Immunoinformatics Analysis and *In-silico* Design of Multi- Epitopes Vaccine Against Lassa Virus

Akinyemi Ademola Omoniyi (aaomoniyi@abu.edu.ng)
Ahmadu Bello University
Samuel Sunday Adebisi
Ahmadu Bello University
Sunday Abraham Musa
Ahmadu Bello University
James Oliver Nzalak
Ahmadu Bello University
Barnabas Danborno
Ahmadu Bello University
Zainab Maryam Bauchi
Ahmadu Bello University
Iswat Taiwo Badmus
Ahmadu Bello University
Oluwasegun Davis Olatomide
Ahmadu Bello University
Olalekan Jerry Oladimeji
Ahmadu Bello University
Jens Randel Nyengaard
Aarhus University, Aarhus University Hospital

Research Article

Keywords: Lassa virus, RIG-I, Immunoinformatics, Multi-epitopes vaccine, Molecular docking

Posted Date: March 25th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-355782/v1

Abstract

Lassa virus, an arenavirus, represents the most prevalent human pathogen causing viral haemorrhagic fever. It is endemic in Nigeria and other West African countries. Despite the high burden of the disease, limited treatments are available and no approved vaccine for the prevention of this disease is available. In this study, an immunoinformatics approach was used to predict response of B and T cells from the Lassa virus proteome (GPC, NP, L and Z). The designed chimeric vaccine was modeled, refined, validated and docked with the RIG-I receptor. The docked complex of vaccine-RIG-I was subjected to dynamic stability test and the results suggest that the complex is stable. Validation of the final vaccine construct was done through *in silico* cloning using E. coli as host. A CAI value of 0.99 suggests that the vaccine construct expressed properly in the host. Immune simulation predicted significantly high levels of IgG1, T-helper, T-cytotoxic cells, INF-γ and IL-2. This theoretical study suggests infection control by creating an effective immunological memory against Lassa virus infections. However, both in vitro and in vivo experiments are needed to validate the immunogenicity and safety of the chimeric vaccine.

Introduction

Lassa virus, an arenavirus, historically originated from Lassa village in the north-eastern Nigeria, estimated to infect over 300 000 individuals per year across West Africa and responsible for over 3000 deaths[1]. Lassa fever is endemic to the West African countries like Nigeria, Sierra Leone, Liberia, Benin, Ghana, Guinea and Mali[2]. From January to April 2020, over 963 confirmed cases and 188 deaths were reported by the Nigerian Center for Disease Control with the majority of the confirmed cases in Edo and Ondo States [3]. Lassa virus was first discovered in 1969 [4] and is, carried by persistent infection of the reservoir rodent host *Mastomys natalensis*[5]. Infection possibly involves inhalation of contaminated aerosolized rodent excreta and consumption of contaminated food[6]. The virus is transmitted from one person to another via contact of infected bodily fluids[7–10].

Arenaviruses are enveloped negative-strand RNA viruses[11] and the genome consists of two segments. The small segment encodes the glycoprotein precursor (GPC), expressed on the envelope of the virus in a trimeric state[12] and a nucleoprotein in the reverse direction that encapsulates the viral genome. The large segment encodes the viral matrix protein and the viral RNA-dependent RNA polymerase[5]. The posttranslational modification of GPC produces a stable signal peptide (SSP), N-terminal GP1, and transmembrane GP2[13]. GP1 binds to cellular receptors and is important for cell-to-cell proliferation of infection while GP2 mediates viral fusion and structurally resembles class I viral fusion proteins[11]. The Lassa virus moves into the host cell through receptor-mediated endocytosis, and subsequently transported to the endosomal compartments, where fusion occurs at low pH[14,15].

Dystroglycan (DG), the first Lassa virus receptor to be discovered, is universally expressed in a conserved cellular receptor for extracellular matrix proteins[16,17] and found in most tissues in mammals [11]. Produced as a single polypeptide chain, DG undergoes autoprocessing, yielding the peripheral α-DG, serving as a Lassa virus, Lymphocytic choriomeningitis virus glycoprotein and class C new-world arenaviruses receptor [18–20]. The transmembrane β-DG plays important roles in adhesion of extracellular matrix to the cytoskeleton and as a cell adhesion receptor in both muscle and non-muscle tissues [21]. Activation of Lassa virus-specific T cells is

believed to play a vital role during infection [22,23] as strong T-cell response to Lassa virus is critical for protection against Lassa fever [5,24,25]. Despite the high disease burden in the endemic area, there are no prophylactic measures for Lassa virus. Although progress has been made in the identification of promising preclinical candidates, the majority of them have failed in clinical trials. The lack of approved vaccine and limited treatment options makes the development of novel therapeutic strategies against Lassa virus urgently needed.

The experimental methods for vaccine design are challenging and time-consuming, thus Immunoinformatics has emerged as a new tool for identification of the target antigens for vaccine development[26–28]. Immunoinformatics approach can narrow down a vast number of potential molecules to be tested, therefore increasing the chance of finding better candidates. Epitope based vaccines has been reported to be effective in elucidating protective immunity against influenza A[29], hepatitis B[30] and C virus[31], Leishmania[32], Shigella[33] and Mayaro virus[34].

Multi-epitope vaccines ultimate method for the prevention and treatment of tumors or viral infections[35–39]. Multi-epitope vaccines are composed of multiple MHC-restricted epitopes that can be recognized by TCRs of multiple clones from various T-cell subsets [38], cytotoxic T lymphocytes (CTL), Th and B-cell epitopes. This induce strong cellular and humoral immune responses simultaneously. Components with adjuvant, multi-epitopes vaccine can enhance the immunogenicity, provide long-lasting immune responses and reduce unwanted components that can trigger either pathological immune responses or adverse effects[38–43]. This study is using an immune-informatics-driven screening approach, and proteomic data for Lassa virus to design a multi-epitope vaccine that may elicit protective humoral and cellular immune response against Lassa virus.

Methodology

Lassa virus proteome retriever

The complete amino acid sequence of the Lassa virus glycoprotein precursor (E9K9S4), nucleoprotein (A0A097F4S7), viral matrix protein (A0A5J6BM51) and viral RNA polymerase (A0A2P1JNX2) was retrieved from the UniProt database (https://www.uniprot.org/) in standard FASTA format[34]. The overall flow of the work is presented in Fig. 1.

Screening of virulence factor

The Vaxign-ML (http://www.violinet.org/vaxign2) pipeline was applied to compute the protective antigenicity (protegenicity) score and predict the subcellular localization, transmembrane helix and adhesion probability by a vaccine candidate[44]. Vaxign-ML predicts the protegenicity score based on an optimized supervised machine learning model with manually annotated training data consisting of bacterial and viral protective antigens. This method has been validated through nested five-fold cross validation and leave-one-pathogen-out validation to ensure unbiased performance assessment and the ability to predict vaccine candidates against pathogen[44,45].

Cytotoxic T lymphocyte (CTL) epitope prediction

Prediction of CTL epitopes (for the structural polyprotein of Lassa virus) was performed using the online web tool NetCTL v2.0 (http://www.cbs.dtu.dk/services/NetCTL/). The epitope prediction is based on MHC-I binding peptide prediction, proteasomal C-terminal cleavage, and transportation efficiency of Transporter associated with Antigen Processing (TAP). Prediction of MHC-I binding and proteasomal C-terminal cleavage uses artificial

neural networks, while a weight matrix was used to calculate the efficacy of the TAP transporter. Threshold for the prediction of CTL epitopes was set at default value of 0.75[46,47]. To further prove that predicted epitopes can recognize peptides presented on MHC molecules and can be activated and elicit their effector functions, predicted CTL epitopes were subjected to the Class I immunogenicity server (http://tools.iedb.org/immunogenicity/) using the default setting to select the best epitopes for immunogenicity. The tool was only validated for 9-mer peptides for class I immunogenicity [48].

Helper T lymphocytes (HTL) epitope prediction

HTL epitopes for seven human alleles HLA-DRB1*03:01, HLA-DRB1*07:01, HLA DRB1*015:01, HLA-DRB3*01:01, HLA-DRB3*02:02, HLA-DRB4*01:01 and HLA-DRB5*01:01 were predicted from Lassa virus protein by MHC-II prediction module of IEDB (http://tools.iedb.org/mhcii/). The peptide affinity for each receptor is based on IC50 score given to each epitope. Peptides having higher binding affinity must have an IC50 value < 50 nM, whereas the IC50 score < 500 nM and < 5000 nM point to an intermediate and low binding affinity of predicted epitopes, respectively. The score of the percentile rank is inversely linked to the binding affinity of the epitope, i.e., the lower the percentile ranks the higher the binding affinity[49]. To prove the predicted HTL epitopes will have the ability to activate a Th1 type immune response followed by the Interferon-gamma (IFN-γ) production by MHC class II activated $CD4^+$ T helper cells, the predicted HTL epitopes were subjected to the IFN-γ inducing epitopes server (http://crdd.osdd.net/raghava/ifnepitope/index.php) using the predict option. The motif and SVM hybrid was selected as the approach and IFN-gamma versus other cytokine as a model of prediction[50]. IFNs mostly have a protective role against infectious diseases and minimize host damage[51–53].

Allergenicity prediction

Algorithms were used to predict the allergenic score for the predicted CTL epitopes, HTL epitopes and constructed vaccine sequence with great accuracy using the prediction of allergenic proteins and for mapping IgE epitopes on allergenic proteins (AlgPred) (http://www.imtech.res.in/raghava/algpred/). Accuracy acquired for this approach is about 85% at a 0.4 threshold. The server employs six different approaches for the prediction of allergenicity with very high accuracy[54].

Antigenicity prediction

Antigenicity of the predicted CTL epitopes, HTL epitopes and constructed vaccine sequence were predicted using VaxiJen server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) with high accuracy at 0.4 thresholds for virus selected as target organism. The server is based solely on the physiochemical properties for prediction of antigenicity for a given amino acid sequence[47].

Toxicity prediction

Non-toxic nature of the selected CTL epitopes and HTL epitopes were predicted using ToxinPred (http://crdd.osdd.net/raghava/toxinpred/) an online tool for prediction of toxicity of epitopes. The server predicts the toxicity of epitopes based on their physiochemical properties[55].

B cell epitope prediction

B cell epitopes are identified and bound by the receptors associated with the B- lymphocytes on its surface and are characterized as antigenic factors. B-cell epitopes play a much larger role in the host antibody production plan. ABCpred server was used for prediction of linear B cell epitopes

(https://webs.iiitd.edu.in/raghava/abcpred/index.html). The server predicts linear B cell epitopes in the Lassa virus proteome, using an artificial neural network. It is the first server developed based on recurrent neural network (machine based technique) with 65.9% accuracy using fixed length patterns[56]. Maximum accuracy has been obtained using recurrent neural network (Jordan network) with a single hidden layer of 35 hidden units for a window length of 16.

Also, ElliPro an online server (http://tools.iedb.org/ellipro/) was used for prediction of conformational B-cell epitopes (default threshold was 0.5, default maximum distance was 6). ElliPro results are based on Thornton's technique and residue clustering algorithm, ElliPro assigns protrusion index (PI) score to each predicted epitope. Using several ellipsoids, the 3D shape of the epitope is defined and for each residue, the PI value is accurately described based on the center of mass of each residue, which is located in the outer region of the largest promising ellipsoid. A residue having larger value has better solvent accessibility[57].

Construction of multi-epitope vaccine sequence

A vaccine sequence was carefully constructed using the CTL, HTL and linear B cells epitopes predicted by the above-employed immunoinformatics approaches. These CTL, HTL and B cells epitopes were joined with the help of AAY, GPGPG and KK[58] used as linkers respectively, and adjuvant was added with the help of the EAAK linker[34,59] in Notepad + + version 7.8. (Fig. 2).

Physiochemical properties and domain identification

Physiochemical properties such as amino acid composition of the vaccine, theoretical pI, molecular weight, instability index, half-life (*in vitro* and *in vivo*), aliphatic index and grand average of hydropathicity (GRAVY) of the final vaccine construct were evaluated using an online web tool the ProtParam (http://web.expasy.org/protparam/) [60,61]. Additionally, the solubility of the vaccine construct upon overexpression in E. coli was predicted by the SOLpro (http://scratch.proteomics.ics.uci.edu/) tool in the SCRATCH suite[61,62].

Secondary structure of construct vaccine prediction

PSIPRED, a freely available web tool (http://bioinf.cs.ucl.ac.uk/psipred/) for protein secondary structure prediction for a given amino acid sequences with high accuracy was used for the prediction of the secondary structure of construct vaccine. PSIPRED 4.0 uses position-specific iterated BLAST (Psi-Blast) to identify and select sequences showing significant homology to the vaccine protein and accomplished a normal Q3 score of 81.6% when an exceptionally stringent cross-approval strategy was utilized to assess its performance[63,64].

Tertiary structure of construct vaccine prediction

A free online server, RaptorX (http://raptorx.uchicago.edu/) was used to predict the 3D structure of the vaccine sequence. The server provides automated tools for protein structure prediction and analysis. RaptorX also predicts secondary structure, contacts, solvent accessibility, disordered regions and binding sites for the input sequence[65]. For visualization of the predicted 3D model, Pymol software was used.

Tertiary structure of construct vaccine refinement

The 3D model obtained for the chimeric vaccine peptide was refined using the GalaxyRefine server (http://galaxy.seoklab.org/). The sever utilized the CASP10 refinement method for reconstruction of the protein's side chains, repacking as well as molecular dynamic simulations for relaxation of the 3D structure. Galaxy refine

is one of the best available servers for the overall improvement of both the global and the local quality of the given structure generated by utilizing the best protein structure prediction in available web tools[66,67].

Tertiary Structure of Construct Vaccine Validation

Model validation, a critical step in the model building process as it detects potential errors in the predicted 3D models[64] was done through online servers in three stages. Firstly the ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php), calculated an overall quality score for a specific input structure, and displayed it in the context of all known protein structures. Also, any problematic parts of a structure are shown and highlighted in a 3D molecule viewer. If the calculated score falls outside the range characteristic of native proteins the structure likely contains errors[34,68]. Secondly, the SAVES server v6.0 (https://saves.mbi.ucla.edu/) generated the overall quality score of the modeled protein by analyzing non-bonded atom-atom interactions and compared it to reliable high-resolution crystallography structures[34,69].

Thirdly, PROCHECK analysis was carried out on (https://saves.mbi.ucla.edu/) generating a Ramachandran plot in order to visualize energetically allowed and disallowed dihedral angles psi (ψ) and phi (ϕ) of an amino acid. The calculations are based on the van der Waal radius of the side chains. The server uses the PROCHECK principle to validate a protein structure by using a Ramachandran plot and separates plots for Glycine and Proline residues[70,71].

Molecular Docking of the Refined Vaccine Construct and RIG-I Receptor

Initiation of an appropriate immune response depends on the interaction of an antigenic molecule and a specific immune receptor. Molecular docking is a computational approach that involves interaction between ligand and receptor to provide a stable adduct with a calculated score measuring the degree of binding interaction[71,72]. Retinoic acid-inducible gene RIG-I-like receptors (RLRs) can mediate high proinflammatory responses against Lassa virus infection[73,74]. Hence, the 3D structure of the RIG-I receptor was retrieved from a protein data bank (PDB ID: 2QFB) and used as the receptor[75]. The binding pockets in the RIG-I receptor was predicted by use the CASTp server (http://sts.bioe.uic.edu/castp/calculation.html). In CASTp, voids are defined as buried, unfilled empty spaces inside proteins following removal of all heteroatoms inaccessible to water molecules (modelled as a spherical probe of 1.4 Å) from the outside and pockets are concave caverns with constrictions at the opening on the surface regions of proteins. CASTp is excellent at providing identification and measurements of surface accessible binding pockets along with the information of inner inaccessible cavities for protein molecules[76].

Molecular docking of the refined vaccine peptide with the RIG-I receptor was performed using the PatchDock (http://bioinfo3d.cs.tau.ac.il/PatchDock/) server. The molecular docking calculation by PatchDock involves four basic stages, i.e. atomic shape portrayal, surface fixes coordination as well as separation and scoring. PatchDock separates the surface of both the input molecules into small patches as per the proper surface shape. Only after the identification of all these small patches, their superposition is accomplished by the use of shape matching algorithms [77]. Furthermore, the FireDock (Fast Interaction Refinement in Molecular Docking) web tool (http://bioinfo3d.cs.tau.ac.il/FireDock/) was used to optimize as well as re-score the rigid body molecular docking solutions. It provides the best ten solutions for final refinement based on the binding value[78]. PDBsum (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html) and PDBePisa

(https://www.ebi.ac.uk/pdbe/pisa/) were employed to generate the graphical presentation of interaction between the docked complex[79,80].

Molecular Dynamics Simulation

The study of molecular dynamics (MD) is essential for examining the stability of the protein–protein complex in silico analysis. MD simulation analysis of final vaccine construct and vaccine-RIG1 complex was done using GROMACS software (version 20.4)[81,82]. Charmm36[83] force field was employed for the simulation process, using water as a solvent. Cubic and dodecahedron boxes were made with a minimum distance of 1 Å between the surfaces and edges of the vaccine construct and the vaccine-RIG1 complex respectively, and were solvated with the TIP3 water model[84]. After neutralizing the system with Na^+ and Cl^- ions using the Genion tool, energy minimization was carried for all the systems. The systems were equilibrated for 100 ps isothermal-isochoric (NVT) ensemble followed by 200 ps isothermal-isobaric (NPT) ensemble maintaining a constant 300 K temperature and 1 atm pressure. Finally, MD simulation was carried out for 50 ns with a time interval of 2 fs for both vaccine construct and vaccine-RIG1 complex. Root mean square deviation (RMSD) and root mea square fluctuation (RMSF) analysis were performed to examine the standard deviation and fluctuation of the protein backbone, and radius of gyration was performed to examine the folding compactness of the vaccine construct.

In silico Cloning Optimization of Vaccine Construct Candidate

Reverse translation and codon optimization were performed using the Java Codon Adaptation Tool server (http://www.jcat.de/) to express the multi-epitope vaccine construct in a selected expression vector [85]. Codon optimization was performed to express the final vaccine construct in the E. coli (strain K12) host, as the codon usage of E. coli differs from that of the native host Lassa virus, where the sequence of the final vaccine construct is derived. Three additional options were selected to avoid the rho-independent transcription termination, prokaryote ribosome binding site, and restriction enzymes cleavage sites. JCat output includes the codon adaptation index (CAI) and percentage GC content, which can be used to assess protein expression levels. CAI affords information on codon usage biases; the ideal CAI score is 1.0 but > 0.8 is considered a good score[86]. The GC content of a sequence ranges between 30–70%. GC content values outside this range suggest unfavourable effects on translational and transcriptional efficiencies[87]. Cloning the optimized gene sequence of the final vaccine construct in E. coli pET-30a (+) vector, Nde I and Xho I restriction sites were introduced to the N and C-terminals of the sequence, respectively. Finally, the optimized sequence with restriction sites was inserted into the pET-30a (+) vector using SnapGene tool to ensure vaccine expression[71].

Immune Simulation

Proven the immunogenicity and immune response profile of the vaccine construct further, in silico immune simulations were conducted using the C-ImmSim server (http://150.146.2.1/C-IMMSIM/index.php). C-ImmSim is an agent-based model that uses a position-specific scoring matrix (PSSM) for immune epitope prediction and machine learning techniques for prediction of immune interactions. It “simultaneously simulates three compartments that represent three separate anatomical regions found in mammals: (i) the bone marrow, where hematopoietic stem cells are stimulated to produce new lymphoid and myeloid cells; (ii) the thymus, where naive T cells are selected to avoid auto immunity; and (iii) a tertiary lymphatic organ, such as a lymph node”[88]. Three injections were given at intervals of four weeks[89]. All simulation parameters were set at default with time steps set at 1, 84, and 168 (each time step is 8 hours and time step 1 is injection at time = 0). Furthermore, 12

injections of the designed peptide were given four weeks apart to simulate repeated exposure to the antigen seen in a typical endemic area in order to probe for clonal selection.

Results

Retrieval of Proteins Sequences

Amino acid sequences for four proteins (glycoprotein precursor, nucleoprotein, viral matrix protein and viral RNA polymerase) of Lassa virus were retrieved from the UniProt database and used to predict the B and T cell epitopes for designing the multi-epitope subunit vaccine. Human β-Defensin 1 was retrieved from the UniProt database (P60022) and used as an adjuvant for the immune interaction based on its ability to induce antiviral immune response[90].

Screening of Virulence Factor

Vaxign-ML analysis predicted four Lassa virus proteins (GPC, NP, Z, and L protein) as protective antigen with high protegenicity score. GPC and Z proteins were predicted as adhesive proteins (Table 1). Adhesin plays a vital role in the virus adhering to the host cell and enabling the virus entry to the host cell[91]. All the predicted proteins were different to human, mouse or pig proteins.

Table 1
Vaxign-ML prediction and adhesion probability of Lassa virus proteins.

Protein code	Protein name	Protegenicity score	Adhesion probability
GPC	Glycoprotein precursor	89.6	0.633
Z	Viral matrix	76.3	0.583
NP	nucleoprotein	94.9	0.204
L	Viral RNA polymerase	89.6	0.168

Cytotoxic T Lymphocyte (CTL) Epitope Prediction

A total of 96 unique CTL epitopes (9-mer) were predicted from the four Lassa virus proteins using the NetCTL v2.0 server. Among them, only 42 epitopes were found as antigenic, immunogenic, and non-toxic while 24 epitopes were found to be non-allergenic (Supplementary Table 1). From these, 12 non-allergenic epitopes were selected based on high immunogenicity scores as CTL epitopes for vaccine construction (Table 2).

Table 2
Selected cytotoxic T-lymphocyte (CTL) epitopes for multi-epitope vaccine construction

ID	Peptide Seq.	Combine score	Antigenicity	Immunogenicity	Allergenicity	Toxicity
1	ITEVEYLLY	3.692	0.960	0.152	Non-allergen	Non-toxic
2	LIDGRTTFY	2.943	1.073	0.229	Non-allergen	Non-toxic
3	RTLLEFHYY	1.860	0.723	0.218	Non-allergen	Non-toxic
4	YADLNFGEK	0.953	2.604	0.187	Non-allergen	Non-toxic
5	ESTTNPPPY	1.136	0.793	0.019	Non-allergen	Non-toxic
6	ALTDLGLIY	1.802	1.284	0.124	Non-allergen	Non-toxic
7	MRMAWGGSY	0.853	0.763	0.162	Non-allergen	Non-toxic
8	CSTTYKGVY	2.624	0.412	-0.102	Non-allergen	Non-toxic
9	MTSYQYLII	1.588	0.601	-0.105	Non-allergen	Non-toxic
10	SWDCIMTSY	1.518	1.430	-0.147	Non-allergen	Non-toxic
11	FSRPSPIGY	1.254	1.612	-0.034	Non-allergen	Non-toxic
12	VQYNLSHTY	0.844	0.923	-0.124	Non-allergen	Non-toxic

Helper T Lymphocyte (HTL) Epitope Prediction

HTL epitopes for seven human alleles HLA-DRB1*03:01, HLA-DRB1*07:01, HLA DRB1*015:01, HLA-DRB3*01:01, HLA-DRB3*02:02, HLA-DRB4*01:01 and HLA-DRB5*01:01 were predicted by the MHC-II prediction module of IEDB. A total of 123 HTL epitopes (15-mer) were predicted from the four Lassa virus proteins and were found to be non-allergenic, antigenic and non-toxic (Supplementary Table 2). Among them, only 23 were found to be IFN-γ positive epitopes. From these, 13 non-overlapping epitopes were selected based on high percentile rank scores as HTL epitopes for vaccine construction (Table 3)

Table 3
Selected helper T-lymphocyte (HTL) epitopes for multi-epitope vaccine construction

S/N	Allele	Peptide	Pt rank	Allergenicity	Antigenicity	Toxicity	IFN-γ
1	HLA-DRB1*15:01	EKHILPFMSDLASTK	3.30	-0.55	0.569	Non-toxic	Positive
2	HLA-DRB3*01:01	ERPMFKFDLMMGYAW	0.46	-0.53	0.426	Non-toxic	Positive
3	HLA-DRB1*03:01	IQSSGGSLRLKGRTC	38.00	-0.41	0.698	Non-toxic	Positive
4	HLA-DRB5*01:01	ISYMCHFITKETPDR	34.00	-0.57	0.522	Non-toxic	Positive
5	HLA-DRB3*01:01	QTKDPKAESSPRASL	66.00	-0.81	0.688	Non-toxic	Positive
6	HLA-DRB4*01:01	AQPGLTSAVIEALPR	27.00	-0.10	0.486	Non-toxic	Positive
7	HLA-DRB3*02:02	GWPYIASRTSIVGRA	0.35	-0.66	0.977	Non-toxic	Positive
8	HLA-DRB5*01:01	LVELKSTQQKSILRV	14.00	-0.60	0.4328	Non-toxic	Positive
9	HLA-DRB4*01:01	IRRLKTEAQMSIQLI	0.68	-0.46	0.508	Non-toxic	Positive
10	HLA-DRB1*03:01	RDIYISRRLLGTFTW	1.40	-0.53	0.540	Non-toxic	Positive
11	HLA-DRB1*03:01	DIYISRRLLGTFTWT	3.00	-0.44	0.445	Non-toxic	Positive
12	HLA-DRB1*07:01	WEDHCQFSRPSPIGY	4.70	-0.53	0.892	Non-toxic	Positive
13	HLA-DRB1*07:01	NVATCGLFGLVSFLL	7.80	-0.52	0.536	Non-toxic	Positive
14	HLA-DRB1*07:01	GIYNVATCGLFGLVS	9.70	-0.49	0.605	Non-toxic	Positive

Pt rank: Percentile rank score, IFN-γ: Interferon-gamma

Toxicity prediction

The non-toxic nature of selected 12 CTL and 13 HTL epitopes were confirmed using the ToxinPred server. The results obtained from ToxinPred for the 12 CTL epitopes are shown in Table 1, and 13 HTL epitopes are shown in Table 2. These epitopes are considered safe for use.

Construction of multi-epitope subunit vaccine

The selected 12 CTL epitopes were connected by an AAY linker, adjuvant protein (Human β-Defensin 1) joined with the first CTL epitopes using an EAAAK linker and 13 HTL selected epitopes connected by GPGPG linker to form the final vaccine construct. The final vaccine construct contained 494 amino acid residues and the structure is shown in Fig. 2.

B Cell Epitopes Prediction for Lassa Virus Protein

Among Linear B cell epitopes predicted by the ABCpred server, six epitopes of 16-mer lengths that were predicted to be antigen, non - allergenic and non – toxic were selected (Table 3). The Ellipro suit was utilized to predict the conformational B-cell epitopes. A total of five discontinuous B-cell epitopes of varying residue length were predicted, out of which three epitopes of residue length 48, 47 and 55 residues with a score of 0.721, 0.688 and 0.693 respectively, were evaluated (Fig. 3). Moreover, in the first chain of the discontinuous epitope, the start and end residues were methionine and glycine with residue score of 0.630 and 0.073, respectively. The second chain started with alanine and ended with lysine showing a residue score of 0.444 and 0.955. The third chain started with alanine and ended with proline with the residue score of 0.357 and 0.357.

Table 4
Linear B-cell epitopes with 16-mer lengths predicted by the ABCPred server

S/N	Sequence	Score	Antigenicity	Antigenicity	Toxicity
1	TKDPKAESSPRASLIP	0.91	0.8703	-0.57	Non-toxic
2	QRTRDIYISRRLLGTF	0.81	0.6425	-0.41	Non-toxic
3	DHCQFSRPSPIGYLGL	0.75	0.9244	-0.43	Non-toxic
4	LGLVDLFVFSTSFYLI	0.63	0.4118	-0.52	Non-toxic
5	RPSPIGYLGLLSQRTR	0.62	2.0552	-0.40	Non-toxic
6	EYIDRQGKTPLGLVDL	0.61	0.9843	-0.47	Non-toxic

Allergenicity of the vaccine construct

The AlgPred server was used to check the allergenicity of the vaccine construct. Allergenicity score of the constructed multi-epitope vaccine was − 0.846. The scores obtained for each CTL and HTL epitopes from the server are presented in supplementary Tables 1 and 2.

Antigenicity of the Vaccine Construct

VaxiJen server was used to predict the antigenic nature of our final multi-epitope vaccine. The predicted score was 0.65. The results suggest that the vaccine candidate possess strong antigenic properties that will help to provoke the immune response. The scores obtained for each CTL and HTL epitopes from the server are presented in supplementary Tables 1 and 2.

Prediction of Physiochemical Parameters and Solubility

Physiochemical properties of the final vaccine construct obtained from ProtParam server indicate a molecular weight of 65.2 kDa and theoretical protrusion index (PI) of 9.51 revealing that the vaccine construct is basic in nature. The estimated *in vivo* half-life in E. coli was greater than 10 hours, instability index was 38.9, which implies that the sequence of the construct will remain stable after expression. The aliphatic index was 74.3

revealing a thermostable nature at varying temperatures. The grand average of hydropathicity (GRAVY) was calculated to be negative (− 0.185). This negative value indicates a hydrophilic nature of the protein; thus, this protein tends to have better interaction with other proteins. The solubility as calculated by the SOLpro server was 0.60, Table 4.

Table 4
Physiochemical properties of the vaccine construct protein

S/N	Physiochemical Properties	Results
1	Number of amino acids	602
2	Molecular weight	65.23 kDa
3	Theoretical protrusion index (PI)	9.51
4	Total number of negatively charged residues (Asp + Glu)	35
5	Total number of positively charged residues (Arg + Lys)	66
6	Chemical Formula	$C_{2964}H_{4560}N_{782}O_{826}S_{27}$
7	Total number of atoms	9159
8	Estimated half-life (mammalian reticulocytes, in vitro)	30 hours
9	Estimated half-life (yeast, in vivo)	> 20 hours
10	Estimated half-life (Escherichia coli, in vivo)	> 10 hours
11	Instability index	38.9
12	Aliphatic index	74.3
13	Grand average of hydropathicity (GRAVY)	-0.185
14	Solubility upon overexpression	0.60

Secondary Structure Prediction

The PSIPRED server was used to predict the secondary structure of the vaccine construct. The results obtained from the server revealed that the multi-epitopes vaccine contains alpha helix (32.1%), beta-strand (15.6%) and coil (52.3%) (Fig. 4). The multi-epitope vaccine also contains small nonpolar (43.7%), hydrophobic (19.8%), polar (21.9%) and aromatics plus cysteine (14.6%) residues.

Figure 4: Graphical representation of secondary structure prediction of the multi-epitope vaccine. The protein is predicted to have an alpha helix (34.4%), a beta-strand (15.8%), and a coil (49.8%).

Modelling of Tertiary Structure of Construct Vaccine

The RaptorX server predicted five tertiary structure models of the designed chimeric protein based on Multi-template based approach and was used for the homology model. The five predicted models had estimated RMSD ranging from 10.50 to 12.53 Å. The model with the best estimated RMSD from the homology modelling was selected for further refinement.

Tertiary Structure Model Refinement

Refinement of the vaccine construct model 3D structure using GalaxyRefine server generated five models. Based on model quality scores for all refined models, model 1 was selected to be the best based on various parameters including GDT-HA (0.897), RMSD (0.543), MolProbity (3.141), Clash score (59.2), Poor rotamers score (4.0), and Ramachandran plot (93.2 %). This model was chosen as the final vaccine model for further analysis. Figure 5.

Tertiary Structure Model Validation

The quality and potential errors in the vaccine construct 3D model were verified by ProSA-web and SAVES. ProSA-web analysis of the chosen model after refinement had a Z-score of -8.27 (Fig. 6A) while SAVES revealed an overall quality factor of 66.7%. The Ramachandran plot analysis of the model protein revealed that 86.8% of residues in the protein are in the most favoured regions and 12.4% in additional allowed regions. Additionally, 0.6% of the residues were predicted to be in generously allowed regions and only 0.2% in disallowed regions (Fig. 6B).

Molecular Docking of Refined Model Vaccine with Immune Receptor (RIG-I).

The CASTp server was used to determine the protein binding and hydrophobic interaction sites on the protein surface. The molecular surface area of the pocket was 3858 $Å^2$ with a molecular surface volume of 4679 $Å^3$, the mouth molecular area was about 7014 $Å^2$, and the molecular circumference sum was 109 Å. To assess the interaction between the refined model and the RIG-I (PDB ID: 2QFB) immune receptor, molecular docking was performed by the use of the PatchDock server. Hundred models were generated, which were scored based on their protein surface, geometry and electrostatic complementarity. The top 10 complexes were further submitted to the FireDock server for refinement and rescoring the docking solutions of RIG-I and multi-epitope vaccine candidates. The best docking solution was selected for further analysis from the molecular docking studies based on the binding score. The selected complex of the multi-epitope vaccine candidate and RIG-I had a binding energy of − 0.01 kcal/mol, the attractive van der Waals force was − 28.69 kcal/mol, the repulsive van der Waals force was 21.42 kcal/mol, and the atomic contact energy was 8.17 kcal/mol. The online database PDBePISA was used to generate the interaction between the docked complexes and ten interaction places with hydrogen bonds were noted. Structural analysis of RIG-I and multi-epitope vaccine also revealed that Arg886 formed a hydrogen bond with Met172, Gln300 and Thr301 at a distance of 2.58 Å, 3.80 Å and 3.80 Å respectively, while Tyr819 – Gly295 formed hydrogen bond at 3.42 Å. Similarly, Thr899 – Ser326 and Ile897-Ser326 develops hydrogen bond at 3.14 Å and 2.09 Å, Asp836 - Arg288 and Glu840 – Lys245 at 2.37 Å and 3.43 Å. Also, Gln867 – Arg171 formed hydrogen bond at 3.19 Å and Ile897 – Ala327 at 3.70 Å. (see Fig. 7).

Molecular Dynamics Simulation of the RIG-I - Multi-Epitope Vaccine Complex

The RIG-I-final vaccine complex was subjected to molecular dynamics simulation for stability and residual fluctuation test. RMSD and RMSF of this complex was calculated and presented in Fig. 8 while RMSD and RMSF graphs for the final vaccine construct are given in supplementary Fig. 1. The average RMSD for the system for the time period of 50 ns was found to be 0.88 Å. RMSF for the side chain was found to show stability except for a few places, which are located in the loop region of the vaccine construct, showing higher fluctuation. The radius of gyration is also within the acceptable range.

Codon Optimization and *In silico* Cloning of the Final Vaccine Construct

The Java codon adaptation tool was used to optimize codon usage of the vaccine construct in E. coli (strain K12) for maximal protein expression. The optimized codon sequence was 1,806 nucleotides long. The Codon adaptation index (CAI) of the optimized nucleotide sequence was 0.99, and the average GC content of the adapted sequence was 50.7% indicating good expression of the vaccine candidate in the E. coli host. The optimal percentage range of GC content (54.9%) is between 30% and 70%. Lastly, the sequence of the recombinant plasmid was designed using the SnapGene software by inserting the adapted codon sequences into the pET30a (+) vector (Fig. 9).

In Silico Immune Simulation

The immune simulation response was comparable with actual immune responses (Fig. 10), with the tertiary and secondary responses higher compared to the primary response. High concentrations of antigen count was observed at the primary response. In both the secondary and tertiary reactions, the typical high levels of immunoglobulin activities (i.e., IgG1 + IgG2, IgM, and IgG + IgM antibodies) were obvious with associated antigen reduction (Fig. 10A). This indicates the development of immune memory and consequently increased antigen clearance upon subsequent exposures (Fig. 10E). Additionaly, several long-lasting B-cell isotypes were observed, suggestive of the potential for isotype switching and memory formation (Fig. 10B). For the T-cytotoxic cell populations with the respective memory development, a similarly elevated response was noticed (Fig. 10F). In the course of exposure, continuous proliferating dendritic cells was observed (Fig. 10C), high levels of IFN-γ and IL-2 were also apparent with a low Simpson index (D) (Fig. 10D).

Discussion

Following Lassa virus infection in humans, 80% of the cases are mostly presented as a mild non-descript disease while 20% of infections result in severe hemorrhagic fever with multi-organ failure[92]. Lassa fever incubation period is usually 1–3 weeks with accompanying fever, fatigue, hemorrhaging, gastrointestinal symptoms (vomiting, diarrhea, and stomachache), respiratory symptoms (cough, chest pain, and dyspnea), and neurologic symptoms (disorientation, seizures, and unconsciousness)[7]. Presently, treatment is limited to supportive care and the use of ribavirin, which reduces mortality if administered early[93]. The use of Ribavirin is restricted to high-risk patients due to its adverse effects and mixed efficacy[93,94], making the development and testing of effective vaccines for Lassa virus critical. The severe spike in Lassa virus infection cases observe in West Africa, particularly Nigeria, usually reaches its peak during the dry season (December – May) annually[7,95]. Due to its high lethality, Lassa virus is considered a category ‘A’ agent by the Centers for Disease Control and Prevention[96].

The Coalition for Epidemic Preparedness Innovations and World Health Organization has identified Lassa virus as an emerging threat to global health and as a disease that is in crucial need of attention in the area of research and development[97,98] and this has lead to a push for several preclinical vaccine candidates toward clinical trials. The Bill and Melinda Gates Foundation and their collaborators founded the Coalition for Epidemic Preparedness Innovations (CEPI) along with the World Health Organization (WHO) with the aim the of advancing three priority pathogens with likely preclinical candidates to Phase III clinical trials and Lassa virus was part of the top priority pathogens[5]. Most of the vaccine candidates target the glycoprotein of Lassa virus, thus only expressing this protein and not those known to modulate innate immunity. ML29 reassortant vaccine and

recombinant yellow fever vaccine (YF17D) have generated substantial safety concerns with regards to the genetic stability[99,100]. Sterilizing immunity was not achieved[101] while vesicular stomatitis virus vaccine (VSV) awaits clinical trials [102]. Despite the sizeable list of preclinical candidates for Lassa virus prevention, none have yet shown combined efficacy and safety in humans.

Recently, attention has shifted towards the development of subunit vaccines because of improved safety profiles and logistical feasibility associated with them[103]. With the immunoinformatics approach, multi-epitope based vaccines represent a novel approach for producing a specific immune response and avoiding responses against other unfavourable epitope antigens[104]. Other potential advantages of multi-epitope based vaccines also include better safety, the chance to rationally engineer the epitopes for increased potency, and the ability to focus immune responses on preserved epitopes[105].

This work therefore focussed on the *in silico* design and development of a potential multi-epitope antigen peptide against Lassa virus using four proteins expressed in the uniprot database. It has been reported that immunity to Lassa virus is dependent on T-cells[22,23]. The role for antibody-dependent cell-mediated cytotoxicity[22,106] and the RIG-I pathway in immunity to Lassa virus has also been reported[5,107,108]. Using several servers, CTL and HTL epitopes were selected based on their antigenicity, allergenicity, immunogenicity and toxicity for the mult-epitope candidate. Additionally, only HTL epitopes that release other types of cytokines such as interferon-gamma (IFN-γ) were chosen for the multi-epitopes construct as IFN-γ secretion has been shown to be an an important mediator of protection against the Lassa virus. A multi-epitope peptide was formed by joining the selected epitopes with appropriate linkers[109]. The AAY and GPGPG linkers[34,71] were applied between the predicted epitopes to generate sequences with minimized junctional immunogenicity, therefore, allowing the rational design of a potent multi-epitope vaccine[109]. The EAAAK linker[110], was also introduced between the adjuvant sequence and the fused epitopes in order to generate a high level of expression and improved bioactivity of the fusion protein[71]. Adjuvants are used as supplements in vaccine formulations to develop specific immune responses to antigens, and enhance the stability, longevity of the vaccine against infection[111].

The molecular weight (~ 65 kDa) of our vaccine candidate is within the average molecular weight (40–70 kDa) for a multiepitope vaccine[34]. Given that solubility of the overexpressed recombinant protein in the E. coli host is one of the vital requirements for many biochemical and functional analysis[64], the vaccine construct was predicted to be soluble upon expression suggesting easy access to the host. The theoretical pI value suggests that the vaccine is basic in nature and the instability index revealing that the vaccine will remain stable after expression[112], therefore, enhancing its usage capacity further. The hydrophilicity and thermostability as revealed by the GRAVY score and aliphatic index makes it well-matched for use in endemic areas, most of which are found in West Africa. Naturally unfolded protein regions and alpha-helical coiled peptides have been reported to be important forms of structural antigens[113], making the knowledge of secondary and tertiary structures of the target protein crucial to vaccine design[109]. Secondary structure analyses showed that the protein consisted predominantly of coils (52.3%). Ramachandran plot of the refined 3D structure of the vaccine candidate shows that 86.8% of the residues are located within the favoured and allowed regions with very few residues in the outlier region, demonstrating that the quality of the overall model is satisfactory[114]. Different structure validation tools were applied to identify errors in the modelled vaccine construct. The Z-score of − 8.27 and ERRAT quality factor of 86.5% from SAVES server revealed that the overall structure of the refined vaccine is acceptable[34].

Since the RIG-I receptor has proven recognition capability in Lassa virus[74], the molecular interaction of vaccine construct through docking analysis suggested that the candidate vaccine has significant affinity to RIG-I to act as a sensor for recognizing molecular patterns of pathogen and initiating immune response[73]. Consequently, the vaccine construct RIG-I complex is capable of generating an effective innate and adaptive immune response against Lassa virus.

Molecular docking analysis and dynamics simulation carried out established potential immune interaction and stability between RIG-I and the multi-epitopes vaccine construct. Energy minimization conducted to minimize the potential energy of the whole system confirmed the complete conformational stabilization of the RIG-I-multi-epitopes vaccine construct docked complex. The energy minimizes the inappropriate structural geometry by exchanging individual protein atoms, therefore making the structure more stable with suitable stereochemistry. The derived RMSD and RMSF indicates the complex's stability[34]. Candidate vaccine construct demands validation by screening for immunoreactivity through serological analysis[115]. This includes the expression of the recombinant protein in Escherichia coli expression systems, as the systems are suitable for the production of recombinant proteins[116,117]. Codon optimization carried out to attain high-level expression of our recombinant vaccine protein in E. coli (strain K12), shows that the codon adaptability index (0.99) and the GC content (54.9%) were favourable for high-level expression of the protein in bacteria. Immune simulation results revealed consistent reaction with typical immune responses[112]. Following repeated exposure to the antigen, an increase in the immune responses was observed. Development of memory B-cells and T- cells was apparent, with memory B-cells lasting for several months and helper T cells stimulated. Increase in levels of IFN-γ and IL-2 following the first injection maintained the peak levels after repeated exposures to the antigen. This indicates high levels of TH cells and thus efficient Ig production, associated with humoral response[118]. The Simpson index D for analysis of clonal specificity suggests a possible diverse immune response[119]. This is reasonable considering that the constructed chimeric peptide is composed of several B and T epitopes.

Conclusion

In this study, an immunoinformatics approach was used to design a potential vaccine peptide coding for multiple T-cell (HTL and CTL) epitopes. Given that the proteins containing these epitopes are expressed in Lassa virus, the vaccine peptide will potentially provide prophylactic benefits. The interaction and binding potential between chimeric vaccine protein and immune receptor were high and stable. Effective immune responses in real life were modeled and observed in immune simulation. This study will potetially aid infection control by creating an effective immunological memory against Lassa virus infections. The next step is to express this peptide in a bacterial system and perform the various immunological assays needed to validate the results obtained.

Declarations

Author Contributions: AAO, SAM and ZMB conceived and designed the analysis; AAO, ITB, ODO and OJO performed immunoinformatic analyses; AAO prepared illustrations and wrote the manuscript; SSA, JON, BD and JRN contributed to the critical revision of the manuscript; JRN supervised the whole work; and all authors approved the final manuscript.

Compliance with Ethical Standards

Conflict of interest: All authors have declared that they have no conflict of interest.

Research Involving Human Participants and/or Animals: This article does not contain any studies involved with human participants or animals performed by any of the authors.

Data Availability

All data generated or analysed during the study are included in the submitted manuscript. The sequences of the protein analysed can be retrieved from UniProt database using their UniProt ID.

References

1. Geisbert, T. W. Predicting outcome and improving treatment for Lassa fever. *Lancet Infect. Dis.* **18**, 594–595 (2018).
2. Ajayi, N. A. *et al.* Containing a Lassa fever epidemic in a resource-limited setting: Outbreak description and lessons learned from Abakaliki, Nigeria (January-March 2012). *Int. J. Infect. Dis.* (2013) doi:10.1016/j.ijid.2013.05.015.
3. NCDC. An update of Lassa fever outbreak in Nigeria. https://ncdc.gov.ng/reports/138/2018-august-week-32%0Ahttps://ncdc.gov.ng/news/165/decline-in-lassa-fever-cases-as-ncdc-sustains-intensive-response%0Ahttps://ncdc.gov.ng/projects (2020).
4. Frame, J. D., Baldwin, J. M., Gocke, D. J. & Troup, J. M. Lassa fever, a new virus disease of man from West Africa. I. Clinical description and pathological findings. *Am. J. Trop. Med. Hyg.* (1970) doi:10.4269/ajtmh.1970.19.670.
5. Warner, B. M., Safronetz, D. & Stein, D. R. Current research for a vaccine against Lassa hemorrhagic fever virus. *Drug Des. Devel. Ther.* **12**, 2519–2527 (2018).
6. Grant, D. S., Khan, H., Schieffelin, J. & Bausch, D. G. Lassa Fever. in *Emerging Infectious Diseases: Clinical Case Studies* (2014). doi:10.1016/B978-0-12-416975-3.00004-2.
7. Richmond, J. K. & Baglole, D. J. Lassa fever: Epidemiology, clinical features, and social consequences. *British Medical Journal* (2003) doi:10.1136/bmj.327.7426.1271.
8. Gibb, R., Moses, L. M., Redding, D. W. & Jones, K. E. Understanding the cryptic nature of Lassa fever in West Africa. *Pathogens and Global Health* (2017) doi:10.1080/20477724.2017.1369643.
9. WHO. Lassa fever fact sheet. https://www.who.int/en/news-room/fact-sheets/detail/lassa-fever (2019).
10. Nunberg, J. H. & York, J. The curious case of arenavirus entry, and its inhibition. *Viruses* **4**, 83–101 (2012).
11. Torriani, G., Galan-Navarro, C. & Kunz, S. Lassa Virus Cell Entry Reveals New Aspects of Virus-Host Cell Interaction. *J. Virol.* **91**, 1–8 (2017).
12. Schlie, K. *et al.* Characterization of Lassa Virus Glycoprotein Oligomerization and Influence of Cholesterol on Virus Replication. *J. Virol.* (2010) doi:10.1128/jvi.02039-09.
13. Rojek, J. M. *et al.* Targeting the Proteolytic Processing of the Viral Glycoprotein Precursor Is a Promising Novel Antiviral Strategy against Arenaviruses. *J. Virol.* **84**, 573–584 (2010).

14. Cosset, F.-L. *et al.* Characterization of Lassa Virus Cell Entry and Neutralization with Lassa Virus Pseudoparticles. *J. Virol.* (2009) doi:10.1128/jvi.01711-08.
15. Klewitz, C., Klenk, H. D. & ter Meulen, J. Amino acids from both N-terminal hydrophobic regions of the Lassa virus envelope glycoprotein GP-2 are critical for pH-dependent membrane fusion and infectivity. *J. Gen. Virol.* (2007) doi:10.1099/vir.0.82950-0.
16. Cao, W. *et al.* Identification of α-dystroglycan as a receptor for lymphocytic choriomeningitis virus and Lassa fever virus. *Science (80-.).* (1998) doi:10.1126/science.282.5396.2079.
17. Kunz, S., Rojek, J. M., Perez, M., Spiropoulou, C. F. & Oldstone, M. B. A. Characterization of the Interaction of Lassa Fever Virus with Its Cellular Receptor -Dystroglycan. *J. Virol.* (2005) doi:10.1128/jvi.79.10.5979-5987.2005.
18. Imperiali, M. *et al.* O Mannosylation of -Dystroglycan Is Essential for Lymphocytic Choriomeningitis Virus Receptor Function. *J. Virol.* (2005) doi:10.1128/jvi.79.22.14297-14308.2005.
19. Kunz, S. Receptor binding and cell entry of Old World arenaviruses reveal novel aspects of virus-host interaction. *Virology* (2009) doi:10.1016/j.virol.2009.02.042.
20. Rojek, J. M., Spiropoulou, C. F., Campbell, K. P. & Kunz, S. Old World and Clade C New World Arenaviruses Mimic the Molecular Mechanism of Receptor Recognition Used by -Dystroglycan's Host-Derived Ligands. *J. Virol.* (2007) doi:10.1128/jvi.02574-06.
21. Barresi, R. & Campbell, K. P. Dystroglycan: From biosynthesis to pathogenesis of human disease. *J. Cell Sci.* (2006) doi:10.1242/jcs.02814.
22. Baize, S. *et al.* Early and Strong Immune Responses Are Associated with Control of Viral Replication and Recovery in Lassa Virus-Infected Cynomolgus Monkeys. *J. Virol.* (2009) doi:10.1128/jvi.01948-08.
23. Carrion, R. *et al.* Lassa Virus Infection in Experimentally Infected Marmosets: Liver Pathology and Immunophenotypic Alterations in Target Tissues. *J. Virol.* (2007) doi:10.1128/jvi.02876-06.
24. Ter Meulen, J. *et al.* Old and New World arenaviruses share a highly conserved epitope in the fusion domain of the glycoprotein 2, which is recognized by Lassa virus-specific human CD4+ T-cell clones. *Virology* (2004) doi:10.1016/j.virol.2003.12.013.
25. Fisher-Hoch, S. P., Hutwagner, L., Brown, B. & McCormick, J. B. Effective Vaccine for Lassa Fever. *J. Virol.* (2000) doi:10.1128/jvi.74.15.6777-6783.2000.
26. Davies, M. N. *et al.* Using databases and data mining in vaccinology. *Expert Opin. Drug Discov.* **2**, 19–35 (2007).
27. Backert, L. & Kohlbacher, O. Immunoinformatics and epitope prediction in the age of genomic medicine. *Genome Medicine* (2015) doi:10.1186/s13073-015-0245-0.
28. Tomar, N. & De, R. K. Immunoinformatics: a brief review. *Methods Mol. Biol.* **1184**, 23–55 (2014).
29. Stanekov, Z. & Varekov, E. Conserved epitopes of influenza A virus inducing protective immunity and their prospects for universal vaccine development. *Virology Journal* (2010) doi:10.1186/1743-422X-7-351.
30. Sominskaya, I. *et al.* Construction and immunological evaluation of multivalent hepatitis B virus (HBV) core virus-like particles carrying HBV and HCV epitopes. *Clin. Vaccine Immunol.* (2010) doi:10.1128/CVI.00468-09.
31. He, L. *et al.* Approaching rational epitope vaccine design for hepatitis C virus with meta-server and multivalent scaffolding. *Sci. Rep.* (2015) doi:10.1038/srep12501.

32. E Silva, R. de F. *et al.* Combination of In Silico Methods in the Search for Potential CD4(+) and CD8(+) T Cell Epitopes in the Proteome of Leishmania braziliensis. *Front. Immunol.* **7**, 327 (2016).
33. Pahil, S., Taneja, N., Ansari, H. R. & Raghava, G. P. S. In silico analysis to identify vaccine candidates common to multiple serotypes of Shigella and evaluation of their immunogenicity. *PLoS One* (2017) doi:10.1371/journal.pone.0180505.
34. Khan, S. *et al.* Immunoinformatics and structural vaccinology driven prediction of multi-epitope vaccine against Mayaro virus and validation through in-silico expression. *Infect. Genet. Evol.* (2019) doi:10.1016/j.meegid.2019.06.006.
35. Brennick, C. A., George, M. M., Corwin, W. L., Srivastava, P. K. & Ebrahimi-Nik, H. Neoepitopes as cancer immunotherapy targets: Key challenges and opportunities. *Immunotherapy* **9**, 361–371 (2017).
36. Buonaguro, L. Developments in cancer vaccines for hepatocellular carcinoma. *Cancer Immunol. Immunother.* **65**, 93–99 (2016).
37. Kuo, T., Wang, C., Badakhshan, T., Chilukuri, S. & BenMohamed, L. The challenges and opportunities for the development of a T-cell epitope-based herpes simplex vaccine. *Vaccine* **32**, 6733–6745 (2014).
38. Lu, I. N., Farinelle, S., Sausy, A. & Muller, C. P. Identification of a CD4 T-cell epitope in the hemagglutinin stalk domain of pandemic H1N1 influenza virus and its antigen-driven TCR usage signature in BALB/c mice. *Cell. Mol. Immunol.* **14**, 511–520 (2017).
39. He, R. *et al.* Efficient control of chronic LCMV infection by a CD4 T cell epitope-based heterologous prime-boost vaccination in a murine model. *Cell. Mol. Immunol.* **15**, 815–826 (2018).
40. Jiang, P. *et al.* Evaluation of tandem Chlamydia trachomatis MOMP multi-epitopes vaccine in BALB/c mice model. *Vaccine* **35**, 3096–3103 (2017).
41. Lennerz, V. *et al.* Immunologic response to the survivin-derived multi-epitope vaccine EMD640744 in patients with advanced solid tumors. *Cancer Immunol. Immunother.* (2014) doi:10.1007/s00262-013-1516-5.
42. Lin, X. *et al.* Chimerically fused antigen rich of overlapped epitopes from latent membrane protein 2 (LMP2) of Epstein-Barr virus as a potential vaccine and diagnostic agent. *Cell. Mol. Immunol.* (2016) doi:10.1038/cmi.2015.29.
43. Saadi, M., Karkhah, A. & Nouri, H. R. Development of a multi-epitope peptide vaccine inducing robust T cell responses against brucellosis using immunoinformatics based approaches. *Infect. Genet. Evol.* (2017) doi:10.1016/j.meegid.2017.04.009.
44. Ong, E. *et al.* Vaxign-ML: Supervised machine learning reverse vaccinology model for improved prediction of bacterial protective antigens. *Bioinformatics* **36**, 3185–3191 (2020).
45. He, Y., Xiang, Z. & Mobley, H. L. T. Vaxign: The first web-based vaccine design program for reverse vaccinology and applications for vaccine development. *J. Biomed. Biotechnol.* **2010**, 297505 (2010).
46. Larsen, M. V. *et al.* Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics* (2007) doi:10.1186/1471-2105-8-424.
47. Doytchinova, I. A. & Flower, D. R. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* (2007) doi:10.1186/1471-2105-8-4.
48. Calis, J. J. A. *et al.* Properties of MHC Class I Presented Peptides That Enhance Immunogenicity. *PLoS Comput. Biol.* (2013) doi:10.1371/journal.pcbi.1003266.

49. Vita, R. *et al.* The immune epitope database (IEDB) 3.0. *Nucleic Acids Res.* (2015) doi:10.1093/nar/gku938.
50. Dhanda, S. K., Vir, P. & Raghava, G. P. S. Designing of interferon-gamma inducing MHC class-II binders. *Biol. Direct* (2013) doi:10.1186/1745-6150-8-30.
51. Green, D. S., Nunes, A. T., Annunziata, C. M. & Zoon, K. C. Monocyte and interferon based therapy for the treatment of ovarian cancer. *Cytokine and Growth Factor Reviews* (2016) doi:10.1016/j.cytogfr.2016.02.006.
52. James, C. M. *et al.* Differential activities of alpha/beta IFN subtypes against influenza virus in vivo and enhancement of specific immune responses in DNA vaccinated mice expressing haemagglutinin and nucleoprotein. *Vaccine* **25**, 1856–1867 (2007).
53. Jiang, W., Ren, L. & Jin, N. HIV-1 DNA vaccine efficacy is enhanced by coadministration with plasmid encoding IFN-α. *J. Virol. Methods* (2007) doi:10.1016/j.jviromet.2007.07.016.
54. Saha, S. & Raghava, G. P. S. AlgPred: Prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res.* (2006) doi:10.1093/nar/gkl343.
55. Gupta, S. *et al.* In Silico Approach for Predicting Toxicity of Peptides and Proteins. *PLoS One* (2013) doi:10.1371/journal.pone.0073957.
56. Saha, S. & Raghava, G. P. S. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins Struct. Funct. Genet.* (2006) doi:10.1002/prot.21078.
57. Ponomarenko, J. *et al.* ElliPro: A new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics* (2008) doi:10.1186/1471-2105-9-514.
58. Sabourin, M., Tuzon, C. T., Fisher, T. S. & Zakian, V. A. A flexible protein linker improves the function of epitope-tagged proteins in Saccharomyces cerevisiae. *Yeast* **24**, 39–45 (2007).
59. Lee, S. J. *et al.* A potential protein adjuvant derived from Mycobacterium tuberculosis Rv0652 enhances dendritic cells-based tumor immunotherapy. *PLoS One* (2014) doi:10.1371/journal.pone.0104351.
60. Gasteiger, E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server. in *The Proteomics Protocols Handbook* (2005). doi:10.1385/1-59259-890-0:571.
61. Darvishi, F., Zarei, A. & Madzak, C. In silico and in vivo analysis of signal peptides effect on recombinant glucose oxidase production in nonconventional yeast Yarrowia lipolytica. *World J. Microbiol. Biotechnol.* **34**, 128 (2018).
62. Magnan, C. N., Randall, A. & Baldi, P. SOLpro: Accurate sequence-based prediction of protein solubility. *Bioinformatics* (2009) doi:10.1093/bioinformatics/btp386.
63. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* (2000) doi:10.1093/bioinformatics/16.4.404.
64. Khatoon, N., Pandey, R. K. & Prajapati, V. K. Exploring Leishmania secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach. *Sci. Rep.* (2017) doi:10.1038/s41598-017-08842-w.
65. Källberg, M. *et al.* Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* **7**, 1511–1522 (2012).
66. Ko, J., Park, H., Heo, L. & Seok, C. GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Res.* (2012) doi:10.1093/nar/gks493.

67. Lee, G. R., Won, J., Heo, L. & Seok, C. GalaxyRefine2: Simultaneous refinement of inaccurate local regions and overall protein structure. *Nucleic Acids Res.* **47**, W451–W455 (2019).
68. Wiederstein, M. & Sippl, M. J. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* (2007) doi:10.1093/nar/gkm290.
69. Colovos, C. & Yeates, T. O. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Sci.* (1993) doi:10.1002/pro.5560020916.
70. Lovell, S. C. *et al.* Structure validation by Calpha geometry: Phi,psi and Cbeta deviation. *Proteins* (2003).
71. Shey, R. A. *et al.* In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Sci. Rep.* (2019) doi:10.1038/s41598-019-40833-x.
72. Lengauer, T. & Rarey, M. Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.* (1996) doi:10.1016/S0959-440X(96)80061-3.
73. Meyer, B. & Ly, H. Inhibition of Innate Immune Responses Is Key to Pathogenesis by Arenaviruses. *J. Virol.* (2016) doi:10.1128/jvi.03049-15.
74. Loo, Y. M. & Gale, M. Immune Signaling by RIG-I-like Receptors. *Immunity* (2011) doi:10.1016/j.immuni.2011.05.003.
75. Cui, S. *et al.* The C-Terminal Regulatory Domain Is the RNA 5'-Triphosphate Sensor of RIG-I. *Mol. Cell* **29**, 169–179 (2008).
76. Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: Computed atlas of surface topography of proteins. *Nucleic Acids Res.* (2018) doi:10.1093/nar/gky473.
77. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Res.* (2005) doi:10.1093/nar/gki481.
78. Mashiach, E., Schneidman-Duhovny, D., Andrusier, N., Nussinov, R. & Wolfson, H. J. FireDock: a web server for fast interaction refinement in molecular docking. *Nucleic Acids Res.* (2008) doi:10.1093/nar/gkn186.
79. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**, 774–797 (2007).
80. Evgeny, K. Crystal contacts as nature's docking solutions. *J. Comput. Chem.* **31**, 133–143 (2010).
81. Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
82. van Gunsteren, W. F. & Berendsen, H. J. C. Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. *Angew. Chemie Int. Ed. English* **29**, 992–1023 (1990).
83. Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ1 and χ2 Dihedral Angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
84. Van Der Spoel, D., Van Maaren, P. J. & Berendsen, H. J. C. A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field. *J. Chem. Phys.* **108**, 10220–10230 (1998).
85. Grote, A. *et al.* JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* (2005) doi:10.1093/nar/gki376.

86. Morla, S., Makhija, A. & Kumar, S. Synonymous codon usage pattern in glycoprotein gene of rabies virus. *Gene* (2016) doi:10.1016/j.gene.2016.02.047.
87. Ali, M. *et al.* Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Sci. Rep.* (2017) doi:10.1038/s41598-017-09199-w.
88. Rapin, N., Lund, O., Bernaschi, M. & Castiglione, F. Computational immunology meets bioinformatics: The use of prediction tools for molecular binding in the simulation of the immune system. *PLoS One* (2010) doi:10.1371/journal.pone.0009862.
89. Castiglione, F., Mantile, F., De Berardinis, P. & Prisco, A. How the interval between prime and boost injection affects the immune response in a computational model of the immune system. *Comput. Math. Methods Med.* (2012) doi:10.1155/2012/842329.
90. Corleis, B. *et al.* Early type i Interferon response induces upregulation of human β-defensin 1 during acute HIV-1 infection. *PLoS One* (2017) doi:10.1371/journal.pone.0173161.
91. Ribet, D. & Cossart, P. How bacterial pathogens colonize their hosts and invade deeper tissues. *Microbes Infect.* **17**, 173–183 (2015).
92. VHFC. Lassa | VHFC. https://vhfc.org/diseases/lassa/ (2020).
93. Eberhardt, K. A. *et al.* Ribavirin for the treatment of Lassa fever: A systematic review and meta-analysis. *Int. J. Infect. Dis.* **87**, 15–20 (2019).
94. Günther, S. & Lenz, O. Lassa virus. *Critical Reviews in Clinical Laboratory Sciences* (2004) doi:10.1080/10408360490497456.
95. Ilori, E. A. *et al.* Epidemiologic and clinical features of lassa fever outbreak in Nigeria, january 1-may 6, 2018. *Emerg. Infect. Dis.* **25**, 1066–1074 (2019).
96. Borio, L. *et al.* Hemorrhagic fever viruses as biological weapons: Medical and public health management. *Journal of the American Medical Association* (2002) doi:10.1001/jama.287.18.2391.
97. The Lancet Infectious Diseases. Lassa fever and global health security. *The Lancet Infectious Diseases* (2018) doi:10.1016/S1473-3099(18)30179-8.
98. Siddle, K. J. *et al.* Genomic Analysis of Lassa Virus during an Increase in Cases in Nigeria in 2018. *N. Engl. J. Med.* **379**, 1745–1753 (2018).
99. Bredenbeek, P. J. *et al.* A recombinant Yellow Fever 17D vaccine expressing Lassa virus glycoproteins. *Virology* (2006) doi:10.1016/j.virol.2005.12.001.
100. Lukashevich, I. S. & Pushko, P. Vaccine platforms to control Lassa fever. *Expert Review of Vaccines* (2016) doi:10.1080/14760584.2016.1184575.
101. Lukashevich, I. S. Advanced vaccine candidates for Lassa Fever. *Viruses* (2012) doi:10.3390/v4112514.
102. Marzi, A., Feldmann, F., Geisbert, T. W., Feldmann, H. & Safronetz, D. Vesicular stomatitis virus–based vaccines against lassa and ebola viruses. *Emerg. Infect. Dis.* (2015) doi:10.3201/eid2102.141649.
103. Vartak, A. & Sucheck, S. J. Recent advances in subunit vaccine carriers. *Vaccines* (2016) doi:10.3390/vaccines4020012.
104. Moise, L. *et al.* Ivax: An integrated toolkit for the selection and optimization of antigens and the design of epitope-driven vaccines. *Hum. Vaccines Immunother.* (2015) doi:10.1080/21645515.2015.1061159.

105. Zhou, W.-Y. *et al.* Therapeutic efficacy of a multi-epitope vaccine against Helicobacter pylori infection in BALB/c mice model. *Vaccine* **27**, 5013–5019 (2009).
106. Günther, S. *et al.* Antibodies to Lassa virus Z protein and nucleoprotein co-occur in human sera from Lassa fever endemic regions. *Med. Microbiol. Immunol.* (2001) doi:10.1007/s004300100061.
107. Prescott, J. B. *et al.* Immunobiology of Ebola and Lassa virus infections. *Nature Reviews Immunology* (2017) doi:10.1038/nri.2016.138.
108. Russier, M., Pannetier, D. & Baize, S. Immune responses and Lassa virus infection. *Viruses* (2012) doi:10.3390/v4112766.
109. Meza, B., Ascencio, F., Sierra-Beltrán, A. P., Torres, J. & Angulo, C. A novel design of a multi-antigenic, multistage and multi-epitope vaccine against Helicobacter pylori: An in silico approach. *Infect. Genet. Evol.* (2017) doi:10.1016/j.meegid.2017.02.007.
110. Arai, R., Ueda, H., Kitayama, A., Kamiya, N. & Nagamune, T. Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Eng.* (2001) doi:10.1093/protein/14.8.529.
111. Lee, S. & Nguyen, M. T. Recent Advances of Vaccine Adjuvants for Infectious Diseases. *Immune Netw.* (2015) doi:10.4110/in.2015.15.2.51.
112. Singh, A., Thakur, M., Sharma, L. K. & Chandra, K. Designing a multi-epitope peptide based vaccine against SARS-CoV-2. *Sci. Rep.* **10**, (2020).
113. Corradin, G., Villard, V. & Kajava, A. V. Protein structure based strategies for antigen discovery and vaccine development against malaria and other pathogens. *Endocr. Metab. Immune Disord. Drug Targets* **7**, 259–265 (2007).
114. Wlodawer, A. Stereochemistry and validation of macromolecular structures. *Methods Mol. Biol.* **1607**, 595–610 (2017).
115. Gori, A., Longhi, R., Peri, C. & Colombo, G. Peptides for immunological purposes: Design, strategies and applications. *Amino Acids* (2013) doi:10.1007/s00726-013-1526-9.
116. Chen, R. Bacterial expression systems for recombinant protein production: E. coli and beyond. *Biotechnol. Adv.* (2012) doi:10.1016/j.biotechadv.2011.09.013.
117. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: Advances and challenges. *Frontiers in Microbiology* (2014) doi:10.3389/fmicb.2014.00172.
118. Chaplin, D. D. Overview of the immune response. *J. Allergy Clin. Immunol.* **125**, S3–S23 (2010).
119. Kaplinsky, J. & Arnaout, R. Robust estimates of overall immune-repertoire diversity from high-throughput measurements on samples. *Nat. Commun.* **7**, 11881 (2016).

Figures

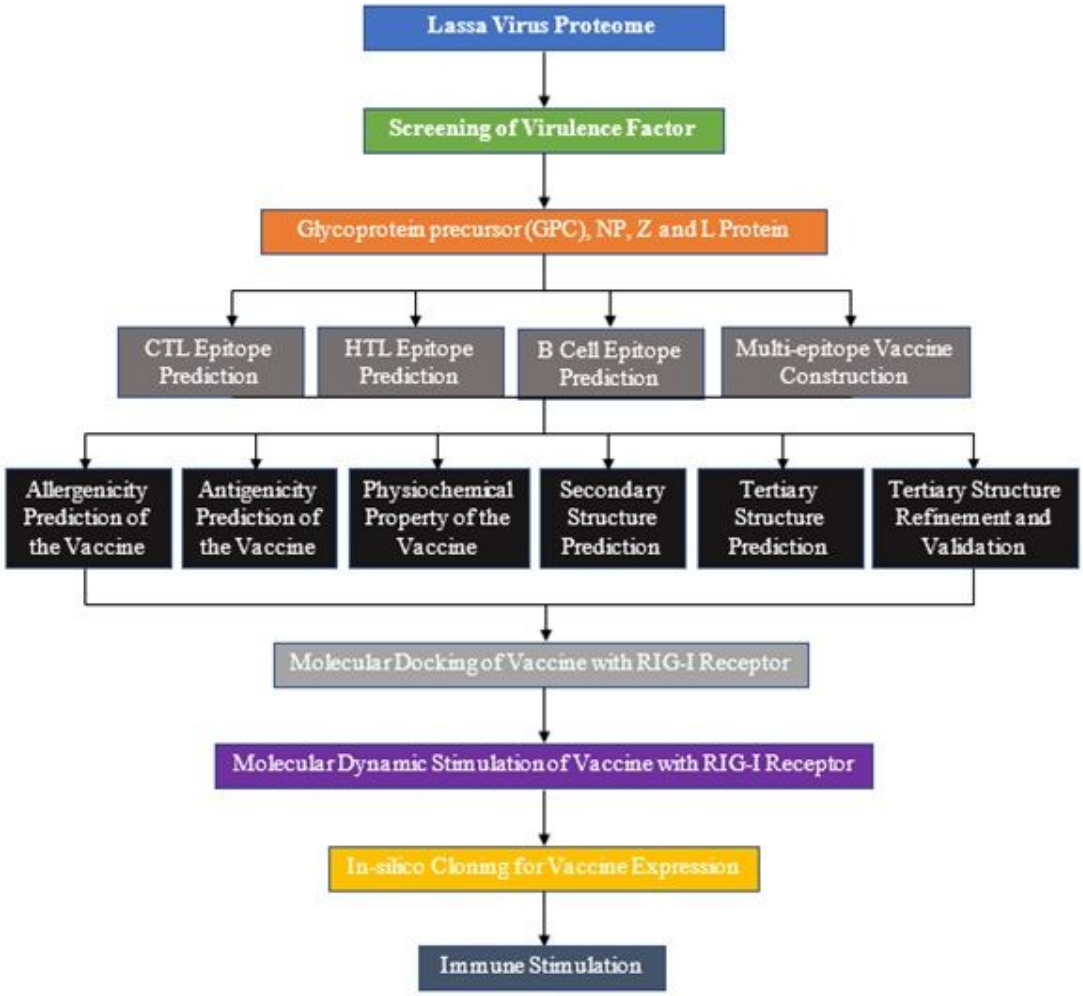

Figure 1

Presentation of the overall flow. The methodological strategy was divided into six parts: (i) Lassa virus proteome retrieval, (ii) prediction of epitopes, (iii) multi-epitope vaccine construction, (iv) vaccine modelling, (v) molecular docking and dynamic simulation, (vi) in-silico expression and immune stimulation.

Figure 2

Structure of the final multi-epitope vaccine construct. CTL and HTL epitopes along with linkers are depicted.

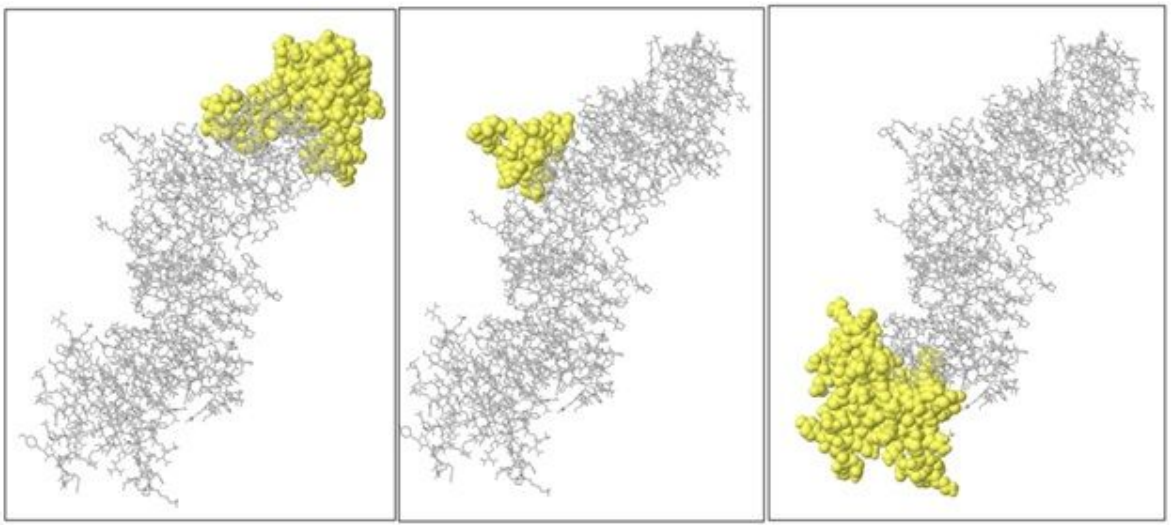

Figure 3

Predicted conformational B-cell epitopes are portrayed as coloured spheres in the final vaccine structure.

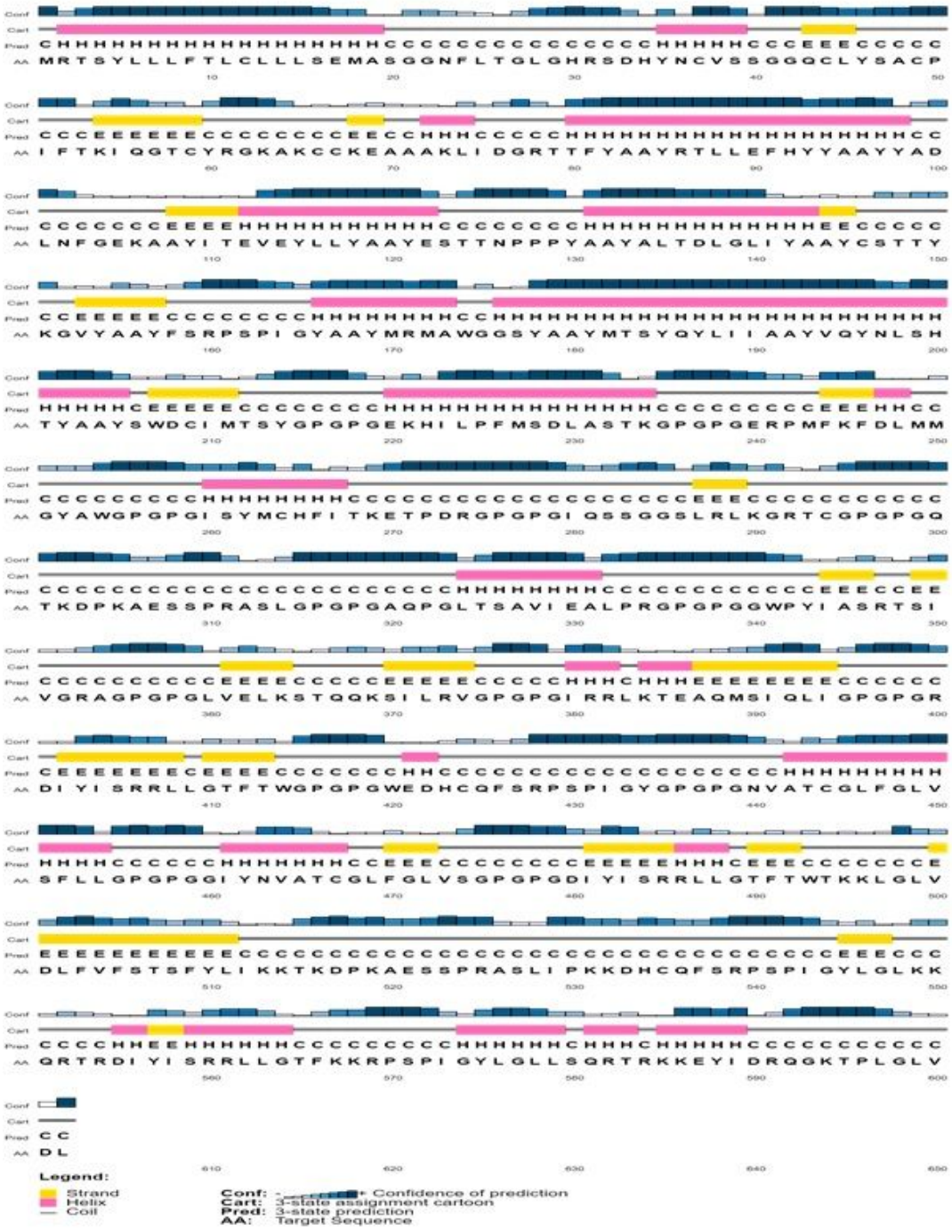

Figure 4

Graphical representation of secondary structure prediction of the multi-epitope vaccine. The protein is predicted to have an alpha helix (34.4%), a beta-strand (15.8%), and a coil (49.8%).

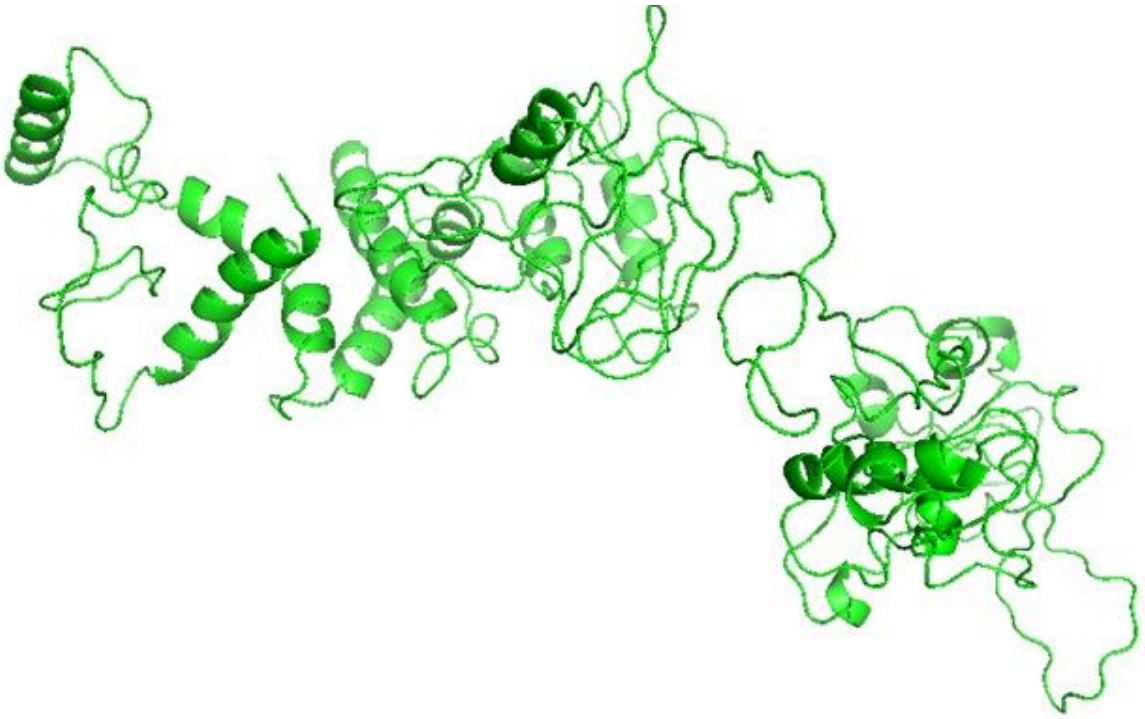

Figure 5

3D structure final subunit vaccine model.

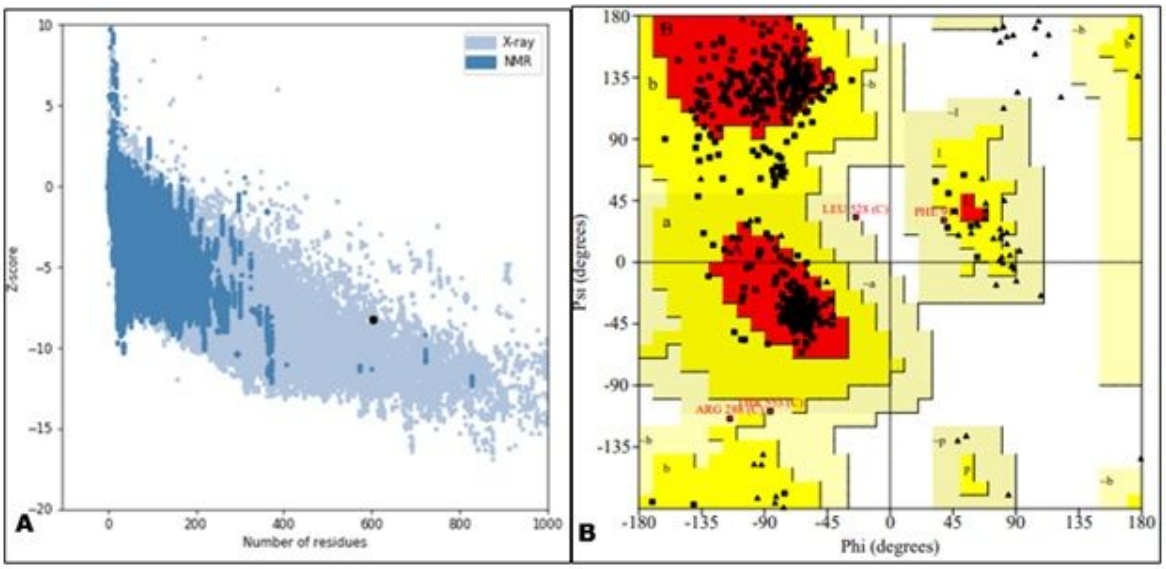

Figure 6

(A) Validation of refined model with ProSA-web, giving a Z-score of -8.27, (B) Ramachandran plot analysis of the refined model protein revealed that 86.8% of residues in the protein are in most favoured regions and 12.4% in additional allowed regions, 0.6% of the residues in generously allowed regions and only 0.2% in disallowed region.

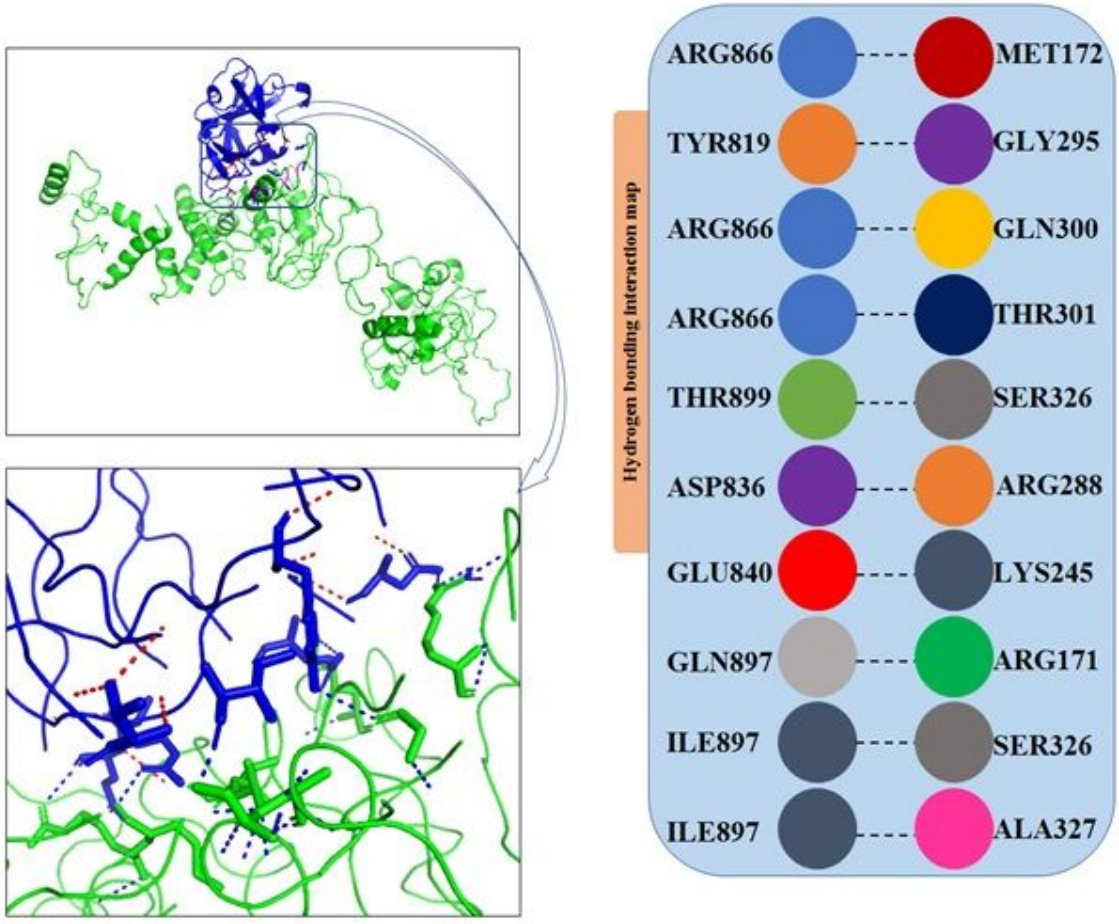

Figure 7

RIG-I (PDB ID: 2QFB)-vaccine docked complex and hydrogen bonding interaction map. The RIG-I (receptor) is shown in blue, while the green colour represents the multi-epitope subunit vaccine (left figure). RIG-I receptor on the left and final vaccine construct on the right (right figure: hydrogen bond map).

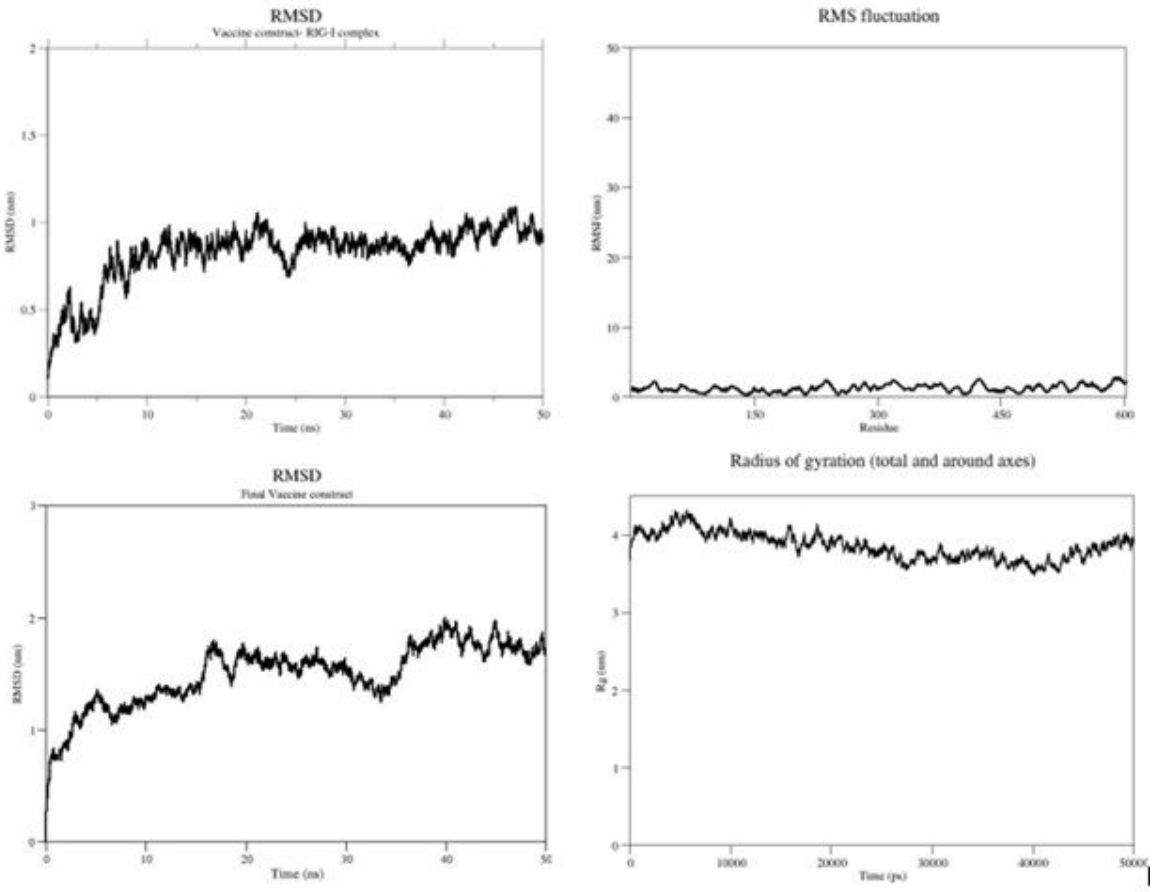

Figure 8

Molecular dynamics simulation of the RIG-I-multi-epitope subunit vaccine complex with RMSD of the final vaccine construct and the complex, RMSF and radius of gyration of the final vaccine construct.

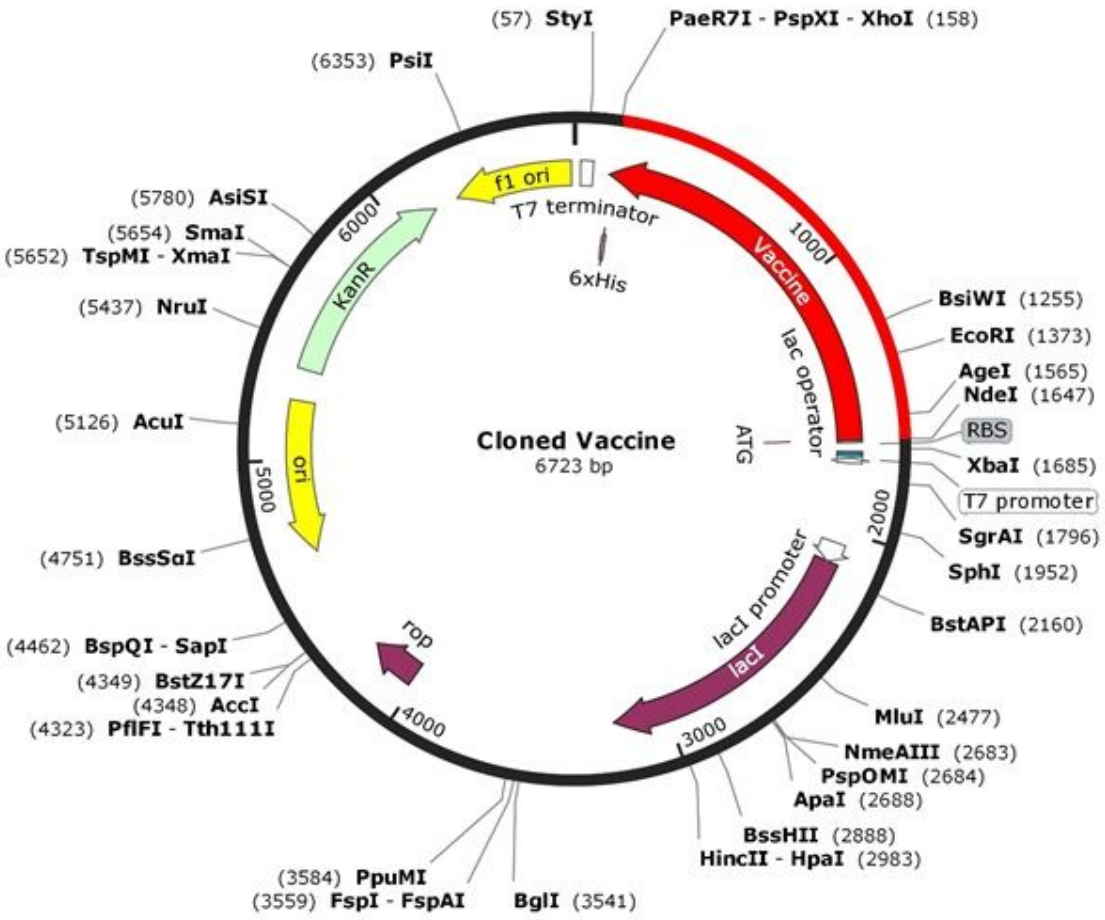

Figure 9

In silico restriction cloning of the final vaccine sequence into the pET30a (+) expression vector. The red part represents the gene coding for the vaccine and the black circle represents the vector backbone.

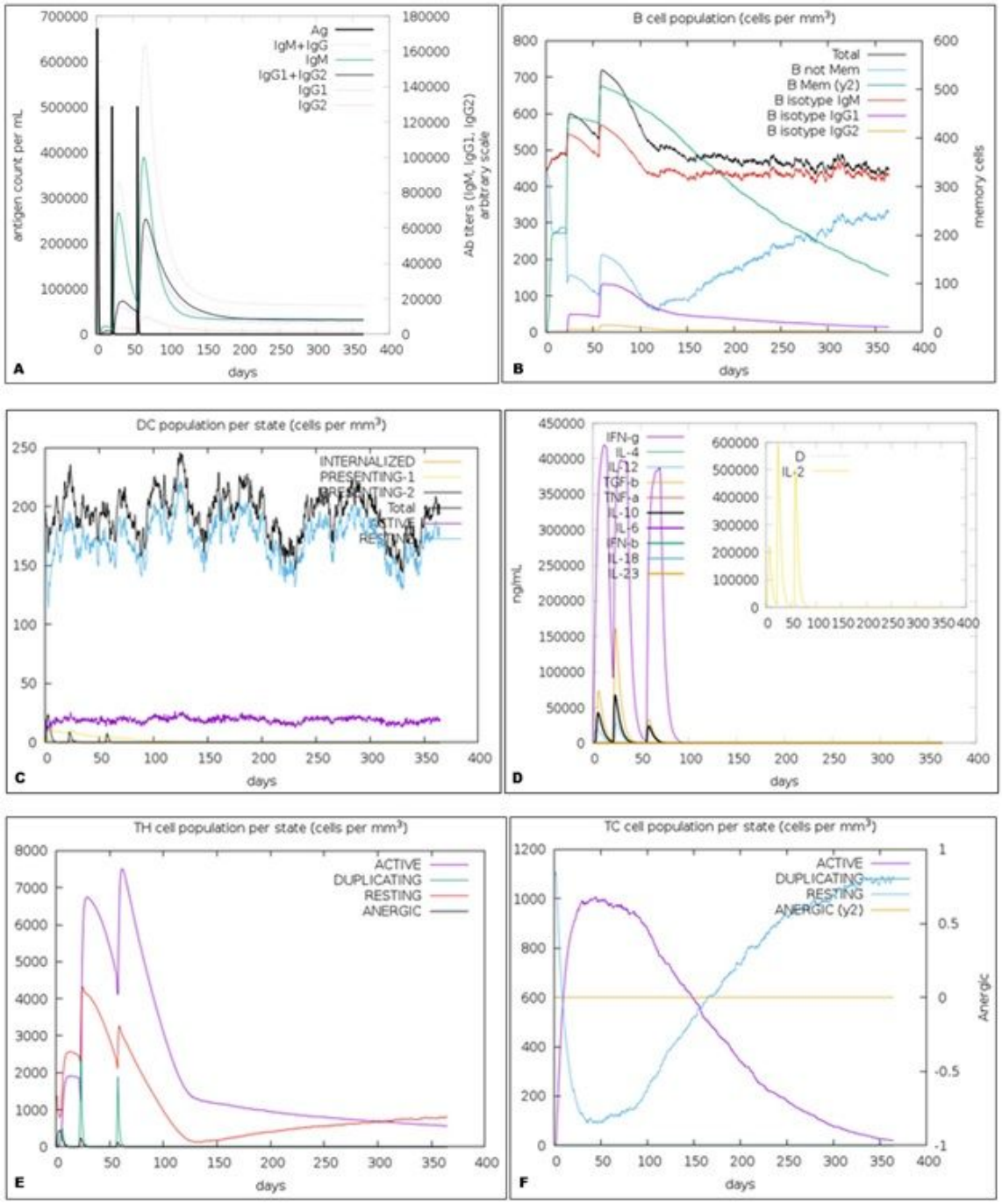

Figure 10

C-ImmSim presentation of an in silico simulation of immune response using the vaccine construct peptide as antigen. A: Immunoglobulin production in response to antigen injections (black vertical lines); specific subclasses are indicated as coloured peaks. B: The evolution of B-cell populations after three injections. C: Dendritic cell population per state, D: production of cytokine and interleukins with Simpson index (D), E: The evolution of T-helper, and F: T-cytotoxic cell populations per state after the injections. The resting state represents cells not presented with the antigen while the anergic state represents tolerance of the T-cells to the antigen due to repeated exposures.