in production of IgG specific for MARV-GP, which is susceptible to moderate levels of mutations [2] and thereby,

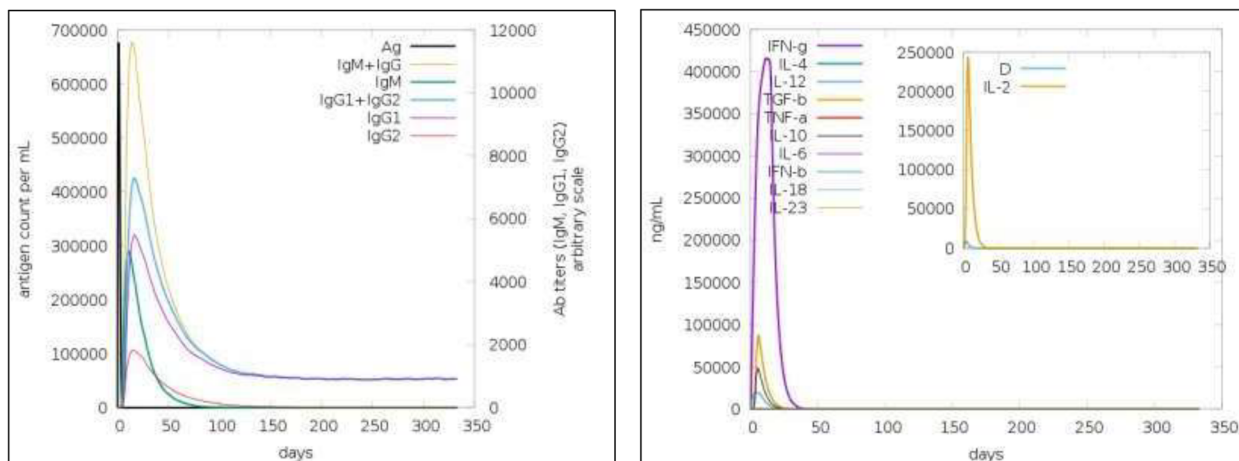


Fig. 5 Production of immunoglobulin (Ig) and cytokines against the immunogenic vaccine molecule MarVax

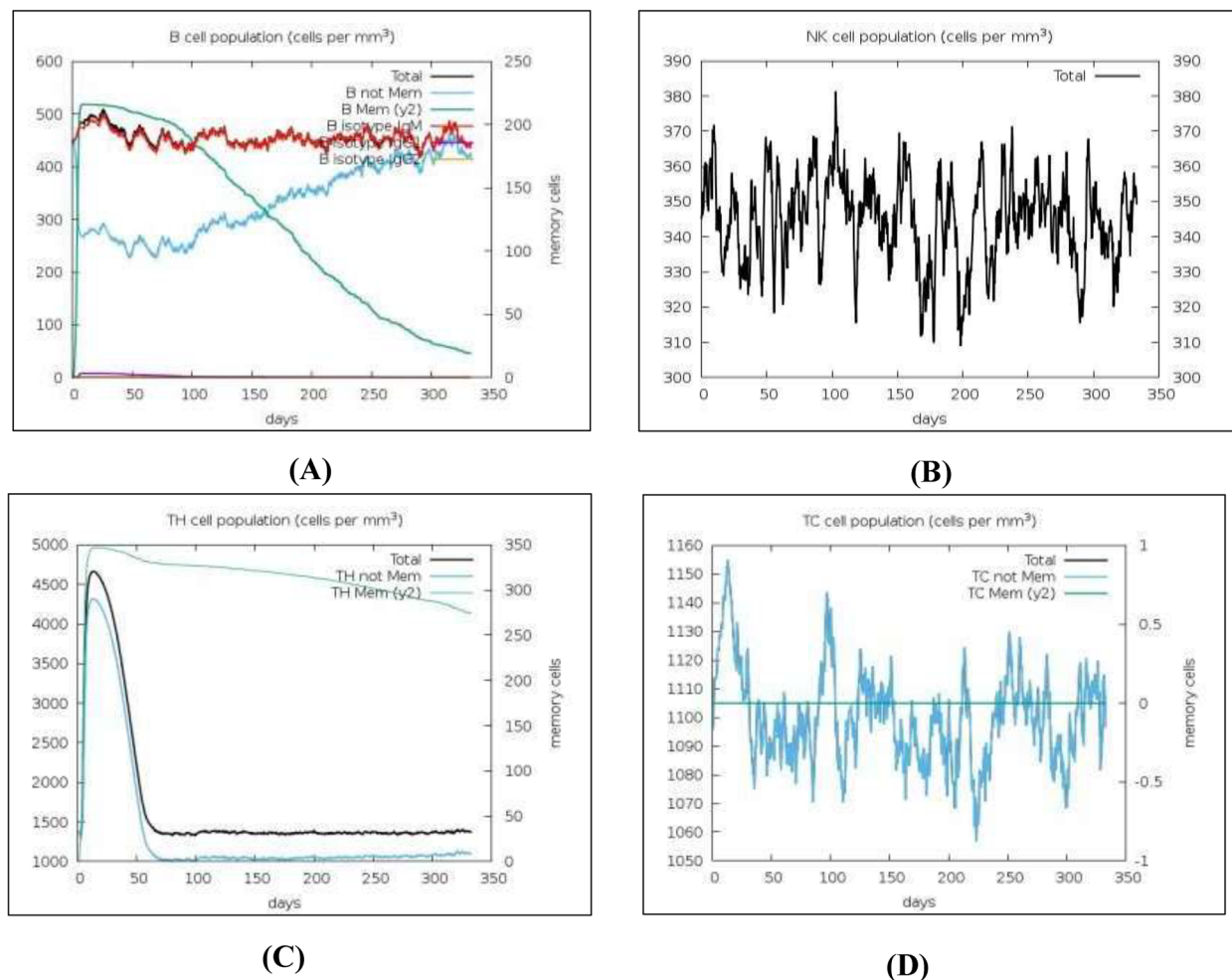


Fig. 6 Production of immune cells against the antigenic protein. (A) B-cell population per mm^3 . (B) NK cell population per mm^3 . (C) T_H -cell population per mm^3 . (D) T_C -cell population per mm^3

potentially impinges on the sustained efficacy of the vaccine in the long term. In our study, the predicted MHC-I- and MHC-II-restricted T-cell epitopes as well as B-cell epitopes for antigenic targets were screened based on two main parameters: antigenicity score and conservancy levels. The selected T-cell and B-cell epitopes had average antigenicity scores of > 0.5 and were almost 100% conserved in five reported strains of MARV, which signifies that the epitopes are highly antigenic and are highly conserved. Hence, the vaccine construct is capable to maintain its antigenic integrity, notwithstanding the elevated mutation rate of VP40 and VP35, as well as the moderate mutation rate of GP [23]. All these factors consolidate our work pioneering, as we employed epitopes encompassing conserved regions of five endogenous and exogenous viral proteins, namely, GP, VP24, VP35, VP40 and RdRp, which indicate towards long term stability and efficacy of the vaccine construct in mediating an immune response against MARV infection.

Epitopes used for the combinatorial vaccine construct were selected after further screening based on docking studies, hydrophobicity, allergenicity, and other parameters. The final modeled multi-epitope vaccine construct was found to be highly antigenic and non-allergenic. The modeled protein construct has 324 amino acids and an average molecular weight of 34 kDa and is globular in nature. A Ramachandran favored score of 97.8 %, C-score of 14, and Z-score of 7.98 further consolidate the structural confidence and stability of the protein. According to WHO guidelines, a suitable vaccine candidate should have less than 1 transmembrane helices. Our designed vaccine construct was devoid of any transmembrane helix, as predicted using the DeepTMHMM server. This suggests ease of expression and purification of the protein vaccine. The estimated half-life in mammalian reticulocytes was 30 h, which is sufficient for generating an immune response. Hence, the results demonstrate that it is a strong antigenic vaccine protein that can efficiently stimulate innate immunity.

To analyze the molecular interaction of vaccine construct MarVax, molecular docking was carried out with TLR 2, 4, and 5. TLRs 2 and 5 are surface-exposed and recognize various PAMPs, and TLR4 plays an important role in amplification of inflammatory response and lipopolysaccharide recognition. Hence, interactions with these TLRs can indicate a potential offset of a stable inflammatory response. The interaction of the vaccine construct with TLR 5 had the strongest binding score of $- 20.8$ kcal/mol due to hydrogen bond formations between eight residues. The binding energy was moderately compared to the previously modeled vaccines [67], which might elicit a higher affinity and efficiency in mediating immune response.

The final molecular dynamics simulation showed that the protein, upon interaction with the cell surface TLRs, maintains its stability over time. Immune simulation-predicted immune responses of the host body towards the antigenic protein revealed the production of antibodies, memory B-cells, cytokines, and other immune cells within 2 weeks of vaccination, with antibody-releasing B-cells persisting for almost 60 days following the first vaccine. This further indicated that the vaccine molecule MarVax confers stable long-term immunity against the viral infection and is effective against different strains of virus due to the high conservancy of epitopes. This protein's immunological simulation study provided further evidence that it has the potential to be a stable and robust antigenic vaccine protein to combat the deadly infection. Hence, based on various physicochemical and molecular dynamic interaction-based studies, we can say that the designed multi-epitope protein construct MarVax can be a stable, specific, and antigenic vaccine against the Marburg virus, which needs to be further consolidated by *in vitro* and *in vivo* studies for approval.

Conclusion

Recurring outbreaks and fatality rates going as high as 88% have put Marburg virus disease at the forefront of research for the cure against the viral infection. Categorization of MARV as a priority pathogen A by The National Institute of Allergy and Infectious Diseases (NIAID) and category A bioterrorism agent by the Centers for Disease Control and Prevention (CDC) further necessitate the commercialization of vaccine and therapy against MARV infection. While previous studies have predicted multi-epitope vaccines, a vast majority of them do not address the long-term antigenicity of the vaccine construct, which might be affected owing to high mutation rates in certain viral proteins.

The multi-epitope vaccine designed by us combines epitopes from GP, VP24, VP35, VP40, and RdRp and is highly antigenic and immunogenic. Docking studies and results exhibited by molecular dynamics establish the vaccine construct as a stable molecule with higher affinity as compared to predecessors. The immune simulation of this vaccine revealed a strong immune response with significant immune cell and cytokine secretion within 2 weeks after vaccination, and the immune system being active for up to 60 days afterward. The population of memory B-cells, helper T-cells, and NK cells also remained stable for up to 350 days, establishing the vaccine construct's efficacy in generating long-term stable immunity. Hence, the designed multi-epitope MarVax shows a strong potential to be an effective vaccine against MARV and is suitable for *in vitro* and *in vivo* validation.

Abbreviations

MVD	Marburg virus disease
MARV	Marburg virus
EBOV	Ebola virus
GP	Glycoprotein
NP	Nucleoprotein
VP	Viral protein
RdRp	RNA-dependent RNA polymerase
TLR	Toll-like receptor
CTL	Cytotoxic T lymphocyte
HTL	Helper T lymphocyte
IEDB	Immune epitope database
PDB	Protein Data Bank
MD	Molecular dynamics
RMSD	Root mean square deviation
SASA	Solvent-accessible surface area
RMSF	Root mean square fluctuation

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s43141-023-00575-w>.

Additional file 1: Supplementary Figure 1. Epitope distribution in the modeled multi-epitope vaccine. B-cell and T-cell epitopes are highlighted in red and blue color, respectively. **Supplementary Figure 2.** Protein-TLR2 interaction. TLR is highlighted in green color, side chain in yellow while protein in cyan color. **Supplementary Figure 3.** Protein-TLR4 interaction. TLR is highlighted in green color, side chain in yellow while protein in cyan color. **Supplementary Figure 4.** Ramachandran plot of vaccine construct molecule, validates using the PROCHECK server **Supplementary Figure 5.** Z-score plot of vaccine construct molecule.

Additional file 2: Supplementary Table 1. T-cell MHC class I epitopes. **Supplementary Table 2.** T-cell MHC class II epitopes. **Supplementary Table 3.** B-cell epitopes.

Acknowledgements

We would like to thank all the web servers, softwares, and open access data bases used in this study.

Data availability

All relevant data generated and analyzed during this study are included in the article and its supplementary information files.

Authors' contributions

Kuntal Pal did conceptualization, study design, formula analysis, investigation, resource, writing, and editing of the original draft, visualization, supervision, project administration, and funding acquisition. Bishal Debroj and Sribas Chowdhury did methodology and study design, software validation, investigation, resource, writing original draft, and visualization.

Funding

This work was supported by the Department of Science and Technology (DST)-SERB grant, Government of India (File Number: SRG/2020/002215) awarded to Kuntal Pal.

Declarations**Conflict of interest**

The authors declare that they have no competing interests.

Ethics approval and consent to participate

No ethical issues are applicable in this work.

Consent for publication

Not applicable

Author details

¹Department of Biological Sciences, School of Life Science and Biotechnology, Adamas University, Barasat-Barrackpore Road, Kolkata, West Bengal 700126, India. ²Department of Biotechnology, School of Life Science and Biotechnology, Adamas University, Barasat-Barrackpore Road, Kolkata, West Bengal 700126, India. ³Cancer Biology Laboratory, Adamas University, Barasat-Barrackpore Road, Kolkata, West Bengal 700126, India. ⁴School of Biosciences and Technology (SBST), Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India.

Received: 20 May 2023 Accepted: 26 October 2023

Published online: 28 November 2023

References

- Zhao F, He Y, Lu H (2022) Marburg virus disease: a deadly rare virus is coming. *Biosci Trends* 16(4):312–316. <https://doi.org/10.5582/bst.2022.01333>
- Shifflett K, Marzi A (2019) Marburg virus pathogenesis-differences and similarities in humans and animal models. *Virology* 16(1):165. <https://doi.org/10.1186/s12985-019-1272-z>
- Kiley MP et al (1982) Filoviridae: a taxonomic home for Marburg and Ebola viruses? *Intervirology* 18(1–2):24–32. <https://doi.org/10.1159/000149300>
- Amman BR et al (2012) Seasonal pulses of Marburg virus circulation in juvenile *Rousettus aegyptiacus* bats coincide with periods of increased risk of human infection. *PLoS Pathog* 8(10):e1002877. <https://doi.org/10.1371/journal.ppat.1002877>
- Towner JS et al (2007) Marburg virus infection detected in a common African bat. *PLoS One* 2(8):e764. <https://doi.org/10.1371/journal.pone.0000764>
- Ristanovic ES et al (2020) A forgotten episode of Marburg virus disease: Belgrade, Yugoslavia, 1967. *Microbiol Mol Biol Rev* 84(2). <https://doi.org/10.1128/MMBR.00095-19>
- Bharat TA et al (2011) Cryo-electron tomography of Marburg virus particles and their morphogenesis within infected cells. *PLoS Biol* 9(11):e1001196. <https://doi.org/10.1371/journal.pbio.1001196>
- Welsch S et al (2010) Electron tomography reveals the steps in filovirus budding. *PLoS Pathog* 6(4):e1000875. <https://doi.org/10.1371/journal.ppat.1000875>
- Abir MH et al (2022) Pathogenicity and virulence of Marburg virus. *Virology* 13(1):609–633. <https://doi.org/10.1080/21505594.2022.2054760>
- Feldmann H et al (1992) Marburg virus, a filovirus: messenger RNAs, gene order, and regulatory elements of the replication cycle. *Virus Res* 24(1):1–19. [https://doi.org/10.1016/0168-1702\(92\)90027-7](https://doi.org/10.1016/0168-1702(92)90027-7)
- Becker S et al (1998) Interactions of Marburg virus nucleocapsid proteins. *Virology* 249(2):406–417. <https://doi.org/10.1006/viro.1998.9328>
- Muhlberger E et al (1999) Comparison of the transcription and replication strategies of Marburg virus and Ebola virus by using artificial replication systems. *J Virol* 73(3):2333–2342. <https://doi.org/10.1128/JVI.73.3.2333-2342.1999>
- Feldmann H et al (1991) Glycosylation and oligomerization of the spike protein of Marburg virus. *Virology* 182(1):353–356. [https://doi.org/10.1016/0042-6822\(91\)90680-A](https://doi.org/10.1016/0042-6822(91)90680-A)
- Kolesnikova L et al (2004) The matrix protein of Marburg virus is transported to the plasma membrane along cellular membranes: exploiting the retrograde late endosomal pathway. *J Virol* 78(5):2382–2393. <https://doi.org/10.1128/JVI.78.5.2382-2393.2004>
- Bamberg S et al (2005) VP24 of Marburg virus influences formation of infectious particles. *J Virol* 79(21):13421–13433. <https://doi.org/10.1128/JVI.79.21.13421-13433.2005>
- Kortepeter MG et al (2020) Marburg virus disease: a summary for clinicians. *Int J Infect Dis* 99:233–242. <https://doi.org/10.1016/j.ijid.2020.07.042>
- Paweska JT et al (2012) Virological and serological findings in *Rousettus aegyptiacus* experimentally inoculated with vero cells-adapted hogan strain of Marburg virus. *PLoS One* 7(9):e45479. <https://doi.org/10.1371/journal.pone.0045479>
- Kolesnikova L et al (2007) Budding of Marburgvirus is associated with filopodia. *Cell Microbiol* 9(4):939–951. <https://doi.org/10.1111/j.1462-5822.2006.00842.x>

19. Borchert M et al (2007) Use of protective gear and the occurrence of occupational Marburg hemorrhagic fever in health workers from Watsa health zone, Democratic Republic of the Congo. *J Infect Dis* 196(Suppl 2):S168–S175. <https://doi.org/10.1086/520540>
20. Kuhn JH (2008) Filoviruses, a compendium of 40 years of epidemiological, clinical, and laboratory studies. *Arch Virol Suppl* 20:13–360
21. Martini GA et al (1968) On the hitherto unknown, in monkeys originating infectious disease: Marburg virus disease. *Dtsch Med Wochenschr* 93(12):559–571. <https://doi.org/10.1055/s-0028-1105098>
22. Barman N et al (2022) Strategy to configure multi-epitope recombinant immunogens with weightage on proinflammatory response using SARS-CoV-2 spike glycoprotein (S-protein) and RNA-dependent RNA polymerase (RdRp) as model targets. *J Pure Appl Microbiol* 16(1):281–295. <https://doi.org/10.22207/JPAM.16.1.17.2022>
23. Wei H et al (2017) Deep-sequencing of Marburg virus genome during sequential mouse passaging and cell-culture adaptation reveals extensive changes over time. *J Sci Rep* 7(1):3390. <https://doi.org/10.1038/s41598-017-03318-3>
24. Zhou Z, Zhu Y, Chu MJFJ (2022) Role of covid-19 vaccines in sars-Cov-2 variants. *Front Immunol* 2273. <https://doi.org/10.1093/bioinformatics/bti125>
25. Soding J (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21(7):951–960. <https://doi.org/10.1093/bioinformatics/bti125>
26. Raman S et al (2009) Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins* 77(9):89–99. <https://doi.org/10.1002/prot.22540>
27. Yang Y et al (2011) Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates. *Bioinformatics* 27(15):2076–2082. <https://doi.org/10.1093/bioinformatics/btr350>
28. Kallberg M et al (2012) Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* 7(8):1511–1522. <https://doi.org/10.1038/nprot.2012.085>
29. Song Y et al (2013) High-resolution comparative modeling with RosettaCM. *Structure* 21(10):1735–1742. <https://doi.org/10.1016/j.str.2013.08.005>
30. Ovchinnikov S et al (2017) Protein structure determination using metagenome sequence data. *Science* 355(6322):294–298. <https://doi.org/10.1126/science.aah4043>
31. Yang J et al (2020) Improved protein structure prediction using predicted interresidue orientations. *Proc Natl Acad Sci U S A* 117(3):1496–1503. <https://doi.org/10.1073/pnas.1914677117>
32. Baek M et al (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373(6557):871–876. <https://doi.org/10.1126/science.abj8754>
33. Hiranuma N et al (2021) Improved protein structure refinement guided by deep learning based accuracy estimation. *Nat Commun* 12(1):1340. <https://doi.org/10.1038/s41467-021-21511-x>
34. Waterhouse A et al (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 46(W1):W296–W303. <https://doi.org/10.1093/nar/gky427>
35. Morris AL et al (1992) Stereochemical quality of protein structure coordinates. *Proteins* 12(4):345–364. <https://doi.org/10.1002/prot.340120407>
36. Laskowski RA et al (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8(4):477–486. <https://doi.org/10.1007/BF00228148>
37. Buus S et al (2003) Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach. *Tissue Antigens* 62(5):378–384. <https://doi.org/10.1034/j.1399-0039.2003.00112.x>
38. Nielsen M et al (2003) Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci* 12(5):1007–1017. <https://doi.org/10.1110/ps.0239403>
39. Lundegaard C, Nielsen M, Lund O (2006) The validity of predicted T-cell epitopes. *Trends Biotechnol* 24(12):537–538. <https://doi.org/10.1016/j.tibtech.2006.10.001>
40. Lundegaard C et al (2008) NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. *Nucleic Acids Res* 36(Web Server issue):W509–W512. <https://doi.org/10.1093/nar/gkn202>
41. Lundegaard C, Lund O, Nielsen M (2008) Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers. *Bioinformatics* 24(11):1397–1398. <https://doi.org/10.1093/bioinformatics/btn128>
42. Paul S et al (2016) TepiTool: a pipeline for computational prediction of T cell epitope candidates. *Curr Protoc Immunol* 114:18 19 1–18 19 24. <https://doi.org/10.1002/cpim.12>
43. Ponomarenko J et al (2008) ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics* 9:514. <https://doi.org/10.1186/1471-2105-9-514>
44. Yan C, Xu X, Zou X (2016) Fully blind docking at the atomic level for protein-peptide complex structure prediction. *Structure* 24(10):1842–1853. <https://doi.org/10.1016/j.str.2016.07.021>
45. Xu X, Yan C, Zou X (2018) MDockPeP: an ab-initio protein-peptide docking server. *J Comput Chem* 39(28):2409–2413. <https://doi.org/10.1002/jcc.25555>
46. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292(2):195–202. <https://doi.org/10.1006/jmbi.1999.3091>
47. Buchan DWA, Jones DT (2019) The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res* 47(W1):W402–W407. <https://doi.org/10.1093/nar/gkz297>
48. Kozakov D et al (2013) How good is automated protein docking? *Proteins* 81(12):2159–2166. <https://doi.org/10.1002/prot.24403>
49. Lee GR, Heo L, Seok C (2016) Effective protein model structure refinement by loop modeling and overall relaxation. *Proteins* 84(Suppl 1):293–301. <https://doi.org/10.1002/prot.24858>
50. Kozakov D et al (2017) The ClusPro web server for protein-protein docking. *Nat Protoc* 12(2):255–278. <https://doi.org/10.1038/nprot.2016.169>
51. Vajda S et al (2017) New additions to the ClusPro server motivated by CAPRI. *Proteins* 85(3):435–444. <https://doi.org/10.1002/prot.25219>
52. Wilkins MR et al (1999) Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* 112:531–552
53. Cheng J et al (2005) SCRATCH: a protein structure and structural feature prediction server. *J Nucleic acids research* 33(suppl_2):W72–W76. <https://doi.org/10.1093/nar/gki396>
54. Kesmir C et al (2002) Prediction of proteasome cleavage motifs by neural networks. *Protein Eng* 15(4):287–296. <https://doi.org/10.1093/protein/15.4.287>
55. Nielsen M et al (2005) The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* 57(1–2):33–41. <https://doi.org/10.1007/s00251-005-0781-7>
56. Bekker H et al (1993) Gromacs—a parallel computer for molecular-dynamics simulations. 4th International Conference on Computational Physics (PC 92). World Scientific Publishing
57. Abraham MJ et al (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *J SoftwareX* 1:19–25. <https://doi.org/10.1016/j.softx.2015.06.001>
58. Rapin N et al (2010) Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PLoS One* 5(4):e9862. <https://doi.org/10.1371/journal.pone.0009862>
59. Castiglione F et al (2021) From infection to immunity: understanding the response to SARS-CoV2 through in-silico modeling. *Front Immunol* 12:646972. <https://doi.org/10.3389/fimmu.2021.646972>
60. Chen X, Zaro JL, Shen WC (2013) Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev* 65(10):1357–1369. <https://doi.org/10.1016/j.addr.2012.09.039>
61. Argos P (1990) An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. *J Mol Biol* 211(4):943–958. [https://doi.org/10.1016/0022-2836\(90\)90085-Z](https://doi.org/10.1016/0022-2836(90)90085-Z)
62. George RA, Heringa J (2002) An analysis of protein domain linkers: their classification and role in protein folding. *Protein Eng* 15(11):871–879. <https://doi.org/10.1093/protein/15.11.871>
63. Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42(Web Server issue):W320–W324. <https://doi.org/10.1093/nar/gku316>
64. Mariani V et al (2013) IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests.

Bioinformatics 29(21):2722–2728. <https://doi.org/10.1093/bioinformatics/btt473>

65. Magdeldin S et al (2012) Murine colon proteome and characterization of the protein pathways. *BioData Min* 5(1):11. <https://doi.org/10.1186/1756-0381-5-11>
66. Hunegnaw R et al (2022) A single-shot ChAd3-MARV vaccine confers rapid and durable protection against Marburg virus in nonhuman primates. *J Sci Transl Med* 14(675):eabq6364. <https://doi.org/10.1126/scitranslmed.abq6364>
67. Sami SA et al (2021) Designing of a multi-epitope vaccine against the structural proteins of Marburg virus exploiting the immunoinformatics approach. *J ACS Omega* 6(47):32043–32071. <https://doi.org/10.1021/acsomega.1c04817>

Publisher's Note

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

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
