

Machine Learning Optimization Approach to Design Multi-Epitope Marburg Vaccine Construct

Shreyansh Suyash¹, Wajihul Hasan Khan², Priyasha Maitra¹,
Vinod Jangid¹, Parveen Punia³ and Avinash Mishra^{1*}

¹Growdea Technologies Pvt. Ltd., Gurugram., Haryana, India.

²All India Institute of Medical Sciences, New Delhi, India.

³Pt. Neki Ram Sharma Government College, Rohtak, Haryana, India.

<https://dx.doi.org/10.13005/bbra/3318>

(Received: 30 September 2024; accepted: 04 November 2024)

The Marburg virus (MARV) causes severe hemorrhagic fevers with life-threatening symptoms. A study aimed to design a multi-epitope vaccine (MEV) using immunoinformatic approaches for treatment for MARV infection. A comprehensive screening procedure was used to identify immunogenic protein sequences within seven crucial proteins from MARV that could trigger T-cell and B-cell responses. A computational analysis of these epitopes showed a non-allergenic nature and significant antigenicity, validating the structural parameters. The final construct of virus-like particle (VLP) was used for mutation using machine-learning model. A machine learning model, DeepPurpose framework was developed and trained to screen out the best vaccine construct/VLP sequence among all the generated sequences. Best variant VLP had the predicted IC₅₀ of 0.021 nM with the receptor TLR4. Model structures of the native and mutant VLP with prediction confidence scores of 96.2% and 88.5% were selected for molecular docking and molecular dynamic simulation to assess stability. RMSD of native construct ranged from 1.75 to 2 nm, while variant had 1.5 to 1.75 nm which was lower than the wild type, suggesting more stable conformation. The VLPs when bound with the toll-like receptor-4 (TLR4), plays a role in innate immunity. Designed VLP-TLR4 complex showed high stability post MD simulation of 500 ns and had strong average binding free energy (?G) of -520.13 (kcal/mol). The vaccine's stability helps it trigger a tailored immune response, making it an attractive candidate for viral neutralization strategies. The study showed a computational pipeline for designing and validating MARV multi-epitope vaccines using physics and machine learning. Additionally, the variant VLP exhibited favourable properties, suggesting its potential suitability for experimental validation, which could provide valuable insights. Nonetheless, the present study relies on in silico methodologies instead of in vivo or in vitro investigations, which is a limitation. This approach has promising applicability in the design of novel peptide vaccines against the MARV.

Keywords: Immunoinformatic; Marburg virus; Molecular Docking; Molecular Dynamics Simulation; Vaccine Design; Virus-like particle (VLP).

The Marburg virus (MARV) is classified within the Filoviridae family, specifically within the genus known as Marburg virus. This genus encompasses both the MARV and the Ravn virus¹.

The primary reservoir for Marburg infection is bats, whereas monkeys serve as the intermediate host. The transmission of the virus occurs through various mechanisms, including aerosol transmission, direct

*Corresponding author E-mail: avish2k@gmail.com



contact, and ingestion². Marburg virus Infection causes severe haemorrhagic fever outbreak in Germany and Belgrade, Yugoslavia (now Serbia) in 1967. The outbreak began from non-human primates that were introduced from Africa³. The two major outbreak of Marburg infection occurred in 1998 and 2004, at Durba, DRC and Uige Province, Angola respectively⁴. Currently, the outbreak of MARV disease was documented in August 2021 inside the Guéckédou prefecture in Guinea. A total of 173 individuals were identified as contacts, of which 14 were classified as high-risk contacts due to their level of exposure⁵. Additionally, in Feb 2023 the outbreak was occurred in Africa, which was officially reported from Equatorial Guinea, that further confirmed by the Institute Pasteur Laboratory in Dakar, Senegal. During this outbreak, the fatality rate of Marburg virus disease (MVD) reached a maximum of 88%⁶. This showed the Marburg virus infection is associated with significant mortality rates, and though multiple studies have been performed, there is devoid of specific treatment for this deadly virus. This remains a necessity for further progress in the drug discovery process pertaining to the sickness caused by Marburg infection. In recent times, there have been notable advancements in *in-silico* investigations, leading to the development of several prediction servers that employ specific algorithms for protein analysis. These servers, together with numerous *in-silico* tools, have proven to be valuable in predicting successful drugs against Marburg illness, as evidenced by multiple studies⁷. Immunoinformatics is also employed in the development of chimeric vaccines that are based on T cells and B cells epitopes targeting diverse diseases including MARV infection⁸.

Here, for the construction of a subunit vaccine on multiple epitopes against the MARV, ideal protein markers were selected. MARV is composed of seven key proteins^{9,10}. However, proteins like VP35 and VP40 have shown indication to trigger host immune response and thus considered ideal for vaccine construction⁶. Majorly, utilising T cell and B cell epitope regions derived from proteins has the potential to further enhance the development of a highly effective and widely applicable vaccine, which might serve as an active strategy to hinder the progression of the virus¹¹. In the present study, to develop a vaccine

candidate targeting the MARV, a selection process was applied to identify various antigenic epitopes to construct a virus-like-particle (VLP). A machine-learning based model was also developed to perform mutation and screen the best variant VLP. In an innovative approach to vaccine development targeting the optimization of a vaccine construct or VLP sequence, our study undertook the systematic modification of specific linker regions within the sequence. Mutations were added using a custom Python script.

A machine learning model was developed and improved using the DeepPurpose framework to explore the extensive range of potential vaccine candidates/VLPs created by these mutations. Furthermore, molecular dynamics simulation was performed for protein-protein interaction analysis of the constructed vaccine candidate against its toll-like receptor. These findings provide a foundation for the development of a subunit vaccine that can potentially show a robust immunological response against the MARV infection. The integration of advanced computational techniques in the selection and analysis of epitopes ensures a comprehensive approach towards vaccine design. The strategy employed can produce highly effective vaccine construct. The use of such *in-silico* methods serves as a testament to the evolving landscape of vaccine research, where precision and efficiency are paramount. Additionally, the selected epitopes were subjected to extensive immunoinformatic analysis, calculating factors that includes antigenicity, allergenicity, and conservancy, to validate their efficacy as vaccine candidates. This multi-faceted approach underscores the importance of a thorough and methodical process in vaccine development, majorly for the highly virulent pathogens, MARV.

MATERIAL AND METHODS

An effectual set of computational procedure was employed utilizing various immunoinformatic and other *in silico* approach, as illustrated in Figure 1 and elaborated in subsequent sub-sections.

Protein Selection and Analysis

Data from experimental epitope determination assays in MARV protein were utilised to select vaccine candidates on the BV-BRC (<https://www.bv-brc.org/>) server¹². This server listed

MARV proteins that have performed acceptable in experimental testing and may be targeted for epitopes. Protein sequences were obtained from UniProt¹³ using the Ids P27588 (Nucleoprotein), P35253 (Envelope glycoprotein), P35260 (Matrix protein VP40), P35259 (Polymerase cofactor VP35), P31352 (RNA-directed RNA polymerase L), and P35258 (Transcriptional activator VP30) and saved in FASTA format.

Prediction of Antigenicity, Allergenicity and Transmembrane Helices

The evaluation of antigenic properties for the extracted Marburg protein sequences was performed using the VaxiJen v2.0 platform¹⁴. Additionally, the AllergenFP v1.0¹⁵ platform was employed to determine allergenic potential. Furthermore, for insights into the transmembrane helices, the TMHMM - 2.0 platform was utilized¹⁶. Subsequently, three proteins were selected for epitope prediction based on their antigenicity prediction scores.

Epitopes Prediction and Screening

This study found numerous epitopes for three selected viral proteins. NetCTL v1.2 predicted CTL epitopes¹⁷. Next, the IEDB tool was used to find MHC-I alleles that can bind to a viral peptide¹⁸. Additionally, the IEDB server predicted HTL epitopes and alleles for the three selected proteins^{19,20}. Later, ABCpred predicted linear B lymphocyte (LBL) epitopes^{21,22}. The MHC allele-containing epitopes were tested for immunogenicity, antigenicity, allergenicity, and toxicity utilising platforms (<http://tools.iedb.org/immunogenicity/>), VaxiJen 2.0, AlgPred 2.0, and ToxinPred 2, mostly for CTL epitopes. Similar study was performed for the HTL epitopes. IFNepitope²³, IL4pred²⁴, and IL10pred²⁵ were utilised to predict IFN, IL4, and IL10 cytokines with threshold values of 0.2, 0.2, and 0.3, respectively. Immunogenicity, antigenicity, allergenicity, and toxicity were anticipated using web-based servers VaxiJen 2.0, AlgPred 2.0, and ToxinPred 2.0 to screen LBL epitopes.

Vaccine Construct Formulation

The design of the vaccine involved the incorporation of specific epitopes, including HTL, CTL, and LBL, derived from precisely selected MARV proteins. The TLR agonist (50s ribosomal protein L7/L12) referenced as NCBI - P9WHE3 was employed as an adjuvant²⁶. The L7/

L12 (P9WHE3) was combined to the vaccine's front using the bi-functional linker, EAAAK. Conversely, the chosen LBL, CTL, and HTL were integrated using Lys-Lys (KK), Ala-Ala-Tyr (AAY), and Gly-Pro-Gly-Pro-Gly (GPGPG) linkers, respectively²⁷. The AAY linker amplifies the immunogenic response of the multi-epitope vaccination²⁸. Livingston and colleagues (2002) conceived the GPGPG linker aiming to function as a flexible gap. The efficacy of GPGPG linker in triggering TH lymphocyte (HTL) reactions, crucial for designing a multi-epitope vaccine^{29,30}.

Machine Learning Guided Mutation

The vaccine construct/VLP sequence was taken and linkers from it were extracted out. Mutations were made at specific residues to serve as linkers to strengthen the interaction of VLP. Each variant was then examined for its affinity and the mutation was carried out by isolating all the linker residues. The linkers underwent mutation using a Python script. Linkers: EAAAK, AAY, GPGPG, KK. Variants created for each linker - Linker_1 – EAAAK generated 3200000 variants, Linker_2 – AAY generated 8000 variants, Linker_3 – GPGPG generated 3200000 variants, Linker_4 – KK generated 400 variants. Once the variants stretch were generated, they were screened based on their affinity using the trained ML model. A machine learning model using DeepPurpose framework³¹ was developed and trained to screen out the best vaccine construct/VLP sequence among all the generated sequences (Figure 3). The model architecture involves a dimension of hidden layers: 64, 32, train epochs: 150, learning rate: 0.001 and batch size: 16. This predictive model was trained on a dataset obtained from Skempi v2.0³², which contained 344 datapoints. The training dataset included protein and peptide sequences and the affinity between them. This dataset was divided into training, validation and test dataset with 70% of the data for training, 10% for validation and 20% for testing. The sequences were encoded using several encoders where Conjoint triad encoder provided by DeepPurpose outperformed. The affinity value was normalized before proceeding with the model training. The normalization of the affinity was achieved by calculating negative logarithmic values for each and dividing the resultant with corresponding peptide sequence length. The encoded sequences along with

normalized affinity scores were fed to the model for training and coefficient of determination (R^2) was calculated to evaluate the accuracy of the model. The multiple generated variant sequences were encoded and fed to the trained model to predict the best variant linker sequence among them. A list of best predicted linkers in each case was formed and based on this list multiple variants of vaccine construct was generated. The generated variants also included wild forms of the linkers to include combination effect of linkers in the study. The variants were fed to the trained model and best predicted variant was selected for further analysis. Figure 2 explained the pipeline for selecting the variant vaccine construct/VLP through ML.

Antigenicity, and Allergenicity Evaluation

The constructed vaccine/VLP for both wild and variant was put through multiple analyses as depicted in Figure 2. It started with amino acid composition analysis of the constructed vaccine sequence. Furthermore, several platforms were used to predict the antigenic and allergenic properties of the vaccine design, this includes Vaxijen v2.0, Scratch protein prediction platform, AllerTOP v. 2.0, and AllergenFP v.1.0.

3D Structure Modeling and Validation

The protein tertiary (3D) structure of wild and variant VLP was predicted using Alpha Fold³³. The predicted models underwent validation through the ProCheck online platform³⁴. The modelled structure was evaluated using ProCheck to assess the quality of the structure based on the fraction of residues in most favourable region of the Ramachandran plot.

Molecular Dynamic Simulation of Vaccine Construct

Based on the ProCheck results the wild and variant VLP was used for MD simulation study. The GROMACS 2021.2 software package was used to perform molecular dynamics simulations on both the VLPs. The simulation software uses the CHARMM36 force field for the parametrization^{35,36}. The protein was solvated within a cubic simulation box with recurring boundary parameters using the TIP3P hydration model³⁷, maintaining a solvent density at 0.997 g/L. The complete system was solvated at a pH of 7.4, after incorporating sodium (Na) and chloride (Cl) ions and setting the temperature at 310 K. Using the steepest descent technique, the complex's energy

was optimized over 50,000 cycles. The extended electrostatic interactions were conducted using the particle mesh Ewald technique³⁸. The simulation was performed over a duration of 300 ns using a V-scale thermostat and under constant pressure conditions using Parrinello-Rahman pressure coupling method. The simulated trajectory was captured at consistent spans of 10 ps of time frame to calculate various parameters including the root mean square deviations (RMSD), root mean square variations (RMSF), the surface area accessible to solvents (SASA) and principal component analysis (PCA). The post MD analysis was performed on the visual platform called "Analogue" developed by Growdea Technologies^{39,40} (<https://growdeatech.com/Analogue/>). Binding energy was calculated by utilizing the GROMACS plugin known as gmx_MMPBSA⁴¹. The MM/GBSA (Molecular Mechanics/Generalized Born Surface Area) approach was used for the determination of the binding free energy of the complex over the last 20 nanoseconds of the simulation.

Molecular Docking of Vaccine Construct

Structural coordinate for the TLR4 complexes was collected from the Protein Data Bank, with PDB ID: 4G8A. The extraneous heteroatoms and chains B, C, and D were then removed using the Pymol software⁴². Later, Swiss PDB viewer tool was used to fix the missing residues⁴³. Both mutated and original vaccine constructs (VLPs) were docked with TLR4 (PDB ID: 4G8A) using ClusPro tool⁴⁴. This application is a web-based automated protein-protein or peptide-peptide docking system. The docking programme assesses the potential surface complementarities of putative complexes. As a result of the clustering properties, a concise inventory of putative complexes is generated by the programme.

MD Simulation of Docked Complex

Furthermore, the docked complex of the best performing VLP with TLR4 was used for MD simulation of 500 ns to evaluate the stability and binding ability. MD simulation was performed with the same method as mentioned in the section 2.7. Additionally, the binding free energy and the energy contribution of the residues were also calculated using the same protocol (MM/GBSA) as mentioned in the section 2.7.

Molecular Dynamics Vaccine-TLR4 Complex

Simulation of Vaccine-TLR4 complex was performed using the same protocol as discussed in the section 2.7. However, here the simulation time for the production run was set as 500 ns.

RESULTS

Primary Analysis

A comprehensive search was performed on BV BRC database that resulted in total of 266 assays for the epitopes. Among the various

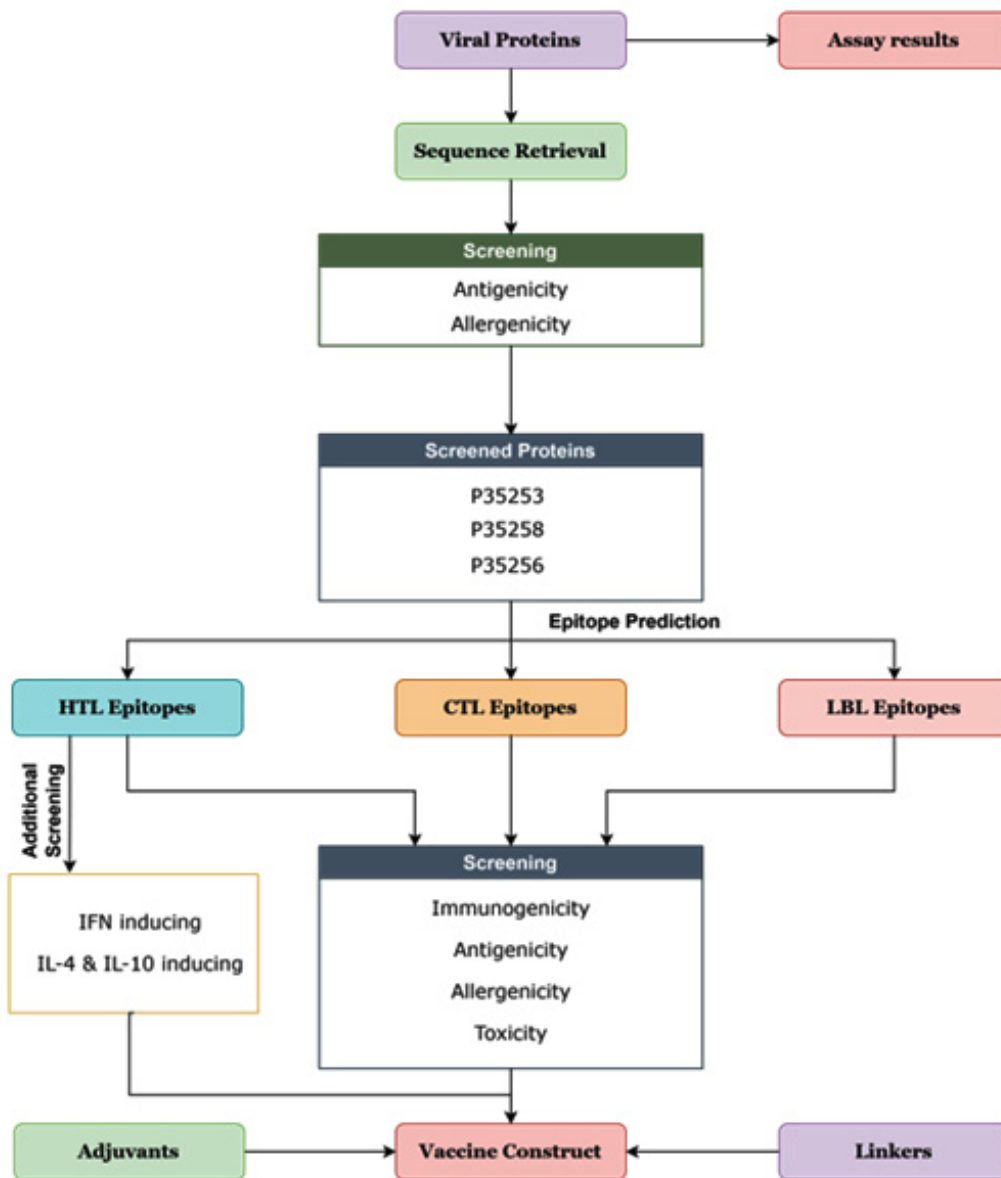


Fig. 1. Immunogenicity Screening and Epitope Prediction Process for Vaccine Development: This flowchart details the stages involved in selecting viral proteins for vaccine formulation, starting with sequence retrieval and initial screening for antigenicity and allergenicity. Screened proteins are then analyzed for epitope prediction, identifying helper T-lymphocyte (HTL), cytotoxic T-lymphocyte (CTL), and linear B-cell lymphocyte (LBL) epitopes. Further screening evaluates immunogenicity, antigenicity, allergenicity, and toxicity, leading to the final vaccine construct, with the integration of suitable adjuvants and linkers.

experimental assays conducted for determining these epitopes, a total of 57 assays demonstrated positive outcomes. The positive indications were subsequently classified into four distinct groups, distinguished by the precise viral components that they selectively targeted. The distribution of positive assays is: (a) viral envelope protein showed positive outcomes in a total of 22 assays, (b) viral nucleoprotein showed positive outcomes in 11 of the assays, (c) viral matrix protein showed a positive result in a 1 assay, and (d) viral polymerase showed positive indications in 23 cases. This study showed that envelope protein and polymerase had maximum number of assays for the epitope

determination. Polymerase is a heavy protein and most likely to find epitopes on it. However, envelope protein has significantly smaller sequence and the protein is on the surface of the virus and thus should be considered as most potential protein for vaccine design.

Protein Retrieval and Analysis

The MARV contains seven structural proteins: (1) nucleoprotein, (2) envelope glycoprotein, (3) matrix protein (VP40), (4) polymerase co-factor (VP-35), (5) RNA directed RNA polymerase L, (6) transcriptional activator (VP30), and (7) membrane associated protein (VP24). Later, the UniProt database was used to

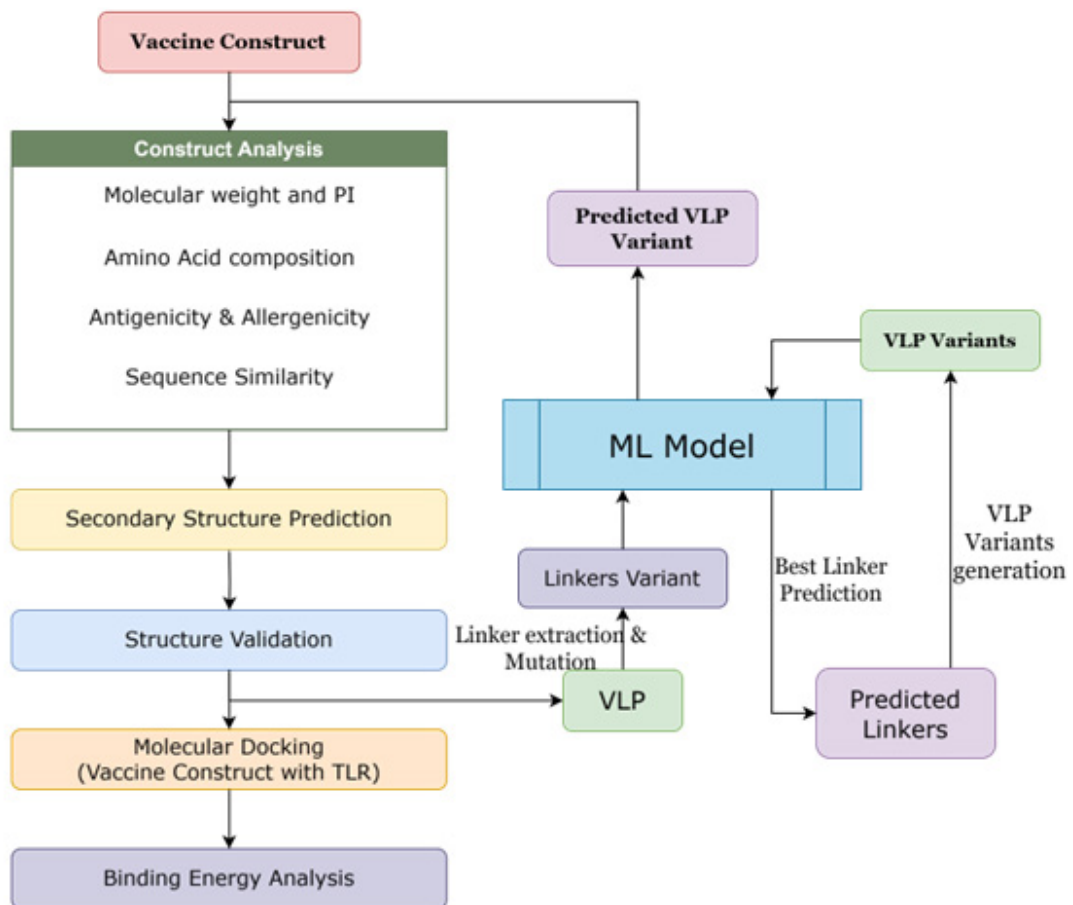


Fig. 2. Workflow for Vaccine Design: This flowchart outlines the step-by-step process in vaccine development, starting from the initial vaccine construct analysis, including assessments of molecular weight, pI, (Isoelectric point), antigenicity and allergenicity. It proceeds through secondary structure prediction and structure validation, which includes molecular docking with Toll-like Receptors (TLRs) and binding energy analysis. Parallel to this is the generation and prediction of VLP variants using a machine learning model, which feeds into the selection of the best linker for the VLP constructs.

retrieve their respective protein sequences as shown in supplementary Table S1.

Antigenicity and Allergenicity

IFN-gamma prediction accuracy was 81.39%, while IL4pred and IL10pred exhibited 75.76% and 81.24% accuracy, respectively. Only one peptide matched all selection criteria against the first protein. However, no peptide met the selection criteria for the other two proteins.

A comprehensive examination was carried out on a group of seven distinct proteins to understand their vaccine related properties. This analysis showed that these proteins possess a high probability of being recognized as antigens while concurrently exhibiting low chances for allergenic reactions. Notably, among these, only one protein showed the presence of transmembrane helical

activity, a feature that differentiated it from other proteins. The detailed attributes of these proteins are systematically catalogued in Table 1, that demonstrate their potential to be considered as potential protein to extract vaccine construct.

The objective of this investigation was to further narrowed down based on the antigenic properties of the proteins. Three proteins, UniProt IDs: P35253 (Envelope glycoprotein), P35258 (Transcriptional activator VP30), and P35256 (Membrane-associated protein VP24) were identified as prime candidates for further examination based on their antigenicity score (shown in Table 1). The selection criterion pivoted mainly on their marked antigenicity, which signifies their potential role in triggering an immune response. This aspect is critical, as the antigenic

Table 1. Selected structural proteins from Marburg virus and their vaccine related parameters including antigenicity, allergenicity and Number of Tm helices

Uniprot ID	Antigenicity	Allergenicity	Number of Tm helices
P27588	0.4761 Probable Antigen	Probable Non-Allergen	0
P35253	0.5481 Probable Antigen	Probable Non-Allergen	1
P35260	0.4107 Probable Antigen	Probable Non-Allergen	0
P35259	0.4360 Probable Antigen	Probable Non-Allergen	0
P31352	0.4518 Probable Antigen	Probable Non-Allergen	0
P35258	0.5636 Probable Antigen	Probable Non-Allergen	0
P35256	0.5423 Probable Antigen	Probable Non-Allergen	0

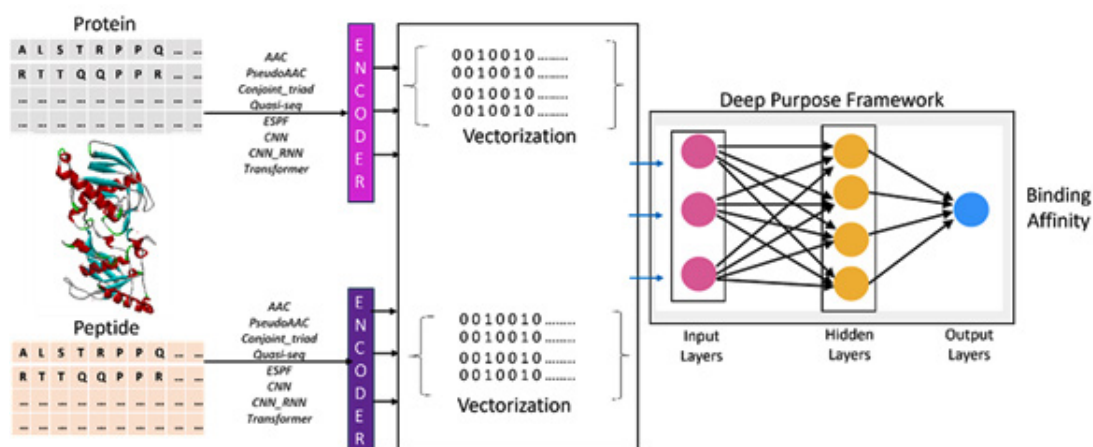


Fig. 3. Neural Network Encoding of Amino Acid Sequences using DeepPurpose framework: This diagram represents a machine learning model where sequences of amino acids are input into an encoder. The encoder's role is to convert the raw sequence data into a format suitable for the neural network to process. Subsequently, these transformed features are passed through various layers of the network (depicted as pink for input and yellow for hidden layers) to eventually arrive at an output (shown in blue). This model was used in this study to predict the affinity of the sequences.

Table 2. Selected CTL epitopes, LBL epitopes, and HTL epitopes from three proteins, P35253 (Envelope glycoprotein), P35258 (Transcriptional activator VP30), and P35256 (Membrane-associated protein VP24) with their relevant vaccine like properties

Protein	Peptide	Immunogenicity	Antigenicity	Allergenicity	Toxicity	Alleles	INF Induction	IL4 & IL10
Selected CTL epitopes								
P35253	FTEGNIAM	0.24146	0.6976	Non-Allergen	Non-Toxin	HLA-A*26:01	NA	NA
	FSLINRHAI	0.2229	0.8741	Non-Allergen	Non-Toxin	HLA-B*08:01	NA	NA
P35258	RSRTRNHQV	0.05979	0.6619	Non-Allergen	Non-Toxin	HLA-A*30:01	NA	NA
	NLGHILSYL	0.02093	0.5076	Non-Allergen	Non-Toxin	HLA-B*08:01	NA	NA
P35256	LEVTSAIHI	0.05941	1.1339	Non-Allergen	Non-Toxin	HLA-A*02:03	NA	NA
	TENSINLIDL	-0.04146	0.8691	Non-Allergen	Non-Toxin	HLA-B*40:01	NA	NA
						HLA-B*44:02		
						HLA-B*44:03		
						HLA-B*40:01	NA	NA
						HLA-B*44:03		
						HLA-B*44:02		
Selected HTL epitopes								
P35253	ASPNISSLTSYFPNI	-0.09305	1.0552	Non-Allergen	Non-Toxin	Positive	Inducing	
P35258	None of the peptide follows all the criteria							
P35256	LSEWLLLEVTSAIHI	0.40544	0.5996	Non-Allergen	Non-Toxin	Negative	Inducing	
Selected LBL epitopes								
P35253	LSWIPFFGPGIEGLYT	0.7257	0.7445	Non-Allergen	Non-Toxin	NA	NA	NA
	TEERTFSLINRHAIDF	0.43179	1.0356	Non-Allergen	Non-Toxin	NA	NA	NA
P35258	TMTELHMNHENLPQDQ	-0.02659	0.6147	Non-Allergen	Non-Toxin	NA	NA	NA
	SKPHYTNYHPRARMS	-0.03105	1.2354	Non-Allergen	Non-Toxin	NA	NA	NA
P35256	QQIITRVNMGFLVEV	0.48772	0.8902	Non-Allergen	Non-Toxin	NA	NA	NA
	MAELSTRYNLPANVTE	0.00647	0.6817	Non-Allergen	Non-Toxin	NA	NA	NA

nature of a protein determines its interaction with the host's immune system, and consequently, its role in various biological processes or therapeutic applications. Moreover, all the proteins were predicted as non-allergen, this provides the safety aspect of these protein to be used as vaccine. The significance of these findings highlights the potential applications of these protein in the immunology for vaccine design. The antigenic properties of these proteins could pave the way for novel approaches in vaccine development or in the design of targeted therapies. Understanding the allergenicity is equally crucial, as it ensures

the safety and efficacy of any derived vaccine candidate. The identification of transmembrane helices in one of the proteins opens up avenues for exploring its role in cellular processes, possibly providing insights into the mechanisms of cellular transport or signal transduction.

Prediction of CTL, HTL, LBL Epitopes

A significant focus was placed on the prediction of Cytotoxic T Lymphocyte (CTL), Helper T Lymphocyte (HTL) and Linear B-Lymphocyte (LBL) epitopes derived from the identified proteins of MARV. This phase of the study was instrumental in predicting a total of 314 potential CTL epitopes. These epitopes represent specific sequences within the complete protein that are capable of triggering an immune response, specifically the activation of CTLs, which play a crucial role in the body's defence mechanism against pathogens. Followed by the identification of these potential CTL epitopes, a rigorous evaluation process was initiated. This involved an in-depth assessment of various characteristics of the epitopes, including their immunogenic properties, antigenic features, and potential allergenic reactions. Additionally, the toxicity of these epitopes was also predicted. The criteria for selection of epitopes were based on the result of these assessments.

Cytotoxic T Lymphocyte (CTL)

The analysis yielded noteworthy results for the protein sequence P35253. Out of the

Table 3. Correlation of determinants (R²) for the training model of using the different encoders. These encoders of the DeepPurpose framework was used for encoding the sequences for using in the ML-model

Encoder	R-squared
AAC	0.415
PseudoAAC	0.825
Conjoint triad	0.966
Quasi-seq	0.885
ESPF	0.680
CNN	0.919
CNN_RNN	0.940
Transformer	0.925

Table 4. Final Wild-Type and Variant Virus-Like Particle (VLP) Construct Sequences. The black font of the sequence represents the adjuvant, yellow represents the HTL epitopes, blue represents the CTL and LBL epitopes and underline amino acids represents the linkers

Wild VLP Construct

MAKLSTDELLDAFKEMTLLELSDFVKKFEETFEVTAAPVAVAAAGAAPAGAAVEAAEEQSEFDVILE
 AAGDKKIGVIKVVREIVSGLGLKEAKDLVDGAPKPLEKVAKEAADEAKAKLEAAGATVTVKEAAAK
 FTEGNIAAMAAYFSLNRHAIAAYRSRTRNHQVAAAYNLGHILSYLAAAYLEVTSAIHIAAYTENSINLDLG
PGPGASPNISLTLSYFPNIGPGPGLSEWLLLEVTSAIHIKKLSWIPFFGPGIEGLYTKTEERTFSLINRHAI
 DFKKTMTELHMNHENLPQDQKKSKPHYNTNYHPRARSMSKKQQIIITRVNMGFLVEVKKMAELSTRYNLPANVTE

Variant VLP Construct

MAKLSTDELLDAFKEMTLLELSDFVKKFEETFEVTAAPVAVAAAGAAPAGAAVEAAEEQSEFDVILE
 AAGDKKIGVIKVVREIVSGLGLKEAKDLVDGAPKPLEKVAKEAADEAKAKLEAAGATVTVKEAAAK
 FTEGNIAAMKSIFSLINRHAIKSIRSRTNRNHQVKSINLGHILSYLKSILEVTSAIHIKSITENSINLDLGPGPG
ASPNISLTLSYFPNIGPGPGLSEWLLLEVTSAIHIKKLSWIPFFGPGIEGLYTKTEERTFSLINRHAIDFKK
 TMTTELHMNHENLPQDQKKSKPHYNTNYHPRARSMSKKQQIIITRVNMGFLVEVKKMAELSTRYNLPANVTE

assessed epitopes, 10 epitope peptides were found to adhere strictly to the established selection criteria for CTL. These peptides showed promise in context to their immunogenic potential, indicating their capability to effectively stimulate an immune response. Out of this subset, the top two peptides were chosen based on their superior immunogenicity. This selection underscores the potential of these peptides in eliciting a robust

immune response, which is a key consideration in vaccine development and immunotherapy. In the case of protein sequence P35258, a similar pattern was observed. Seven peptides from this sequence followed the stringent selection criteria. As with P35253, the top two peptides were selected in this case, primarily based on their heightened immunogenic properties. This choice reflects a focused approach towards identifying epitopes with

Table 5. Comparative biophysical property analysis of wild-type and variant VLP Constructs. The biophysical properties, include amino acid count, molecular weight, theoretical isoelectric point (pI), charge distribution, instability index, and average hydrophaticity (GRAVY).

Characteristics	Wild VLP	Variant VLP
Number of amino acids	352	352
Molecular weight	38307.04 Da	38422.43 Da
Theoretical pI	6.58	8.73
Total number of negatively charged residues (Asp + Glu)	42	42
Total number of positively charged residues (Arg + Lys):	40	45
The instability index (II)	38.01	38.12
Grand average of hydrophaticity (GRAVY)	-0.101	-0.137

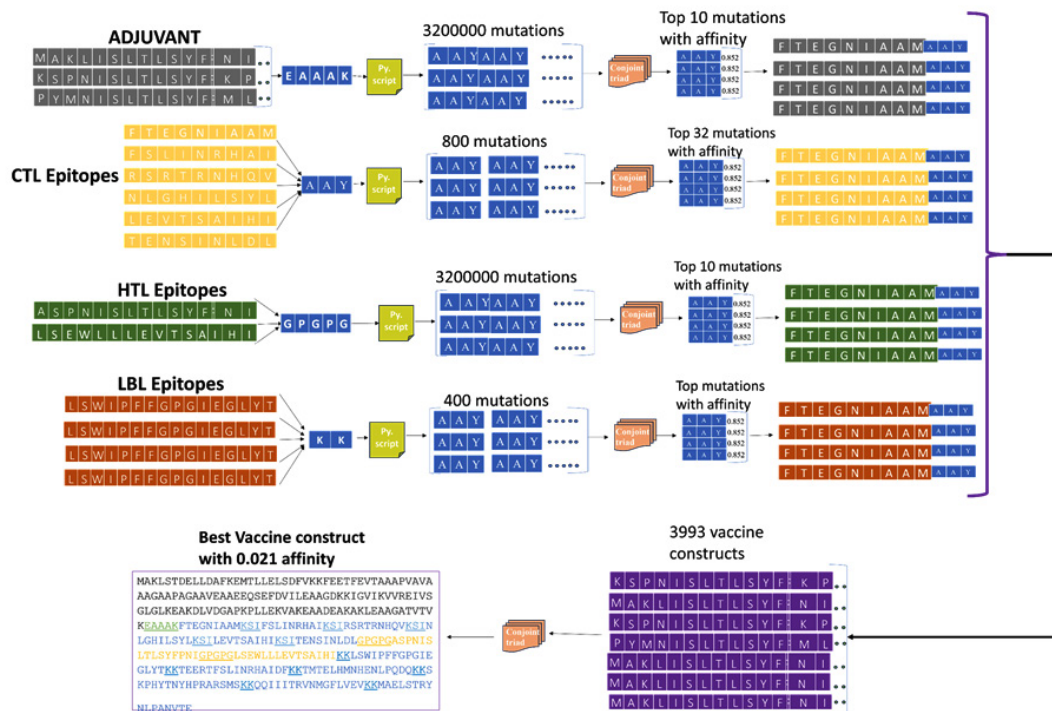


Fig. 4. Mutation of VLPs using ML model: This illustration represents the process of mutation of the linkers used by the adjuvants, CTL, HTL and LBL epitopes using the Python script. The ‘Conjoint triad’ of the DeepPurpose framework was further used for screening sequences based on affinity. The VLP variants were generated and using the conjoint triad screened for the best affinity.

optimal immune response triggering capabilities. Similarly, for the protein sequence P35256, the analysis showed four peptides that satisfied the selection criteria. Following the established protocol, the top two peptides were selected, with their selection being primarily driven by their immunogenicity. This selection process was critical in narrowing down the candidates to those with the highest potential for practical application in immunological study. Table 2 shows the selected epitope peptide from the three proteins with their vaccine related properties.

Helper T Lymphocyte (HTL)

Once the potential HTL epitopes were identified, a comprehensive assessment was conducted. This evaluation was multifaceted, examining the immunogenicity, antigenicity, allergenic potential, and toxicity of the epitopes. The criteria for this assessment were meticulously defined in Table 2, providing a structured properties for evaluating the epitopes. The assessment showed that for P35253 protein, it was found that only one epitope peptide satisfied all the predefined criteria of vaccine candidate. This peptide stood out due to its ability to follow the stringent requirements,

marking it as a promising candidate for vaccine construct. Remarkably, none of the epitope peptides from the P35258 protein sequence met the established criteria, indicating a divergence in their potential as effective HTL epitopes. This result underscores the complex nature of protein-epitope interactions and the variability in immunogenic potential across different protein sequences. Regarding the protein sequence of P35256, the findings were mixed. While one peptide corresponding to this protein satisfied most of the specified requirements, it fell short in one crucial aspect i.e. its ability to induce Interferon (IFN). This negative performance in IFN induction is significant, as IFN plays a vital role in the immune response, particularly in antiviral defence mechanisms.

Linear B-Lymphocyte (LBL)

Subsequently, the peptides underwent a LBL screening process for the specified proteins, focusing on criteria of immunogenicity, antigenicity, allergenicity, and toxicity. The results of this screening are detailed in Table 2, which was utilized for the development of the vaccine. As a result, in the case of protein P35253, 13 peptides are

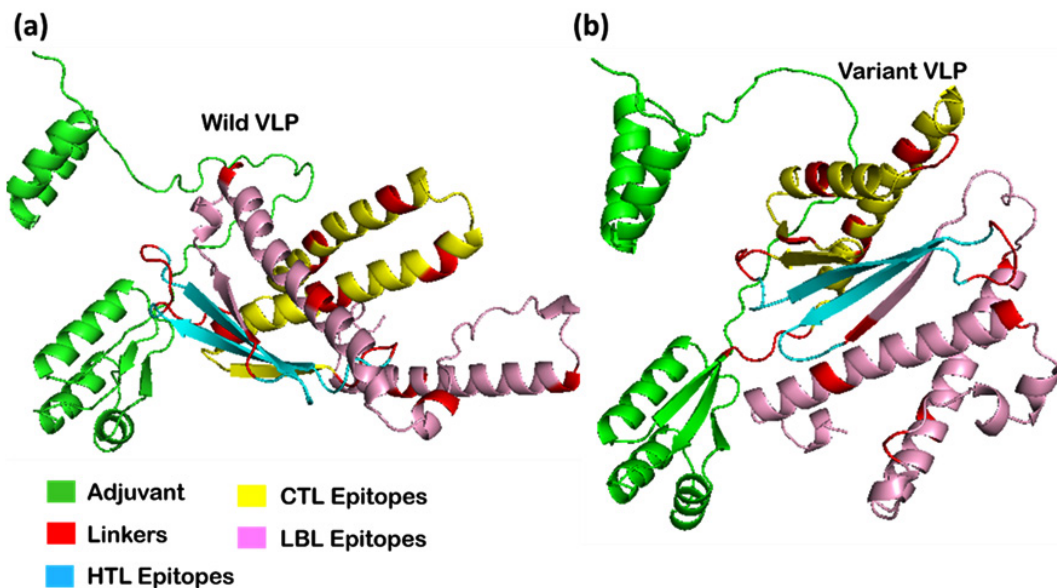


Fig. 5. 3D structural comparison of vaccine construct predicted by AlphaFold for (a) Wild VLP and (b) Variant VLP. This illustration presents the three-dimensional structures of wild-type and variant VLPs, highlighting differences in secondary structures such as alpha-helices and beta-sheets with the adjuvants, CTL, HTL, LBL epitopes and linkers highlighted in the structures.

identified that adhere to the established selection criteria, and subsequently, the top 2 peptides are chosen based on their immunogenicity profile. Similarly, for protein P35258, 4 peptides meet the selection criteria, and the top 2 peptides are selected based on their immunogenicity characteristics. Likewise, in the context of protein P35256, 3 peptides are found to satisfy the selection criteria, and the top 2 peptides are selected on the basis of their immunogenicity attributes.

Human Homology Comparison and MHC Cluster Analysis

Identified epitopes (shown in Table 2) were sequentially compared against the complete human proteome to find the similar segments of sequence. The CTL, HTL, and LBL epitopes that had been screened had exhibited no homology with the human proteome, thereby validating their role as antigens or exogenous entities for humans. Using the IEDB platform, the MHC-I alleles

that had interactions with the epitope derived from the chosen structural protein were grouped together. Here, 25 alleles from each group were incorporated.

Designing of Vaccine construct and Evaluation

The construction and assessment of the vaccine involved a strategic assembly for combination of epitopes targeting CTL, HTL, and LBL. These were linked using linkers, “EAAAK” sequence was used to connect the adjuvant with the CTL epitopes. Adjuvants were added to the epitope sequence to improve the immune response. It also assists the vaccine candidate to create long lasting immune response. Moreover, two CTL epitopes were connected using “AAY” sequence linker. Similarly, “GPGPG” was used to connect two HTL epitopes, and “KK” for LBL epitope connectivity. The final vaccine construct is shown in the Table 4. The adjuvant used in the vaccine was identified as the TLR agonist, specifically the 50S ribosomal

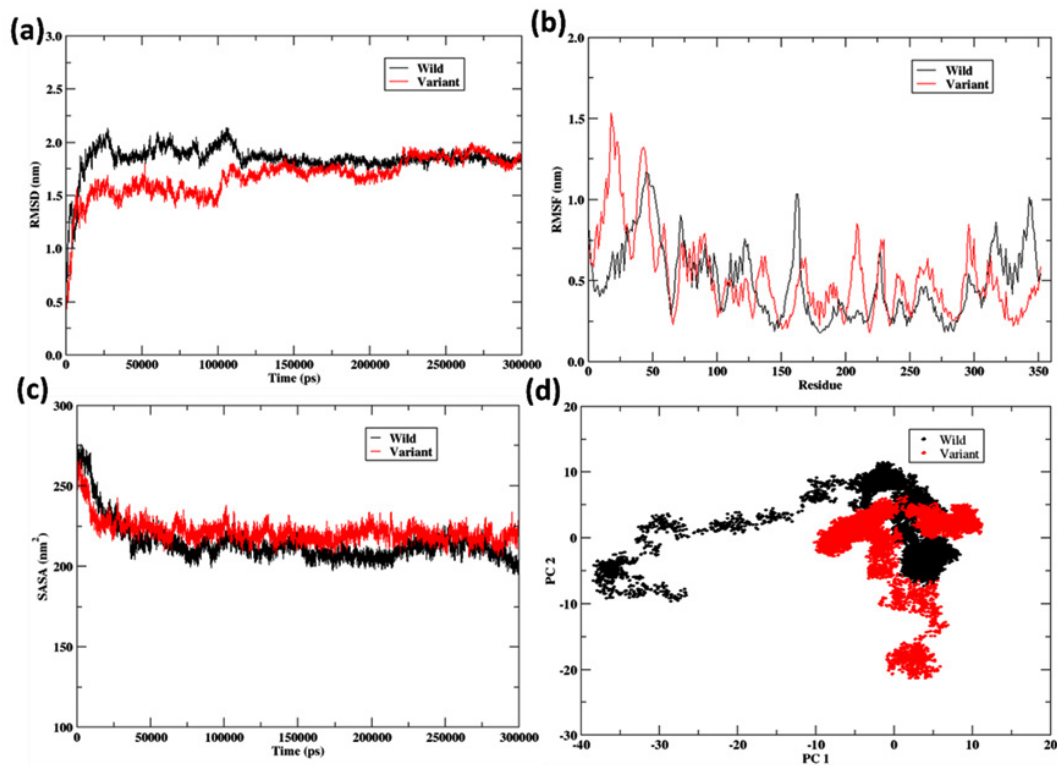


Fig. 6. Comparative post molecular dynamics analysis of wild-type and variant VLP: This set of graphs presents a detailed 300 ns molecular dynamics simulation, comparing wild-type and variant protein forms. The analysis includes root-mean-square deviation (RMSD) over time (a), root-mean-square fluctuation (RMSF) per residue (b), the solvent accessible surface area (SASA) over time (c), and a principal component analysis (PCA) scatter plot (d) to show the conformational space explored by both proteins.

protein L7/L12, catalogued under the NCBI ID P9WHE3.

Machine Learning Based Mutation

The construction of a variant sequence was initiated by extracting linker sequences, followed by their subsequent mutation through a Python script. The linkers investigated in this study were EAAAK, AAY, GPGPG, and KK, resulting in mutations of 3,200,000 for EAAAK and GPGPG,

8,000 for AAY, and 400 for KK, respectively. A machine learning model was developed using the DeepPurpose framework to systematically assess and determine the most effective vaccine construct from the extensive collection of produced mutations. Multiple encoders were used for model training and r-squared for each case was noted as shown in Table 3. Based on R², Conjoint triad was used as encoder for affinity prediction.

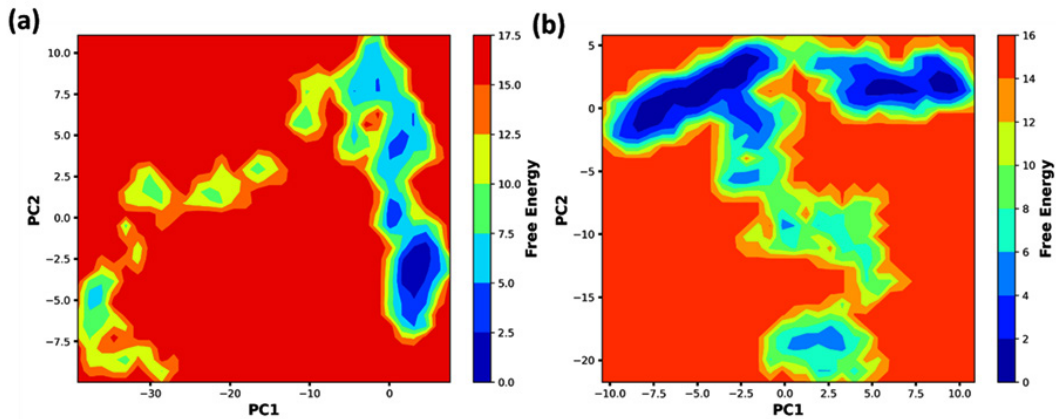


Fig. 7. Free energy landscape for the (a) Wild VLP (b) Variant VLP: These contour maps represent the free energy landscapes of a molecular system as a function of the first two principal components (PC1 and PC2). The color gradients indicate regions of varying free energy, with red areas representing higher energy states and blue areas indicating lower energy states, which are often associated with more stable conformations in the context of molecular dynamics simulations.

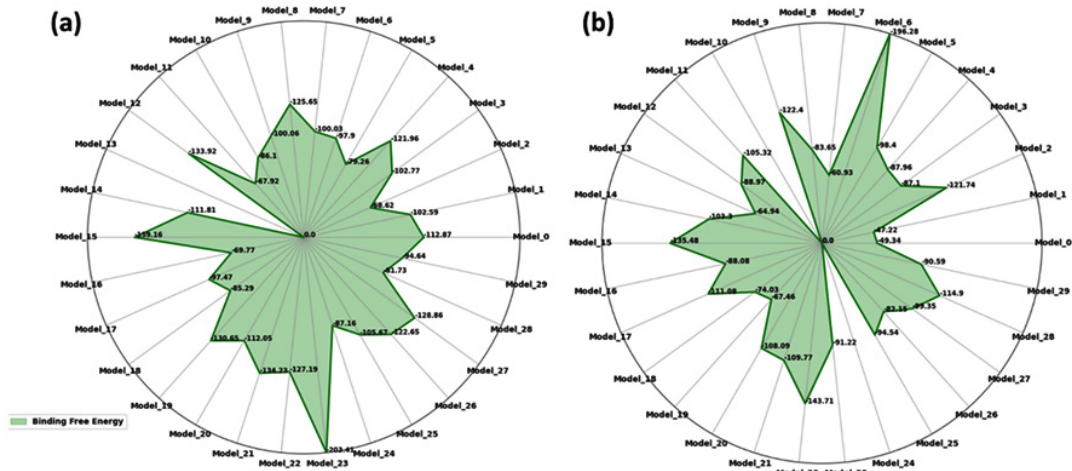


Fig. 8. Comparative radar chart of binding free energy for wild and variant VLP: This dual radar chart illustrates the binding free energy values across a series of molecular models (a) and (b). Each spoke on the radar represents a different model, with numerical values indicating the magnitude of binding free energy, a critical factor in assessing the stability and affinity of molecular interactions.

Sequences encoded using the Conjoint triad encoder, a feature of the DeepPurpose toolkit, were used for model training. Before model training, affinity values were normalised by transforming them into their negative logarithmic equivalents and then normalising these values according to the length of their respective sequences. This technique enabled the model's training phase, during which the encoded sequences and their normalised affinity scores were used as input.

Upon training completion, the model was tasked with evaluating the generated sequences to pinpoint the most promising linker variant. The top 10 variants for EAAAK were, IITTE, MMNMQ, FFHFH, GVIYK, RQLQQ, AAQRA, WVNNN, DMDDE, SCSCC and EPPLP, similarly, for GPGPG the top 10 variants were, SFSSM, TRTYT, DLWDD, MRAAA, GGNEG, HHHII, MEREE, CCQM, VVKKV and YPPLP. The predicted

affinity ranged from 0.77 to 1.79 for EAAAK while it ranged from 0.66 to 1.98 for GPGPG. The AAY linker had top 32 variants showed same predicted score of 0.852 and thus, these all were selected. Unlike, the three mentioned linkers, KK being only two residues long showed same predicted score for all the 400 variants and thus, the wild form itself was used for further study. The selection of notable mutants for each linker was conducted, and these variants were combined into diverse variant VLP configurations. These configurations were further enhanced by integrating wild-type linkers to guarantee their inclusion in the study with entire sequences, hence eliminating any potential bias. A total of 3,993 variant VLPs were produced for the purpose of predicting affinity. The trained model was subsequently provided with these VLP variations to predict the optimal sequence, which was then utilised for further research. The

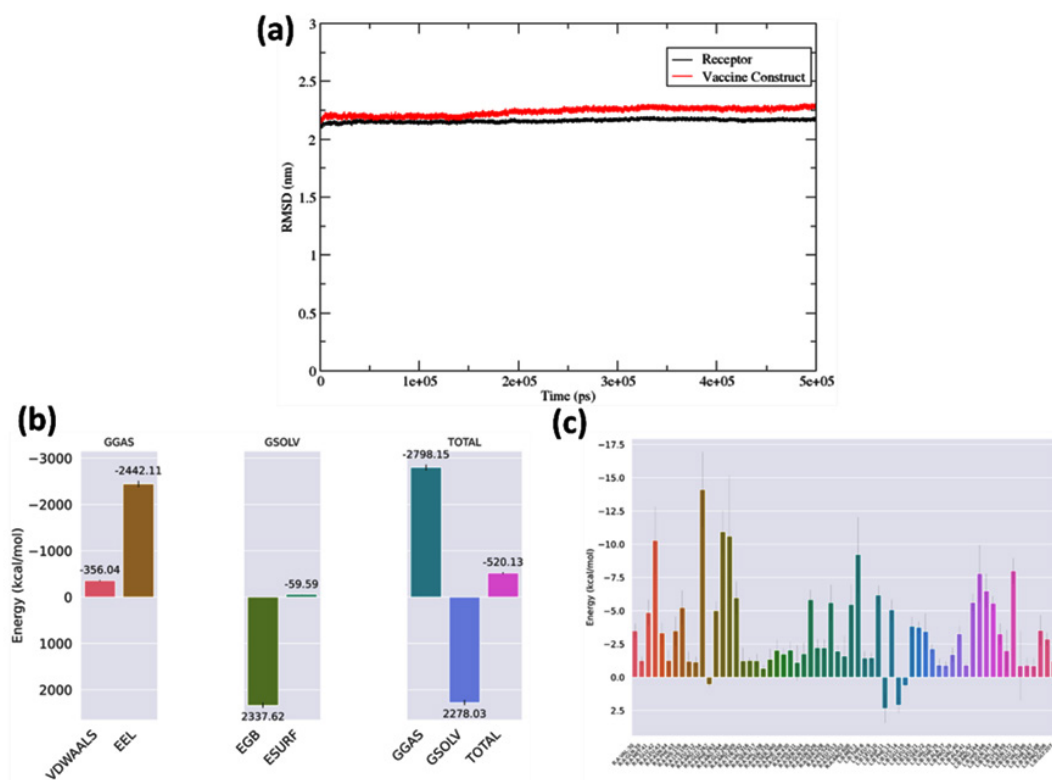


Fig. 9. Post MD analysis of docked variant VLP-TLR complex over the 500 ns MD simulation: (a) RMSD curve for variant VLP-TLR complex calculated during the 500 ns MD simulation. (b) Bar plot representing binding free energy (\AA G) of the variant VLP-TLR complex. (c) Energy Contribution for residues of variant VLP-TLR complex.

optimal vaccine design had an affinity of 0.021, as anticipated by the machine learning model. The best predicted variant VLP sequence is shown in Table 4. Illustration in the Figure 4 explained the screening of the best variant VLP using ML-based model. This approach underscores the integration of computational biology with machine learning to streamline the identification and optimization of VLPs, potentially expediting the development of effective immunogens.

VLP Properties Evaluation

The properties of the wild and variant VLP were analysed and listed in the Table 5. This wild VLP and variant VLP is composed of 352 amino acids, culminating in a molecular mass of 38,307.04 Da and 38422.43 Da, respectively. It has an isoelectric point (pI) of 6.58 and 8.73 for wild and variant VLP. Its molecular architecture comprises 42 negatively charged residues including Aspartic acid (Asp) and Glutamic acid (Glu), and 40 positively charged residues such as Arginine (Arg) and Lysine (Lys). The wild VLP and variant VLP has an instability index is measured at 38.01 and 38.12, respectively. Additionally, the Grand Average of Hydropathicity (GRAVY) index, noted as -0.10 and -0.137, as listed in Table 5, indicates the hydrophilic nature of the protein, suggesting its structural stability.

The antigenic behaviour and potential allergenic reactions were predicted for both wild and variant VLP. Here, findings indicated a notable antigenicity score of 0.740, which strongly suggests the vaccine's capability to trigger an immune response. While variant VLP showed an antigenicity score of 0.6730 which was comparable. Both wild and variant VLP showed probability of being non-allergen. This comprehensive analysis is instrumental in advancing the understanding of the VLP characteristics and ensuring its efficacy and safety. The antigenicity score is particularly encouraging, as it points towards the VLP's potential effectiveness in inducing an immune response. Similarly, the non-allergenic nature reassures its applicability across diverse groups without significant risk of allergic reactions. Collectively, these findings provide a promising outlook for the vaccine's future application and set a foundation for subsequent clinical evaluations.

Three-Dimensional Structure Modelling and Validation

AlphaFold (Colab version) was used to generate the vaccine construct/VLPs 3D structure as shown in Figure 5. As shown in the Figure 5, the two structures were composed of the adjuvant, CTL, HTL and LBL epitopes along with the linkers. These structures were selected for structure validation using Procheck application. In this context, the AlphaFold structures of wild and variant VLP exhibited 96.2% and 88.5%, respectively of amino acid residues in the primary preferred region of the Ramachandran plot. Hence, the both these structures predicted by AlphaFold showed favourable confidence scores from Procheck. These structures were further selected for molecular dynamics simulation.

Molecular Dynamics Simulation

The wild and variant VLP structures were used for molecular dynamics simulation for studying the stability and flexibility. Root Mean Square Deviation (RMSD) is frequently utilised to evaluate the stability of protein structures during molecular dynamics simulations. The measurement calculates the mean distance between the atoms, often the backbone atoms, of aligned proteins. Lower RMSD values generally indicate a more stable protein structure.

RMSD and RMSF

The RMSD graph depicted in Figure 6(a) shows the protein structure's stability throughout a 300 ns timeframe. RMSD of the wild fluctuated between the range 1.75 nm to 2 nm while variant was in the ranged 1.5 nm to 1.75 nm for most of the simulation. Both the wild type (in black) and the variant (in red) exhibit an early increase in RMSD during the simulation, indicating initial structural alterations. The variant reaches a lower RMSD value compared to the wild type, indicating it may attain a more stable conformation faster or possess a less flexible structure overall. The root mean square fluctuation (RMSF) graph illustrates the variability of each residue throughout the simulation. Figure 6(b) displayed the RMSF of wild-type and variant VLP. The variant shows reduced fluctuations compared to the wild type, indicating that the mutations may have caused a more stable structure, perhaps impacting epitope presentation dynamics.

SASA and PCA

The solvent-accessible surface area (SASA) graph quantifies the protein's surface area that can be reached by solvent molecules. In the Figure 6(c), both the wild type and variant show similar trends, with the variant exhibiting slightly higher SASA values throughout the simulation, which could impact how the immune system recognizes and interacts with the VLP. Both showed a SASA of 225 nm² for most of the simulation. The principal component analysis (PCA) scatter plot (Figure 6(d)) shows the dominant patterns of movement in the protein structures. The distinct clustering of the wild type and variant indicates significant differences in their dynamic behaviors, which might have implications for their functional activities. The post MD simulation indicate that the variant VLP possesses distinct structural and dynamic characteristics in comparison to the wild type. These differences may play a critical role in its efficacy as a vaccine candidate, potentially influencing its stability, immune system recognition, and interactions with molecules like toll-like receptors.

FEL

Figure 7 shows the free energy landscapes (FEL) for a wild type VLP in 7(a) and a variation VLP in 7(b), represented on two main components (PC1 and PC2). These landscapes are commonly employed to represent the thermodynamic stability and conformational states of macromolecules, including proteins, in molecular dynamics simulations. The FEL is color-coded based on free energy values, with blue representing lower free energy, suggesting more stability and likelihood of the system existing in that state, and red representing higher free energy, indicating poorer stability. The landscape's topography indicates the quantity of stable states (basins) and the elevation of the barriers separating them.

Figure 7(a) shows that the wild type VLP's FEL has fewer deep blue patches, indicating a reduced amount of very stable conformations. Figure 7(b) displays a larger region of deep blue, suggesting that the variation VLP exhibits a wider array of stable conformations. The variant's FEL shows deeper and more well-defined basins, indicating a rougher and thermodynamically favourable energy landscape that may lead to a larger chance of stable conformations.

The RMSD graph shows that the variant VLP exhibits increased stability, as evidenced by its consistently lower and more stable RMSD values over time compared to the wild type. The red line showing the change VLP rapidly stabilises and retains its initial conformation with minor deviation during the simulation. The variant VLP appears to be structurally more stable than the wild type. The free energy landscapes support the stability of the variety VLP, complementing the data. The variant landscape has a wider range of low-energy conformations, indicated by deeper blue areas, in contrast to the wild type. This suggests a more advantageous and steady structural state, which is crucial for the effectiveness of a virus-Like Particle (VLP) in vaccine development. The computational investigations suggest that the variation VLP reaches structural stability quickly and maintains a stable conformation with favourable energy, highlighting its potential as a promising candidate for vaccine development.

Molecular Docking

The wild and variant VLP structures after MD simulation were analysed for their interaction with TLR4 using molecular docking. The Cluspro programme produced 30 unique clusters with elevated interaction energy as shown in supplementary Table S2. The initial cluster displayed binding scores of -1072.7 kcal/mol for the variant VLP and -934.8 kcal/mol for the wild VLP. This suggests the strong binding of the variant VLP to the TLR4 receptor compared to the wild VLP. This receptor binds to the foreign pathogen, indicating its interaction with the vaccine design, which resembles a virus-like particle. This binding would additionally stimulate the immunological response.

Figure 8 showed a radar chart comparing the binding free energy of 30 docked models for a wild type and variant virus-like particle (VLP). Figures 8(a) and 8(b) indicate that the length of each spoke represents the binding free energy of each model, where shorter lengths imply stronger binding affinity due to lower energy values. The green region in Figure 8(a) representing the binding free energies of various models is smaller than in Figure 8(b), suggesting higher free energy values and therefore weaker binding affinities. The larger green region in Figure 8(b) suggests lower binding free energy values for most models,

indicating stronger binding affinities. The variant VLP showed decreased binding free energy values in most models, suggesting a possible stronger binding affinity to its target. These results suggest that variant VLP is a more suitable choice for advancement, given the lower binding free energies.

MD simulation of Variant VLP- TLR4 Complex RMSD

This section involved conducting a molecular dynamics simulation to gain an understanding of the stability of the TLR4-variant VLP complex and to assess the effectiveness of the developed vaccine. Root-Mean-Square Deviation (RMSD) of both the variant VLP and TLR4 receptor was calculated over the 500 ns trajectory for analysing the deviation of the docked complex. Here the RMSD was calculated by aligning the other protein. Receptor is aligned and RMSD for VLP/vaccine construct calculated and vice versa. Figure 9(a) displays the RMSD plot of the variant VLP-TLR4 complex. Though, it exhibited a greater RMSD, ranging from 2.14 nm to 2.21 nm, it maintains a stable appearance for most part of the simulation frame. The RMSD plot of the vaccine-TLR4 complex showed consistent stability over the simulation, indicating that variations did not cause significant conformational changes and the complex remained stable throughout. The mean RMSD for the receptor and variant VLP was 2.14 ± 0.01 and 2.21 ± 0.03 nm, respectively. The stable connection between the VLP and the receptor indicates the potential effectiveness of the vaccine design in consistently triggering the targeted immune response.

Binding Free Energy

The binding free energy of vaccine construct-TLR complex was assessed using the MM/GBSA methodology. This analysis was conducted based on the trajectory data obtained from the final 20 nanoseconds of the simulation. The complex exhibited a cumulative binding free energy (ΔG) of -520.13 (kcal/mol), as represented in Figure 9(b). The ΔG calculation of the was contributed with the ΔG_{GAS} with -2798.15 kcal/mol while ΔG_{SOLV} with 2278.03 kcal/mol. The negative cumulative ΔG value suggests that the formation of the complex is energetically favourable when both the gas-phase interactions and solvation effects are considered. It was evident

that the complex exhibited high magnitude of minimum binding free energy (ΔG), indicating a robust and stable binding interaction between vaccine construct and toll-like receptor. This observation suggests a favourable and enduring interplay between these two macromolecules.

Energy Contribution for Residues

Further, the energy contribution of the residues of both variant VLP and TLR complex was calculated as shown in Figure 9(c). Critical residues are highlighted in this analysis which contribute to the ability to form stable protein-protein interactions. It was observed that the TLR receptor showed favourable energy contribution for residues ARG234 with -14.11 kcal/mol, ARG264 with -10.94 kcal/mol, GLU266 with -10.6 kcal/mol, and GLU42 with -10.3 kcal/mol. Further, the residues that showed favourable binding free energy for the variant VLP were MET1 with -9.23 kcal/mol, ASN172 with -8 kcal/mol, ARG164 with -7.8 kcal/mol, HIS166 with -6.5 kcal/mol and LEU10 with -6.18 kcal/mol. These findings highlight the complex interplay of multiple types of residue interactions that underpin the stability and specificity of the VLP-TLR complex, which may be pivotal for the VLP's immunogenic efficacy.

DISCUSSION

The progress in computational techniques has resulted in a substantial reduction in both the time and capital expense associated with complicated tasks of vaccine design. The conventional method of vaccine design is time-consuming that require a significant amount of effort even for a single epitope vaccine design. Conversely, *in silico* methods have shown high efficiency and can be effectively used in the design of multi-epitope vaccines. Moreover, the application of well-constructed databases containing comprehensive data greatly facilitates the availability of required information, such as identifying epitopes for a specific protein, as demonstrated in the current investigation. AI-based tools including AlphaFold have shown their efficacy and reliability in the realm of protein structure prediction, illustrating the effectiveness and dependability of computational methodologies in the development and advancement of therapeutic treatments. Moreover, previous studies have shown implementation of *in silico* techniques to yield a

reliable therapeutic outcome. *Sanami*. (2021) have used similar approach to design a multiepitope vaccine against cervical cancer⁴⁵. Furthermore, *A S Mustafa*, in another study demonstrated the development of multiepitope vaccine against *Mycobacterium tuberculosis* showed the accuracy of *in silico* techniques in vaccine design through experimental validation. The experimental results favored the *in silico* approach and thus, adoption of these techniques considered safe and can yield satisfactory results⁴⁶.

In the present study vaccine construct for Marburg virus is constructed using its three of the most antigenic proteins, (1) envelope glycoprotein, (2) transcriptional activator VP30, (3) membrane-associated protein VP24. Whereas in the study by *Yosuaif*, (2023) they used only glycoprotein to form a multiepitope screened vaccine construct against the Marburg virus⁸. Moreover, the final construct of the multiepitope vaccine, which includes adjuvant (TLR agonist) and linkers, was produced through a series of prediction and screening stages. The examination of the final vaccine design showed promising antigenicity, along with the establishment of a strong complex formation with Toll-like receptors (TLR). Similarly, in the study by *Hasan* (2019), TLR-8 used as potential target⁴⁷. Additionally, in the study by *Soltan* (2022) TLR-4 also used as potential vaccine binder⁴⁸. Toll-like receptors (TLRs) are proteins essential in the human innate immune system, identifying and reacting to pathogen-associated molecular patterns (PAMPs) to defend against various illnesses, including viral infection, by activating proinflammatory cytokines and type I interferons. Therefore, as an adjuvants TLR agonist used for influenza, malaria, tuberculosis and cancer vaccines. In this study, the epitopes extracted was validated thorough focusing on their allergenicity, antigenicity, and solubility. The results indicated that while the epitopes were non-allergenic, they exhibited significant antigenicity, which led to their selection for more detailed analysis. Alligning this finding, the study conducted by *Sami*, as well as the investigation by *Mustafa and Shantier*, also showed that the epitopes displayed no allergenic properties^{9,49}. The current study leverages a machine learning model within the DeepPurpose framework, where *in silico* methods were first used to screen potential epitopes based on their binding

affinity to immune receptors. This study advances further by employing an ML model trained on a substantial dataset from Skempi v2.0, enabling the analysis of a more extensive array of mutants at a fraction of the time. Conjoint triad model was used for selecting the best variant VLP based on the affinity. Furthermore, the incorporation of molecular docking and MD simulation techniques established the interaction of the vaccine construct with Toll-like receptors (TLRs). This stability points to the potential efficacy of the vaccine in triggering immune response against MARV.

Building on this framework, this study is integrating multiple antigenic proteins in the vaccine construct that could potentially offer a broader immune response than a vaccine targeting a single protein. This multi-target strategy might confer enhanced protection against diverse strains of the Marburg virus, addressing the challenge of viral genetic variability. The current study's integration of machine learning to identify the best mutant VLP sequence aligns with these innovative approaches, showcasing the potential for rapid and efficient vaccine development through computational methods. Furthermore, the incorporation of TLR agonists as adjuvants in the vaccine design aligns with emerging trends in vaccine development, emphasizing the role of innate immunity in enhancing adaptive immune responses. The non-allergenic nature of our chosen epitopes, coupled with their demonstrated antigenicity, underscores the potential for a strong and safe immune response. This comprehensive approach sets a precedent for future vaccine designs, particularly in tackling pathogens with high mutation rates and complex pathogenicity.

The sole dependence on *in silico* methodologies, which are valuable for initial screening and predictions, is one of the main limitations of this study. However, they are unable to fully replicate the complexities of biological systems. Despite their robustness, the computational models do not encompass the complete spectrum of immune responses that may manifest *in vivo*, including potential unforeseen interactions with other immune cells or variations in human genetic backgrounds. Additionally, the vaccine candidate's immunogenicity, safety, and toxicity must be evaluated in a controlled laboratory environment by conducting *in vitro*

studies to validate its predicted efficacy, stability, and binding affinities. Furthermore, it is imperative to conduct *in vivo* trials in appropriate animal models to assess the vaccine's capacity to induce protective immunity, ascertain the optimal dosage, and identify any adverse effects. The vaccine's real-world applicability is still uncertain in the absence of these critical experimental validations. Consequently, although this investigation establishes a critical foundation, additional experimental research is necessary to verify the vaccine's clinical application potential.

CONCLUSION

In this study, a novel peptide vaccine against the MARV was designed, utilizing immunoinformatics techniques, machine learning that solidified by molecular docking and simulation trials. This innovative vaccine design incorporates a blend of linear B-cell epitopes with epitopes targeting cytotoxic and helper T lymphocytes. Following this, an adjuvant and various linkers were incorporated to create a cohesive epitope structure. This led to the construct of a subunit vaccine construct/ virus-like particle (VLP), combining B and T cell epitopes connected by appropriate linkers. Advanced computational and machine learning methods have enabled the development and screening of multiple vaccine construct/VLP variations that optimise linkers in a VLP sequence. The affinity of the wild and variant vaccine construct/VLP for TLR-4 was confirmed through molecular docking and MM/GBSA that established stronger binding with TLR4 to generate the immune response. The variant VLP emerges as a more promising candidate for subsequent stages of vaccine development, highlighting its potential for eliciting a robust immunological response against its target. The molecular dynamics simulation and MM/GBSA binding free energy analysis shows that the vaccine-TLR4 complex is structurally stable, and has strong and favourable interaction over an extensive simulation period. The variant VLP exhibited favourable properties that suggest its potential suitability for experimental validation. The study highlights the variant VLP's promise as a viable candidate for further *in vitro* and *in vivo* testing. Upon experimental validation, this vaccine candidate may provide novel insights into MARV

neutralisation procedures and aid in the formulation of successful treatments.

ACKNOWLEDGEMENT

The authors are thankful to the staffs of Growdea Technologies Pvt Ltd for continuous support in reviewing the article.

Funding Sources

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Conflict of Interest

The authors do not have any conflict of interest.

Data Availability Statement

This statement does not apply to this article.

Ethics Statement

This research did not involve human participants, animal subjects, or any material that requires ethical approval.

Informed Consent Statement

This study did not involve human participants, and therefore, informed consent was not required.

Author Contributions

Conceptualization, methodology, and editing were performed by Avinash Mishra. Shreyansh Suyash carried out the majority of the calculations and wrote the manuscript. Wajihul Hasan, Vinod Jangid and Priyasha Maitra performed the formal analysis and validation, and Praveen Punia did the validation of the data. All authors participated in the review and editing of the manuscript.

REFERENCES

1. Atkins C, Miao J, Kalveram B, Juelich T, Smith JK, Perez D, Zhang L, Westover JLB, Wettere AJV, Gowen BB, Wang Z, Freiberg AN. Natural History and Pathogenesis of Wild-Type Marburg Virus Infection in STAT2 Knockout Hamsters. *The Journal of Infectious Diseases*, 2018; 218: S438.
2. Markin VA. Marburg virus and the disease it causes. *Journal of microbiology, epidemiology and immunobiology*, 2022; 99: 605–618.
3. Cross RW, Longini IM, Becker S, Bok K, Boucher D, Carroll MW, Diaz JV, Dowling WE, Draghia-Akli R, Duworko JT, Dye JM, Egan MA, Fast P,

- Finan A, Finch C, Fleming TR, Fusco J, Geisbert TW, Griffiths A, Günther S, Hensley LE, Honko A, Hunegnaw R, Jakubik J, Ledgerwood J, Luhn K, Matassov D, Meshulam J, Nelson EV, Parks CL, Rustomjee R, Safronetz D, Schwartz LM, Smith D, Smock P, Sow Y, Spiropoulou CF, Sullivan NJ, Warfield KL, Wolfe D, Woolsey C, Zahn R, Henao-Restrepo AM, Muñoz-Fontela C, Marzi A. An introduction to the Marburg virus vaccine consortium, MARVAC. *PLoS Pathog*, 2022; 18: e1010805.
4. CDC report. Marburg Virus Disease Outbreaks | Marburg (Marburg Virus Disease) | CDC, <https://www.cdc.gov/vhf/marburg/outbreaks/chronology.html> (2023, accessed 27 September 2023).
 5. Aborode AT, Wireko AA, Bel Nono KN, Quarshie LS, Allison M, Bello MA. Marburg virus amidst COVID 19 pandemic in Guinea: Fighting within the looming cases. *Int J Health Plann Manage*, 2022; 37: 553–555.
 6. Idris I, Adesola RO, D'Souza JN. Marburg virus outbreaks in Africa. *Bull Natl Res Cent*, 2023; 47: 96.
 7. Hariprasath R, Akashpriya C, Lakshmaiah VV, Praveen N. In silico studies of viral protein inhibitors of Marburg virus using phytochemicals from *Andrographis paniculata*. *J App Biol Biotech*, 2022; 11: 222–231.
 8. Yousaf H, Naz A, Zaman N, Hassan M, Obaid A, Awan FM, Azam SS. Immunoinformatic and reverse vaccinology-based designing of potent multi-epitope vaccine against Marburgvirus targeting the glycoprotein. *Heliyon*, 2023; 9: e18059.
 9. Sami SA, Marma KKS, Mahmud S, Khan MdAN, Albogami S, El-Shehawi AM, Rakib A, Chakraborty A, Mohiuddin M, Dhama K, Uddin MMN, Hossain MK, Tallei TE, Emran TB. Designing of a Multi-epitope Vaccine against the Structural Proteins of Marburg Virus Exploiting the Immunoinformatics Approach. *ACS Omega*, 2021; 6: 32043–32071.
 10. DiCarlo A, Möller P, Lander A, Kolesnikova L, Becker S. Nucleocapsid formation and RNA synthesis of Marburg virus is dependent on two coiled coil motifs in the nucleoprotein. *Virol J*, 2007; 4: 105.
 11. Pervin T, Oany AR. Vaccinomics approach for scheming potential epitope-based peptide vaccine by targeting l-protein of Marburg virus. *In Silico Pharmacol*, 2021; 9: 21.
 12. Olson RD, Assaf R, Brettin T, Conrad N, Cucinell C, Davis JJ, Dempsey DM, Dickerman A, Dietrich EM, Kenyon RW, Kuscuoglu M, Lefkowitz EJ, Lu J, Machi D, Macken C, Mao C, Niewiadomska A, Nguyen M, Olsen GJ, Overbeek JC, Parrello B, Parrello V, Porter JS, Pusch GD, Shukla M, Singh I, Stewart L, Tan G, Thomas C, VanOeffelen M, Vonstein V, Wallace ZS, Warren AS, Wattam AR, Xia F, Yoo H, Zhang Y, Zmasek CM, Scheuermann RH, Stevens RL. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Research*, 2023; 51: D678–D689.
 13. Bateman A, Martin M-J, Orchard S, Magrane M, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Bye-A-Jee H, Cukura A, Denny P, Dogan T, Ebenezer T, Fan J, Garmiri P, Da Costa Gonzales LJ, Hatton-Ellis E, Hussein A, Ignatchenko A, Insana G, Ishtiaq R, Joshi V, Jyothi D, Kandasamy S, Lock A, Luciani A, Lugaric M, Luo J, Lussi Y, MacDougall A, Madeira F, Mahmoudy M, Mishra A, Moulang K, Nightingale A, Pundir S, Qi G, Raj S, Raposo P, Rice DL, Saidi R, Santos R, Speretta E, Stephenson J, Tootoo P, Turner E, Tyagi N, Vasudev P, Warner K, Watkins X, Zaru R, Zellner H, Bridge AJ, Aimo L, Argoud-Puy G, Auchincloss AH, Axelsen KB, Bansal P, Baratin D, Batista Neto TM, Blatter M-C, Bolleman JT, Boutet E, Breuza L, Gil BC, Casals-Casas C, Echioukh KC, Coudert E, Cuhe B, De Castro E, Estreicher A, Famiglietti ML, Feuermann M, Gasteiger E, Gaudet P, Gehant S, Gerritsen V, Gos A, Gruaz N, Hulo C, Hyka-Nouspikel N, Jungo F, Kerhornou A, Le Mercier P, Lieberherr D, Masson P, Morgat A, Muthukrishnan V, Paesano S, Pedruzzi I, Pilbout S, Pourcel L, Poux S, Pozzato M, Pruess M, Redaschi N, Rivoire C, Sigrist CJA, Sonesson K, Sundaram S, Wu CH, Arighi CN, Arminski L, Chen C, Chen Y, Huang H, Laiho K, McGarvey P, Natale DA, Ross K, Vinayaka CR, Wang Q, Wang Y, Zhang J. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, 2023; 51: D523–D531.
 14. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, 2007; 8: 4.
 15. Dimitrov I, Naneva L, Doytchinova I, Bangov I. AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics*, 2014; 30: 846–851.
 16. Krogh A, Larsson B, Von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology*, 2001; 305: 567–580.
 17. Larsen MV, Lundegaard C, Lamberth K, Buus

- S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC bioinformatics*, 2007; 8: 1–12.
18. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Research*, 2019; 47: D339–D343.
 19. Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, Peters B. Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC bioinformatics*, 2010; 11: 1–12.
 20. Wang P, Sidney J, Dow C, Mothé B, Sette A, Peters B. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS computational biology*, 2008; 4: e1000048.
 21. Saha S, Raghava GPS. Prediction of continuous B cell epitopes in an antigen using recurrent neural network. *Proteins: Structure, Function, and Bioinformatics*, 2006; 65: 40–48.
 22. Saha S, Raghava GP. Prediction methods for B-cell epitopes. *Immunoinformatics: Predicting Immunogenicity In Silico*, 2007; 387–394.
 23. Dhanda SK, Vir P, Raghava GP. Designing of interferon-gamma inducing MHC class-II binders. *Biol Direct*, 2013; 8: 30.
 24. Samad A, Meghla NS, Nain Z, Karpiński TM, Rahman MdS. Immune epitopes identification and designing of a multi-epitope vaccine against bovine leukemia virus: a molecular dynamics and immune simulation approaches. *Cancer Immunol Immunother*, 2022; 71: 2535–2548.
 25. Singh O, Hsu W-L, Su EC-Y. ILeukin10Pred: A Computational Approach for Predicting IL-10-Inducing Immunosuppressive Peptides Using Combinations of Amino Acid Global Features. *Biology (Basel)*, 2021; 11: 5.
 26. Olejnik J, Hume AJ, Mühlberger E. Toll-like receptor 4 in acute viral infection: Too much of a good thing. *PLoS Pathog*, 2018; 14: e1007390.
 27. Dorosti H, Eslami M, Negahdaripour M, Ghoshoon MB, Gholami A, Heidari R, Dehshahri A, Erfani N, Nezafat N, Ghasemi Y. Vaccinomics approach for developing multi-epitope peptide pneumococcal vaccine. *Journal of Biomolecular Structure and Dynamics*, 2019; 37: 3524–3535.
 28. Yang Y, Sun W, Guo J, Zhao G, Sun S, Yu H, Guo Y, Li J, Jin X, Du L, Jiang S, Kou Z, Zhou Y. In silico design of a DNA-based HIV-1 multi-epitope vaccine for Chinese populations. *Human Vaccines & Immunotherapeutics*, 2015; 11: 795–805.
 29. Li X, Guo L, Kong M, Su X, Yang D, Zou M, Liu Y, Lu L. Design and Evaluation of a Multi-Epitope Peptide of Human Metapneumovirus. *Intervirology*, 2015; 58: 403–412.
 30. Livingston B, Crimi C, Newman M, Higashimoto Y, Appella E, Sidney J, Sette A. A Rational Strategy to Design Multi-epitope Immunogens Based on Multiple Th Lymphocyte Epitopes. *The Journal of Immunology*, 2002; 168: 5499–5506.
 31. Huang K, Fu T, Glass LM, Zitnik M, Xiao C, Sun J. DeepPurpose: a deep learning library for drug–target interaction prediction. *Bioinformatics*, 2021; 36: 5545–5547.
 32. Jankauskaite J, Jiménez-García B, Dapkunas J, Fernández-Recio J, Moal IH. SKEMPI 2.0: an updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics*, 2019; 35: 462–469.
 33. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. *Nature*, 2021; 596: 583–589.
 34. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr*, 1993; 26: 283–291.
 35. Huang J, MacKerell AD. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem*, 2013; 34: 2135–2145.
 36. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 2015; 1–2: 19–25.
 37. Harrach MF, Drossel B. Structure and dynamics of TIP3P, TIP4P, and TIP5P water near smooth and atomistic walls of different hydroaffinity. *The Journal of Chemical Physics*, 2014; 140: 174501.
 38. Darden T, York D, Pedersen L. Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics*, 1993; 98: 10089–10092.
 39. Sim(Ana). Analogue Release 2024, Growdea Technologies Pvt. Lt. v1.1, www.growdeatech.com/Analogue (2024).
 40. Trajecta(Ana). Analogue Release 2024, Growdea Technologies Pvt. Lt. v1.1, www.growdeatech.com/Analogue (2024).
 41. Valdés-Tresanco MS, Valdés-Tresanco ME, Valiente PA, Moreno E. gmx_MMPBSA: A

- New Tool to Perform End-State Free Energy Calculations with GROMACS. *J Chem Theory Comput*, 2021; 17: 6281–6291.
42. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8.
43. Kaplan W. Swiss-PDB Viewer (Deep View). *Briefings in Bioinformatics*, 2001; 2: 195–197.
44. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, Beglov D, Vajda S. The ClusPro web server for protein–protein docking. *Nature protocols*, 2017; 12: 255–278.
45. Sanami S, Azadegan-Dehkordi F, Rafieian-Kopaei M, Salehi M, Ghasemi-Dehnoo M, Mahooti M, Alizadeh M, Bagheri N. Design of a multi-epitope vaccine against cervical cancer using immunoinformatics approaches. *Sci Rep*, 2021; 11: 12397.
46. Mustafa AS. In silico Analysis and Experimental Validation of *Mycobacterium tuberculosis*-Specific Proteins and Peptides of *Mycobacterium tuberculosis* for Immunological Diagnosis and Vaccine Development. *Med Princ Pract*, 2013; 22: 43–51.
47. Hasan M, Azim KF, Begum A, Khan NA, Shammi TS, Imran AS, Chowdhury IM, Urme SRA. Vaccinomics strategy for developing a unique multi-epitope monovalent vaccine against Marburg marburgvirus. *Infection, Genetics and Evolution*, 2019; 70: 140–157.
48. Soltan MA, Abdulsahib WK, Amer M, Refaat AM, Bagalagel AA, Dirir RM, Albogami S, Fayad E, Eid RA, Sharaf SMA, Elhady SS, Darwish KM, Eldeen MA. Mining of Marburg Virus Proteome for Designing an Epitope-Based Vaccine. *Front Immunol*, 2022; 13: 907481.
49. Mustafa MI, Shantier SW. Next generation multi epitope based peptide vaccine against Marburg Virus disease combined with molecular docking studies. *Informatics in Medicine Unlocked*, 2022; 33: 101087.