Identifying immunodominant multi-epitopes from the envelope glycoprotein of the Lassa mammarenavirus as vaccine candidate for Lassa fever

Purpose: Lassa fever is a zoonotic acute viral hemorrhagic disease caused by Lassa virus (LASV). There is currently no licensed vaccine for the prevention of the disease. This study is aimed at discovering immunodominant epitopes from the envelope glycoprotein of the Lassa mammarenavirus and designing of a multi-epitope vaccine candidate (VC).

Materials and Methods: The amino acid sequences of the envelope glycoprotein of 26 strains of LASV from five countries were selected. After evaluation for antigenicity, immunogenicity, allergenicity, and toxicity, immunodominant CD8, CD4, and linear B lymphocytes were also selected. The selected epitopes were modelled and a molecular docking with the appropriate major histocompatibility complex (MHC) proteins was performed. Using an adjuvant and linkers, a multi-epitope VC was designed. The VC was evaluated for its physicochemical and immunological properties and structurally refined, validated, and mutated (disulphide engineering). The complex formed by the VC and the toll-like receptor-4 receptor was subjected to molecular dynamic simulation (MDS) followed by in silico cloning in a plasmid vector.

Results: A VC with 203 sequences, 22.13 kDa weight, isoelectric point of 9.85 (basic), instability index value of 27.62, aliphatic index of 68.87, and GRAVY value of -0.455 (hydrophilic) emerged. The VC is predicted to be non-allergenic with antigenicity, MHC I immunogenicity, and solubility upon overexpression values of 0.81, 2.04, and 0.86 respectively. The VC also has an estimated half-life greater than 10 hours in Escherichia coli and showed stability in all the parameters of MDS.

Conclusion: The VC shows good promise in the prevention of Lassa fever but further tests are required to validate its safety and efficacy.

Keywords: Lassa virus, Envelope protein, Major histocompatibility complex, Epitopes, Vaccine candidate

Introduction

Lassa fever (LF) is a zoonotic acute viral hemorrhagic illness caused by a member of the Arenaviridae family of viruses, called the Lassa virus (LASV) [1]. LF is transmitted via ingestion or contact with food, water, and items contaminated with the urine or feces of sick multimammate rat (Mastomys natalensis) and has the ability to kill tens of thousands of people. Like SARS-CoV-2, the LASV can be found in human body fluids including semen, even after recovery [1,2].
Named after a village called Lassa in Borno State, Nigeria, LF was first discovered in 1950s in Sierra Leone but was only linked with its etiological agent, LASV in 1969 when it killed two missionary nurses. LF is endemic across West Africa but prevalent in Nigeria, Benin, Ghana, Guinea, Togo, Sierra Leone, Mali, and Liberia [3].

Usually asymptomatic, LF develops between 6–12 days. However, when symptoms do manifest, they do so gradually from the onset and include fever, headaches, general weakness, nausea, vomiting, and chest and muscle pain. Swelling of the face, fluid in the lungs, bleeding from the mouth, nose, vaginal or gastrointestinal tract, and low blood pressure can all occur in severe cases. Hearing loss is also a symptom found in survivors, which improves with time [4]. LF is critical in late pregnancies, resulting in maternal death and/or fetal loss in as much as 80% of cases in the third trimester [5].

LF is often misdiagnosed as its symptoms are very similar to other diseases like Ebola, malaria, typhoid, and yellow fever [3]. However, confirmatory diagnosis of LF include laboratory tests such as reverse transcriptase polymerase chain reaction, an antigen detection test, virus isolation cell culture, and antibody enzyme-linked immunosorbent assay [3]. The clinical management of LF involves the isolation of affected persons, the administration of the antiviral agent, Ribavirin, and other supportive therapies [3]. There is currently no licensed vaccine for LF; while numerous candidates are in different stages of development and these include live attenuated, DNA, and RNA vaccines and a variety of viral vectored vaccine approaches [6].

This study is aimed at identifying immunodominant multi-epitopes from the envelope glycoprotein of the Lassa mammarenavirus for the design of a vaccine candidate (VC) against the disease.

Materials and Methods

Retrieval of Protein sequences

The amino acid sequences of glycoprotein of several strains of the Lassa mammarenavirus were obtained in FASTA format from the NCBI database (https://www.ncbi.nlm.nih.gov/protein/?term=glycoprotein+Lassa+fever+virus). From five nations in Africa, sequences from 26 strains were chosen. Their accession numbers, country of origin, and year of collection were recorded.

Analyses of sequences

The CLUSTA OMEGA webserver (https://www.ebi.ac.uk/Tools/msa/clustalo/) was used for the phylogenetic analyses of all the 26 retrieved sequences. To get multiple sequence alignments, the phylogenetic tree (neighbor-joining tree without distance corrections), and the percent identity matrix, the output format was set to ClustalW with character counts. The repeat sequences with 100% identity were screened out after a pairwise comparison of the sequences.

Selection of antigenic peptides

The identified protein sequences were entered into the Vaxijen ver. 2.0 webserver (http://www.ddg-pharmfac.net/vaxijen/Vaxijen/Vaxijen.html) in FASTA format for Overall Prediction of the Protective Antigen (OPPA). The target organism was chosen as the virus, and the threshold was set at 0.4.

CTL epitopes: prediction and evaluation

Using the Immune Epitope Database (IEDB) Analysis Resource for major histocompatibility complex (MHC)-I binding predictions (http://tools.iedb.org/mhci/), the binding of peptides to MHC-I alleles was predicted. The prediction method was set at IEDB recommended 2020.09 (NetMHCpan EL 4.1). The sequences of the selected strains were entered into the server in FASTA format for the identification of high-binding epitopes (9 to 10 peptides each) using 54 distinct variations of the human leukocyte antigens (HLA). The HLA include: “HLA-A*01:01, HLA-A*01:01, HLA-A*02:01, HLA-A*02:02, HLA-A*02:03, HLA-A*02:03, HLA-A*02:06, HLA-A*02:06, HLA-A*03:01, HLA-A*03:01, HLA-A*11:01, HLA-A*11:01, HLA-A*23:01, HLA-A*23:01, HLA-A*24:02, HLA-A*24:02, HLA-A*26:01, HLA-A*26:01, HLA-A*30:01, HLA-A*30:01, HLA-A*30:02, HLA-A*30:02, HLA-A*31:01, HLA-A*31:01, HLA-A*32:01, HLA-A*32:01, HLA-A*33:01, HLA-A*33:01, HLA-A*68:01, HLA-A*68:01, HLA-A*68:01, HLA-A*68:02, HLA-A*68:02, HLA-B*07:02, HLA-B*07:02, HLA-B*08:01, HLA-B*08:01, HLA-B*15:01, HLA-B*15:01, HLA-B*35:01, HLA-B*35:01, HLA-B*40:01, HLA-B*40:01, HLA-B*44:02, HLA-B*44:02, HLA-B*44:02, HLA-B*44:03, HLA-B*44:03, HLA-B*51:01, HLA-B*51:01, HLA-B*53:01, HLA-B*53:01, HLA-B*57:01, HLA-B*57:01, HLA-B*58:01, and HLA-B*58:01.” A percentile rank of 0.05 was set as a criterion for strong binders. The antigenicity of selected MHC I epitopes was predicted using the Vaxijen ver. 2.0 webserver with 0.4 as the threshold. The IEDB Analysis Resource (http://tools.iedb.org/immunogenicity/) was used to test for MHC class I immunogenicity using the default set-
ttings for 9- and 10-mer epitopes. Epitopes having immunogenicity score greater than 0.00 were chosen. To perform allergenicity prediction, the AllerTop ver. 2.0 webserver (https://www.ddg-pharmfac.net/AllerTOP/) was utilized, while the ToxinPred webserver (http://crdd.osdd.net/raghava/toxin-pred/) was used to perform toxicity prediction of the selected cytotoxic T lymphocyte (CTL) epitopes. The top four CTL epitopes that are most immunogenic, non-allergic, and non-toxic were selected.

**HTL epitopes: prediction and evaluation**

The MHC-II binding prediction tool from the IEDB Analysis Resource (http://tools.iedb.org/mhcii/) was used to predict helper T lymphocyte (HTL) epitope binding to MHC II alleles. IEDB recommended 2.22 was the prediction method chosen, the whole HLA reference set was selected, and the length of the peptide was set to 15. The MHC II alleles include: *HLA-DRB1*01:01, HLA-DRB1*03:01, HLA-DRB1*04:01, HLA-DRB1*04:05, HLA-DRB1*07:01, HLA-DRB1*08:02, HLA-DRB1*09:01, HLA-DRB1*11:01, HLA-DRB1*12:01, HLA-DRB1*13:02, HLA-DRB1*15:01, HLA-DRB3*01:01, HLA-DRB3*02:02, HLA-DRB4*01:01, HLA-DRB5*01:01, HLA-DQA1*05:01/DQB1*02:01, HLA-DQA1*05:01/DQB1*03:01, HLA-DQA1*03:01/DQB1*03:02, HLA-DQA1*04:01/DQB1*04:02, HLA-DQA1*01:01/DQB1*03:01, HLA-DQA1*01:02/DQB1*06:02, HLA-DPA1*02:01/DPB1*01:01, HLA-DPA1*01:03/DPB1*01:01, HLA-DPA1*03:01/DPB1*04:01, HLA-DPA1*03:01/DPB1*04:02, HLA-DPA1*02:01/DPB1*05:01, and HLA-DPA1*02:01/DPB1*14:01." For good binders, the peptides were sorted by adjusted rank with a cut-off of 0.3. The HTL epitopes were evaluated for antigenicity using the Vaxijen ver. 2.0 server. The IEDB CD4 T cell Immunogenicity prediction tool (http://tools.iedb.org/CD4episcore/) was used to evaluate the immunogenic properties of the selected HTL epitopes. The prediction method was set at IEDB recommended (combined) and peptides were sorted by score/percentile rank. “Show all peptides” was set as maximum combined score threshold. The AllerTop ver. 2.0 and ToxinPred webservers, respectively. The top four most immunodominant, non-allergic, non-toxic, IFN-gamma-inducing, IL-4-inducing, and IL10-inducing HTL epitopes were selected.

**LBL epitopes: prediction and evaluation**

The IEDB Analysis Resource server for B cell epitope prediction (http://tools.iedb.org/main/bcell/) was used. The threshold was 0.5, and the prediction method was Bepipred Linear Epitope Prediction (BepiPred)-2.0 (https://services.healthtech.dtu.dk/service.php?BepiPred-2.0). Selected linear B lymphocyte (LBL) epitopes were further screened for allergenicity and toxicity using the AllerTop ver. 2.0 and ToxinPred webservers, respectively. The four LBL epitopes that are most immunodominant, non-toxic, and non-allergenic were chosen.

**Peptide modeling and molecular docking**

The three-dimensional (3D) structures of the CTL, HTL, and LBL epitopes were predicted using the PEP-FOLD ver. 3.5 server (https://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms:PEP-FOLD3) for de novo peptide structure prediction. The settings used are Run label: PEPFOLD; number of simulations: 100; sort models by: sOPEP, and input style: cartoon model. The model 1 of each of the predicted structures was downloaded in their Protein Data Bank (PDB) formats. The HLA-DRB1 (PDB ID: 6BIY) and HLA-A*3001 (PDB ID: 6IW) crystal structures were obtained from the Protein Data Bank (https://www.rcsb.org). UCSF Chimera (https://www.cgl.ucsf.edu/chimera/), a visualization tool, was used to liberate the proteins from water molecules and their co-crystallized ligands. To calculate the binding affinities between the modeled epitopes and their respective MHC proteins, the HawkDock Server (http://cadd.zju.edu.cn/hawkdock/) was employed. After docking the protein and peptide, the model with the lowest score (model 1) was chosen, and a Molecular Mechanics with Generalized Born and Surface Area Solvation analysis was carried out.

**Construction of the candidate for multiepitope vaccine**

Using the CTL, HTL, and LBL epitopes, a VC was constructed with the proper adjuvant and linker sequences. Several synthetic toll-like receptor-4 (TLR-4) agonist peptides [7] were evaluated using the VaxinPAD webservice (https://webs.iiitd.edu.in/raghava/vaxinpad/) to find an adjuvant. The support vector machine threshold was set to 0.0 on VaxinPAD webservice, and the dipeptide composition approach was chosen. The adjuvant was chosen based on the highest scoring peptide. The 7-mer peptide, HELSVLL was selected as adjuvant.
The EAAAK linker was placed at the adjuvant’s N and C termini for effective separation [8]. To allow for proper epitope presentation, Ala-Ala-Tyr (AAY) linker was placed between the CTL epitopes while the GlyPro-Gly-Pro-Gly (GPGPG) linkers was placed between the HTL and LBL epitopes [9-12].

Physicochemical and immunological properties of vaccine candidate

The ProtParam tool of the ExPasy server (https://web.expasy.org/protparam/) was used to determine the VC’s molecular weight, GRAVY, isoelectric point, and other physicochemical properties. The VC’s immunological profile which includes, MHC class I immunogenicity, antigenicity, allergenicity, toxicity, and solubility upon overexpression in Escherichia coli were predicted using the IEDB (http://tools.iedb.org/immunogenicity/), Vaxijen ver. 2.0 (http://www.ddg-pharmfac.net/vaxijen/Vaxijen.html), AllerTop ver. 2.0 (https://www.ddg-pharmfac.net/AllerTOP/), ToxinPred (https://webs.iitiid.edu.in/raghava/toxinpred/algoph.php), and SOLpro in the SCRATCH (http://scratch.proteomics.ics.uci.edu/explanation.html#SOLpro) protein predictor webservers, respectively.

Prediction of the vaccine candidate’s tertiary structures

The RaptorX server (http://raptorx.uchicago.edu/) was used to predict the VC’s 3D structure. This server employs a deep learning-powered distance-based protein folding algorithm. Based on the distance deviation of residues, model 1 of the VC was selected and downloaded in the PDB format. UCSF Chimera (https://www.cgl.ucsf.edu/chimera/) was used for visualization.

Refinement and validation of vaccine candidate’s 3D structure

The Galaxy-Refine webserver (http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE), which use the CASP10 algorithm for side chain repacking and overall structural relaxation was used to refine the 3D structure of the VC. The appropriate refined model of the VC was selected based on the mild perturbations to the side chain clusters [13]. UCSF Chimera was used for visualization. To validate the structure of the VC, the Ramachandran analysis, and the analysis of nonbonded atom–atom interactions were performed using the Molprobity (http://molprobity.biochem.duke.edu/) and ERRAT (https://servicesn.mbi.ucla.edu/Verify3D/) servers, respectively.

Disulfide engineering of vaccine candidate

Disulphide bonds were added to the VC to increase molecular stability. The disulphide engineering was done using the Disulphide by Design 2 webserver (http://cptweb.cpt.wayne.edu/DbD2/). After uploading the refined structure of the VC for possible residue pair for disulphide engineering, the selected options were intra chain, inter-chain, and build Cβ for glycine. The mutant structure was obtained in PDB format after all potential residue pairs were chosen for mutation.

Molecular docking of the TLR4 receptor and the vaccine candidate

The protein databank was searched for the TLR4 receptor (PDB ID: 4G8A). PyMol software was used to remove the water molecules and co-crystallized ligand [14]. The HDOCK server (http://hdock.phys.hust.edu.cn) was used to perform molecular docking to show the binding affinities between the refined and mutated 3D structure of the VC and the TLR4 receptor. The docking scores of the top 10 models were obtained. Model 1 was selected because it had the lowest docking score and the least ligand root-mean-square deviation (RMSD) value [15].

Molecular dynamics simulation studies

The iMODS server (http://imods.chaconlab.org/) was used to run a molecular dynamics simulation (MDS) of the VC and TLR4 receptor complex. This server performs normal mode analysis by calculating aggregate motions in the internal coordinates of the vaccine–receptor complex. The elastic network model, covariance, variance, eigenvalues, deformability, and B-factors were all calculated.

Codon adaptation and in silico cloning

The codon optimization was carried out using the Java Codon Adaptation Tool server (http://www.jcat.de/) and E. coli K12 was chosen as the host organism. Important E. coli sequences are avoided, such as prokaryotic ribosome binding sites, restriction enzyme cleavage sites, and intrinsic transcription termination sites. The guanine–cytosine (GC) content and the codon adaptation index (CAI) value were used to evaluate the codon-adapted sequences [16]. The original sequences of the vaccine construct and the translated sequences were compared using the NCBI BLAST server (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The adapted nucleotide sequences were cloned into the E. coli DH5α expression vector using the SnapGene ver. 6.0.2 software (SnapGene, San Diego, CA, USA) [17].
Results

Retrieval of protein sequences
Sequences from 26 different strains were selected from five West African countries (Table 1). Specifically, 16 strains were obtained from Guinea, six strains were obtained from Nigeria, two strains from Sierra Leone, and one strain each from Liberia and Mali. In terms of date of collection, seven strains were collected before the year 2000, one strain was collected between the year 2000 and 2010, and 18 strains were collected between 2011 and 2017 (Table 1).

Analyses of sequences
Fig. 1 shows the evolutionary relationship of all the 26 selected strains. All the strains have a common ancestry and fall into five clades (clades 1–5) with 12, four, three, four, and three strains, respectively. All the strains from Nigeria belong to clade 1. Two strains, QKI86418 (Guinea) and QKI86419 (Guinea), belong to clade 2 and they have 100% identity. Also, strains OK186461 (Guinea), OK186463 (Guinea), and OK186464 (Guinea) belonging to clade 5 have 100% sequence identity. The percent identity matrix also reveals that OK186462 (Guinea), OK186465 (Guinea), OK186468 (Guinea), OK186466 (Guinea), and OK186462 (Guinea) have 100% sequence identity. Overall, 18 distinct strain sequences were observed.

Selection of antigenic peptides
From Table 2, 24 of the 26 selected sequences exceed the 0.4 threshold score for OPPA. Strains QKI86458 and QKI86425 are non-antigens. Strain CAI96544 has the highest OPPA value.

Evaluation of predicted cytotoxic T-lymphocyte epitopes
The binding prediction of CTL epitopes obtained from the sequences of the envelope glycoprotein of the Lassa mammarenavirus with the different alleles of the MHC I proteins is observed in Table 3. Table 3 also shows the selected four most immunodominant CTL epitopes. The peptides have an OPPA score greater than 0.4, a percentile rank in MHC I binding ≤0.05, and an MHC class I immunogenicity score greater than 0.00. CTL2 epitope has the highest MHC I binding score while CTL4 has the least. Also, they are all predicted to be non-toxins and non-allergens. The CTL 1 and CTL 2 epitopes exist in 13 each of the 26 strains; CTL3 epitopes exist in 15 out of 26; and CTL4 exists in 21 out of 26 strains.

Evaluation of predicted helper T-lymphocyte epitopes
From Table 4, the four most immunodominant HTL 15-mer
Table 2. Selection of antigenic peptides

| No. | Accession | OPPA   | Prediction |
|-----|-----------|--------|------------|
| 1   | QST04903  | 0.4929 | Antigen    |
| 2   | QKI86479  | 0.6317 | Antigen    |
| 3   | QKI86478  | 0.605  | Antigen    |
| 4   | QKI86468  | 0.626  | Antigen    |
| 5   | QKI86466  | 0.432  | Antigen    |
| 6   | QKI86465  | 0.6277 | Antigen    |
| 7   | QKI86464  | 0.627  | Antigen    |
| 8   | QKI86463  | 0.617  | Antigen    |
| 9   | QKI86462  | 0.6299 | Antigen    |
| 10  | QKI86461  | 0.6298 | Antigen    |
| 11  | QKI86458  | 0.3265 | Non-antigen|
| 12  | QKI86455  | 0.5126 | Antigen    |
| 13  | QKI86451  | 0.6204 | Antigen    |
| 14  | QKI86425  | 0.3643 | Non-antigen|
| 15  | QKI86419  | 0.6148 | Antigen    |
| 16  | QKI86418  | 0.6148 | Antigen    |
| 17  | QKI86385  | 0.445  | Antigen    |
| 18  | CAI96546  | 0.5471 | Antigen    |
| 19  | CAI96545  | 0.5968 | Antigen    |
| 20  | CAI96544  | 0.6683 | Antigen    |
| 21  | CAI96543  | 0.5359 | Antigen    |
| 22  | CAI96542  | 0.5919 | Antigen    |
| 23  | CAI96541  | 0.5378 | Antigen    |
| 24  | QIC50089  | 0.6206 | Antigen    |
| 25  | QIC50079  | 0.6382 | Antigen    |
| 26  | ACO37133  | 0.6009 | Antigen    |

OPPA, Overall Prediction of the Protective Antigen.

epitopes selected have OPPA values greater than 0.4 and are predicted to be high binders with MHC II alleles, non-toxic, non-allergic and inducers of IL-4, IL-10, and IFN-γ. HTL 1 exists in 1 of the 26 strains; HTL 2 exists in 1 of the 26 strains; HTL 3 exists in 1 of the 26 strains; and HTL 4 exists in 1 of the 26 strains.

Selected LBL epitopes

From Table 5, all LBL epitopes are predicted to be antigenic, non-toxic, and non-allergenic. LBL 1 exists in 13 of the 26 strains; LBL 2 exists in one of the 26 strains; LBL 3 exists in 12 of the 26 strains; and LBL 4 exists in six of the 26 strains.

Epitope modeling and molecular docking

The crystal structure of MHC protein (Fig. 2) shows loops in green, alpha helices in red, and beta sheets in yellow: (A) HLA-DRB1 (PDB ID: 6BIY); (B) HLA-A*3001 MHC, class II (PDB ID: 6J1W).

From Fig. 3, the 3D structures of epitopes, CTL1, CTL2, LBL1, LBL2, and LBL3 are made up of loop sequences; CTL3, CTL4, HTL1, HTL2, HTL3, and HTL4 are made up of alpha helices and loops; while LBL4 is made up of beta sheet and loops.

CTL1 binds to the MHC I protein (6BIY) more strongly than the co-crystallized ligands, CTL2, CTL3, and CTL4. Similarly, the four selected HTL epitopes had a higher affinity for the MHC II protein (6BIY) than the co-crystallized ligand (Table 6, Fig. 4). From Fig. 4, the VC has a high percentage of loops. The

Table 3. Antigenicity, immunogenicity, allergenicity, and toxicity prediction for immunodominant CTL epitopes

| Allele | Length | Peptide     | MHC I binding score | Percentile rank in MHC I binding | Antigenicity | Immuno-genicity | Allergenicity | Toxicity | Epitopes |
|--------|--------|-------------|---------------------|----------------------------------|--------------|----------------|---------------|-----------|----------|
| HLA-A*31:01 | 10     | RTRDIVISR   | 0.942767            | 0.01                             | 2.0644       | 0.18102         | Non-allergen | Non-toxin | CTL1     |
| HLA-A*31:01 | 9      | RTRDIVRS    | 0.866385            | 0.01                             | 1.8465       | 0.1809          | Non-allergen | Non-toxin | CTL2     |
| HLA-B*07:02 | 10     | RPSPIGYLGL  | 0.922526            | 0.04                             | 2.346        | 0.1047          | Non-allergen | Non-toxin | CTL3     |
| HLA-A*30:02 | 10     | HLSIPNFNQY  | 0.7102              | 0.05                             | 0.4268       | 0.10018         | Non-allergen | Non-toxin | CTL4     |

CTL, cytotoxic T lymphocyte; MHC, major histocompatibility complex; HLA, human leukocyte antigens.

Table 4. Evaluation of the most immunodominant HTL epitopes

| Allele | Length | Peptide               | OPPA  | Immuno-genicity | Allergenicity | Toxicity | IFN prediction | IL-4 | IL-10 | Epitope |
|--------|--------|-----------------------|-------|----------------|---------------|----------|----------------|------|-------|---------|
| HLA-DRB1*09:01 | 15     | SVQYNLSHAYAVDA     | 0.7054 | 87.8921               | Non-allergen | Non-toxin | Positive        | Inducer | Inducer | HTL1    |
| HLA-DRB4*01:01 | 15     | IMTSIQYLQIRNTTW   | 0.8557 | 83.8261               | Non-allergen | Non-toxin | Positive        | Inducer | Inducer | HTL2    |
| HLA-DRB4*01:01 | 15     | CIMTSIQYLQIRNTT   | 1.0603 | 83.0194               | Non-allergen | Non-toxin | Positive        | Inducer | Inducer | HTL3    |
| HLA-DRB5*01:01 | 15     | TRWMLIEANLKCFGN  | 0.6518 | 78.3517               | Non-allergen | Non-toxin | Positive        | Inducer | Inducer | HTL4    |

HTL, helper T lymphocyte; OPPA, Overall Prediction of the Protective Antigen; IFN, interferon; IL, interleukin; HLA, human leukocyte antigens.
Construction of the vaccine candidate

The following sequences of the VC are arranged in the following order: EAAK linker+adjuvant+EAAK linker+CTL1+AAY linker+CTL2+AAY linker+CTL3+AAY linker+CTL4+AAY linker+HTL1+GPGPG linker+HTL2+GPGPG linker+HTL3+GPGPG linker+HTL4+GPGPG linker+LBL1+GPGPG linker+LBL2+GPGPG linker+LBL3+GPGPG linker+LBL4+GPGPG linker.

The VC weighs 22.13 kDa and it constitutes the following 203 residues: EAAKHELSVLLEAAKRTRDIYISRRAAYRTRDIYISRAAYRPSPIGYLGLAAYHLISIPNFNQYAAYSVQYNLSHAVADAGPGPGIMTSYQYLRIQNTTGPGPGRWMLJEANLKCFCGNGPGPRTRDIYISR-RGPGPTTGRSGPGBPGRDIIYISRPGPGPSYIALDSGRGGWDCIMT.

Table 5. Evaluation of the selected LBL epitopes

| Peptide       | Length | Antigenicity | Allergenicity | Toxicity  | Epitopes |
|---------------|--------|--------------|---------------|-----------|----------|
| RTRDIYISRR   | 10     | 2.0644       | No            | Non-toxin | LBL1     |
| TTGRITS      | 9      | 1.8237       | No            | Non-toxin | LBL2     |
| TRODYISRR    | 9      | 1.7150       | No            | Non-toxin | LBL3     |
| SYIALDSGRGGWDCIMT | 17 | 1.2484       | No            | Non-toxin | LBL4     |

LBL, linear B lymphocyte.

Table 6. Docking results of epitopes and MHC proteins

| Protein-peptide complex | HawkDock score | Binding affinity (kcal/mol) |
|-------------------------|----------------|----------------------------|
| 6BIY (MHC1)+co-crystallized ligand | -3,089.2 | -32.02 |
| 6BIY (MHC1)+CTL1         | -3,453.67 | -39.95 |
| 6BIY (MHC1)+CTL2         | -3,016.71 | -28.21 |
| 6BIY (MHC1)+CTL3         | -2,895.57 | -39.87 |
| 6BIY (MHC1)+CTL4         | -2,450.56 | -16.25 |
| 6J1W (MHC2)+co-crystallized ligand | -2,286.6 | -9.51 |
| 6J1W (MHC2)+HTL1         | -2,616.8  | -25.35 |
| 6J1W (MHC2)+HTL2         | -2,947.51 | -25.68 |
| 6J1W (MHC2)+HTL3         | -3,245.58 | -21.74 |
| 6J1W (MHC2)+HTL4         | -3,038.01 | -37.49 |

MHC, major histocompatibility complex; CTL, cytotoxic T lymphocyte; HTL, helper T lymphocyte.

Fig. 2. (A, B) Crystal structures of major histocompatibility complex proteins.

Fig. 3. Three-dimensional structures of modeled epitopes. (A) Cytotoxic T lymphocyte (CTL)1, (B) CTL2, (C) CTL3, (D) CTL4, (E) helper T lymphocyte (HTL)1, (F) HTL2, (G) HTL3, (H) HTL4, (I) linear B lymphocyte (LBL)1, (J) LBL2, (K) LBL3, and (L) LBL4.

co-crystallized ligand for 6J1W (MHC 2) is an epitope from HIV1 ALA-ILE-PHE GLN-SER-SER-MET-THR-LYS. The co-crystallized ligand for 6BIY (MHC 1) is Histone 2B peptide.

From Fig. 5, CTL2 and CTL3 do not bind with the protein in the same pocket where the co-crystallized ligand binds. Similarly, HTL3 and HTL4 do not bind with the protein in the same pocket where the co-crystallized ligand binds.

The predicted physicochemical and immunological properties of vaccine candidate

From Table 7, the VC has an instability index less than 40, a molecular weight higher than 1 kDa, an isoelectric point (theoretical isoelectric point) higher than 7, an aliphatic index 66.5, and GRAVY value less than 0. The E. coli models have the longest predicted half-life of the VC. In terms of immunological qualities, the VC is a non-allergen, with antigenicity...
Fig. 4. (A, B) Refined three-dimensional structure of vaccine candidate (model 1) showing alpha helices in red, beta sheets in blue, and loops in green.

Table 7. Physicochemical and immunological profile of vaccine candidate

| Variable                         | Value       |
|----------------------------------|-------------|
| No. of amino acids               | 203         |
| Molecular weight                 | 2,213.02    |
| Theoretical isoelectric point    | 9.85        |
| Extinction coefficients          | 41,955      |
| Abs 0.1% (=1 g/L)                | 1.895       |
| Instability index                | 27.62 (stable) |
| Aliphatic index                  | 68.87       |
| Grand average of hydrophaticity  | -0.455      |
| N terminal sequence considered   | E (Glu)     |
| Mammalian reticulocytes, *in vitro* | 1 hr   |
| Yeast, *in vivo*                 | 30 min      |
| *Escherichia coli, in vivo*      | >10 hr      |

**Physicochemical properties**

**Immunological properties**

- Antigenicity: 0.8089
- MHC class I immunogenicity: 2.044270
- Allergenicity: Non allergen
- Solubility upon overexpression: 0.856289

MHC, major histocompatibility complex.

greater than the 0.4 threshold and immunogenicity greater than 0.00 for MHC class I. According to sequence-based prediction, the peptide’s solubility upon overexpression in *E. coli* is more than 0.5.

**Refinement and validation of 3D structure of vaccine candidate**

Of all the models, model 1 showed the lowest clash score, lowest poor rotamer value, and the highest number of residues in the Ramachandran favored region (Table 8). In terms of geometry, the VC has poor rotamers more than 0.3%, favored rotamers greater than 98%, Ramachandran outliers greater than 0.05%, Ramachandran favored greater than 98%, and a Rama distribution Z-score less than 2 (Table 9). The ERRAT score is more than 50%.

**Table 8.**

Refinement and validation of 3D structure of vaccine candidate

| Model   | Clash score | Poor rotamer | Favored rotamers | Outliers | Ramachandran favored | Rama distribution Z-score | ERRAT score |
|---------|-------------|--------------|------------------|----------|-----------------------|---------------------------|-------------|
| Model 1 | 2.5         | 0.2%         | 98%              | 0.3%     | 98%                   | -2                        | 50%         |

**Fig. 5.** Major histocompatibility complex (MHC)-epitope binding complexes. (A) 6BIY (MHC1)+co-crystallized ligand. (B) 6BIY (MHC1)+cytotoxic T lymphocyte (CTL1). (C) 6BIY (MHC1)+CTL2. (D) 6BIY (MHC1)+CTL3. (E) 6BIY (MHC1)+CTL4. (F) 6J1W (MHC2)+co-crystallized ligand. (G) 6J1W (MHC2)+helper T lymphocyte (HTL1). (H) 6J1W (MHC2)+HTL2. (I) 6J1W (MHC2)+HTL3. (J) 6J1W (MHC2)+HTL4.
Table 8. Refinement of model 1 of vaccine candidate

| Model    | GDT-HA | RMSD | Mol probity | Clash score | Poor rotamers | Rama favored |
|----------|--------|------|-------------|-------------|---------------|--------------|
| Initial  | 1      | 0    | 2.405       | 15.6        | 1.3           | 86.6         |
| Model 1  | 0.9458 | 0.423| 2.011       | 10.6        | 0             | 92.5         |
| Model 2  | 0.9458 | 0.414| 2.215       | 16.3        | 0             | 91.5         |
| Model 3  | 0.931  | 0.447| 2.093       | 12.5        | 0             | 92           |
| Model 4  | 0.9433 | 0.431| 2.172       | 16          | 0.6           | 92.5         |
| Model 5  | 0.9544 | 0.408| 2.121       | 12.8        | 0.6           | 91.5         |

GDT-HA, global distance test-high accuracy; RMSD, root-mean-square deviation.

Table 9. Validation of structure of model 1 of refined vaccine candidate

| Protein geometry | Value |
|------------------|-------|
| Poor rotamers    | 1 (0.64) |
| Favored rotamers | 155 (98.73) |
| Ramachandran outliers | 3 (1.49) |
| Ramachandran favored | 186 (92.54) |
| Rama distribution Z-score | -1.28±0.54 |
| ERRAT score      | 90.15 |

Values are presented as number (%), mean±standard deviation, or score.

Table 10. Disulphide engineering of vaccine construct

| Res1 chain | Res1 Seq # | Res1 AA | Res2 chain | Res2 Seq # | Res2 AA | Chi3 | Energy | Sum B-factors |
|------------|------------|---------|------------|------------|---------|------|--------|---------------|
| A          | 14         | ALA     | A          | 20         | ILE     | -109.27 | 6.19   | 0             |
| A          | 14         | ALA     | A          | 40         | TYR     | 120.82  | 5.78   | 0             |
| A          | 35         | ILE     | A          | 71         | ASN     | 123.95  | 7.13   | 0             |
| A          | 43         | SER     | A          | 76         | TYR     | -88.46  | 3.94   | 0             |
| A          | 50         | LEU     | A          | 70         | TYR     | -87.97  | 1.31   | 0             |
| A          | 59         | ASN     | A          | 62         | GLN     | -83.29  | 1.78   | 0             |
| A          | 72         | LEU     | A          | 90         | SER     | 121.6   | 4.39   | 0             |
| A          | 74         | HIS     | A          | 91         | TYR     | -88.25  | 0.89   | 0             |
| A          | 75         | ALA     | A          | 90         | SER     | -103.81 | 2.98   | 0             |
| A          | 77         | ALA     | A          | 85         | PRO     | -105.82 | 1.99   | 0             |
| A          | 79         | ASP     | A          | 85         | PRO     | 85.72   | 3.07   | 0             |
| A          | 92         | GLN     | A          | 116        | ARG     | 95.02   | 3.5    | 0             |
| A          | 98         | ASN     | A          | 101        | TRP     | 115.22  | 5.38   | 0             |
| A          | 113        | GLN     | A          | 134        | ALA     | -109.89 | 4.14   | 0             |
| A          | 131        | LEU     | A          | 134        | ALA     | 105.33  | 4.9    | 0             |
| A          | 143        | PRO     | A          | 147        | ARG     | 90.9    | 1.63   | 0             |
| A          | 154        | SER     | A          | 190        | ALA     | 82.67   | 2.92   | 0             |
| A          | 157        | GLY     | A          | 174        | ARG     | 101.57  | 2.04   | 0             |
| A          | 163        | THR     | A          | 170        | GLY     | 85.52   | 3.78   | 0             |
| A          | 166        | THR     | A          | 170        | GLY     | 118.34  | 5.43   | 0             |
| A          | 169        | PRO     | A          | 173        | THR     | 107.11  | 1.52   | 0             |
| A          | 174        | ARG     | A          | 193        | SER     | 109.84  | 2.06   | 0             |
| A          | 175        | ASP     | A          | 193        | SER     | 126.89  | 2.81   | 0             |
| A          | 177        | TYR     | A          | 190        | ALA     | 112.99  | 3.59   | 0             |
| A          | 179        | SER     | A          | 190        | ALA     | 124.8   | 4.48   | 0             |
| A          | 192        | ASP     | A          | 195        | ARG     | 94.61   | 2.26   | 0             |

Disulphide engineering of vaccine construct

The VC has 21 disulphide bonds (Table 10, Fig. 6A). Model 1 has the least docking score and the least ligand RMSD (Table 11, Fig. 6B).

Molecular dynamics simulation

As shown in Fig. 7A, most of the first 100 residues have their B-factor values higher than 0.6. The VC showed the greatest disorder at the N terminus. All eigenvalues are positive. With each progressive mode, the eigenvalues gradually increase (Fig. 7B). A very small percentage of the residues had deformability values greater than 0.6 (Fig. 7C). Residues with values greater than 0.6 are seen near the residue 600 but the residues at the N terminus are close to the threshold. Individual variance (Fig. 7D) is shown by red bars, whereas cumulative variance is represented by green bars. Overall, when individual variance dropped, cumulative variation increased.

As shown in Fig. 8A, mild stiffness is seen between residues 1,200 and 1,400, and a few other areas. The molecule is primarily stiffness-free. In Fig. 8B, the molecule shows predomi-
nantly anti-correlated and correlated motions of the residues.

**Codon adaptation and in-silico cloning**

GAAGCTGCTAAACACGAAGCCTCTGCTGTAACGTG- (50); CCGTGACATCTATCTCTCCTCTCGTGCTGGTACCTCCTGCTGGAAGCTGCTTACCAGTACCCGCTGACATCT (100); ACATCTCTCGTGCTGCTGCTTACCGTACCCGCTGACCTCCTGCTGGAAGCTGCTTACCAGTACCCGCTGACATCT (150); GCAGCTGACATCTATCTCTCCTCTCGTGCTGGTACCTCCTGCTGGAAGCTGCTTACCAGTACCCGCTGACATCT (200); TGGGGTCCGGGTCCGGGTACCCGTTGGATGCTGATCGAAGCTGCTTACCAGTACCCGCTGACATCT (250); CCAGAACACCA-

**Fig. 6.** Three-dimensional structure of vaccine candidate refined and mutated by disulphide engineering showing alpha helices in red, beta sheets in blue, and loops in green (A), complexed with the toll-like receptor-4 receptor after molecular docking showing alpha helices in red, beta sheets in blue, and loops in green (B).

**Table 11.** Molecular docking scores between the vaccine construct and the toll-like receptor-4 receptor

| Rank | Docking score | Ligand RMSD |
|------|--------------|-------------|
| 1    | -340.91      | 57.26       |
| 2    | -330.66      | 76.13       |
| 3    | -314.9       | 76.35       |
| 4    | -305.59      | 75.70       |
| 5    | -303.07      | 60.14       |
| 6    | -300.95      | 120.21      |
| 7    | -300.46      | 75.42       |
| 8    | -298.24      | 78.02       |
| 9    | -297.21      | 62.59       |
| 10   | -294.55      | 81.08       |

RMSD, root-mean-square deviation.

**Fig. 7.** Charts for toll-like receptor-4 vaccine candidate complex showing (A) the B factor (temperature), (B) eigenvalue chart, (C) deformability chart, and (D) variance char. NMA, normal mode analysis; PDB, Protein Data Bank.
The translated sequence has a 100% resemblance with the input sequence. Also, the VC has a GC-content of the improved sequence between 50–60 and a very high CAI-value (Table 12, Fig. 9).

**Discussion**

LF kills thousands of people annually on the West African subcontinent causing a colossal negative socio-economic impact. The development of a vaccine would go a long way in the prevention of the disease. Through a bioinformatics approach, 12 immunodominant epitopes (4 CTL, 4 HTL, and 4 LBL) from 26 different strains of the envelope glycoprotein of the Lassa mammarenavirus was used to design a VC. Phylogenetic analyses reveal that while most of the strains are closely related, 18 of them have distinct sequences suggesting that genetic mutations could likely affect virulence, transmissibility, and patient’s immune response to new strains after prior exposure to older strains [18].

Screening of strains for antigenicity was based on a threshold score of 0.4 with those having higher values considered as probable antigens [19]. In prospecting for immunodominant epitopes, factors such as antigenicity, immunogenicity, allergenicity, toxicity, and the ability to induce certain cytokines (IFN-γ, IL-4, and IL-10) are important [20]. Fig. 3 shows that the modeled epitopes are structurally diverse, implying that they will elicit distinct immunological responses when combined with the MHC proteins in Fig. 5 [21]. The presence of
alpha helices and beta sheets in the selected epitopes ensures stability of the proposed VC [22]. Hydrophobic interactions and backbone hydrogen bonding help to keep beta sheets stable while backbone hydrogen bonding helps to stabilize alpha helices. These secondary structures will play a crucial role in the determination of global structure and function of the VC [23]. Usually binding to a protein target by a ligand affects its bioactivity. Remarkably, certain selected epitopes (CTL1, HTL1, HTL2, and HTL3) are predicted to bind better with their respective MHC proteins suggesting stronger immunological output. The selected epitopes that do not bind in the same pocket as the co-crystallized ligands suggest diverse immunological responses.

Linkers were employed in the design of the VC to improve expression, folding, and stability ultimately to improve immunogenicity [24]. Adjuvants are used to stimulate both cell-mediated and humoral immune responses and consequently potentiate vaccine and extend the duration of the immunological response. This is because subunit vaccines do not elicit the same level of immunogenicity as live attenuated vaccines [25].

The physicochemical properties of the antigen play a critical role in the recognition by the T cell receptor [26]. In terms of molecular weight, the predicted VC is large enough to elicit significant immune response since it is significantly larger than a hapten [27]. The pH of a solution at which the net charge of a protein becomes zero is known as the isoelectric point. With an isoelectric point of 9.85, the VC shows basicity and this determines its electrophoretic mobility and consequently its purification [28]. With an instability index less than 40, the VC is predicted to be stable after expression [29] and it is also thermo-stable across wide temperature range due to high aliphatic index [29]. A negative GRAVY value indicates that the VC is non-polar suggesting depicting hydrophobicity [29].

From the results of the half-life estimation, the VC is best expressed in an E. coli model due to greater bioavailability. According to sequence-based prediction, the VC’s solubility upon overexpression in E. coli is 0.859289 (more than 0.5), which improves the quality of its function [30].

The VC is predicted to be significantly antigenic, immunogenic, and non-allergenic and suitable for further development. It also shows significant solubility when over expressed in E. coli which would enhance its function [30]. The refinement of the structure of the VC was achieved through side-chain repacking [13]. Model 1 had the lowest clash score, the lowest poor rotamer value, and the greatest number of residues in the Ramachandran favored region of all the models. The high quality of model 1 is further validated with a Ram distribution Z-score less than 2 and a very high ERRAT score [31].

Disulphide engineering improves the 3D structural stability of proteins and with 21 disulphide bonds; the VC is predicted to be very stable [32]. After docking the refined and mutated VC with the TLR4 receptor, model 1 was considered the best because it had the least ligand RMSD and the least docking score.

Since protein dynamics also determines its function, MDS was used to evaluate the VC’s stability with the TLR4 receptor. The B factor predicts the thermostability of a protein molecule. B-factor values higher than 0.6 in the N-terminus suggest disorder. This is expected as amino acid flexibility is greatest in the N and C termini of proteins [33]. The main chain deformability of a molecule is measured by its ability to deform at each of its residues. The stability of the VC is further validated by the deformability values of most of the residues being less than 0.6 [34]. The eigenvalue plot measures the relative modal stiffness of the structure of a protein molecule. All eigenvalues are positive and with each progressive mode, the eigenvalues gradually increased. A gradual increase in the modal index is suggestive of structural stability as greater energy is required to distort the structure [34].

The variance associated with the modes denotes their contribution to the equilibrium motions and the eigenvalue is inversely related to the individual variance associated with each normal mode [34]. Individual variance is depicted by red bars, while cumulative variance is depicted by green bars. Individual variance decreased from mode 1 to mode 20 where it tended to zero, while cumulative variation increased. This shows that the VC is stable [34].

The elastic normal mode measures the stiffness of a molecule. It has been postulated that stiffness may influence function so a protein’s dynamics influence its function [35]. The elastic normal mode specifies which pair of residues is linked by springs. Each dot on the graph indicates a spring between a pair of residues. The stiffness of the springs is indicated by the color of the dots; darker grays indicate stiffer springs and vice versa [34]. The structure of the VC is predominantly stiffness-free except for few patches of stiffness seen between atoms 1,000 and 1,400. The covariance matrix shows which sections of the protein move correlated (red), uncorrelated (white), or anti-correlated (blue) [34]. From Fig. 8B, the mol-
ecule shows predominantly anti-correlated and correlated motions of the residues. Overall, the results suggest that the structure of the VC is of good quality. E. coli K12 strain was selected as the recombinant vector of choice because from earlier results it was predicted to have the longest shelf life. Codon optimization was used to obtain high levels of expression of our recombinant vaccine peptide in this vector [35]. The CAI is widely used to evaluate codon bias. Excluding the termination codons and non-synonymous codons, this relative adaptiveness is defined as the ratio of each codon’s usage to that of the most prevalent codon for the same amino acid [36,37]. CAI values range between 0 and 1, with higher values reflecting a greater fraction of the most abundant codons [36]. A value of 1.0 indicates that all isoleucine codons are the most common (AUU). As you move away from 1.0, the AUC and AUA options become more numerous [38].

The VC has a CAI value of 1.0. In bacteria, increasing GC concentration in the VC improves translation and enhances high-level protein expression [36,37]. GC content should be between 40% and 60% to achieve stable primer-template binding and efficient amplification, with values above 65% generating unsatisfactory results because the templates fold into intricate secondary structures [39,40].

In conclusion, a range of immunoinformatic techniques were used in this study to identify possible CTL, HTL, and LBL epitopes in the envelope glycoprotein of the Lassa mammarenavirus. The VC designed from these epitopes is predicted to have good physicochemical and immunological properties. It is also predicted that binding of the VC to the immune receptor, TLR4 will elicit a significant immune response against LF infection. Overall, our findings reveal that the VC has good promise based on the computational data generated. It is advised that the candidate be evaluated further in-vitro and in-vivo to demonstrate its safety and efficacy in the prophylaxis of LF.

**ORCID**

Adekunle Babajide Rowaiye  [https://orcid.org/0000-0002-8773-5944](https://orcid.org/0000-0002-8773-5944)

Ezinne Janefrances Nwonu [https://orcid.org/0000-0001-8392-9170](https://orcid.org/0000-0001-8392-9170)

Titilayo Mercy Asala [https://orcid.org/0000-0003-3464-2936](https://orcid.org/0000-0003-3464-2936)

Amoge Chidinma Ogu [https://orcid.org/0000-0003-3461-2737](https://orcid.org/0000-0003-3461-2737)

Doofan Bur [https://orcid.org/0000-0001-7625-7118](https://orcid.org/0000-0001-7625-7118)

Chimaobi Chukwu [https://orcid.org/0000-0003-2411-0290](https://orcid.org/0000-0003-2411-0290)

Angus Nnami Oli [https://orcid.org/0000-0001-8519-2555](https://orcid.org/0000-0001-8519-2555)

Tarimoboere Agbalalah [https://orcid.org/0000-0002-5875-4126](https://orcid.org/0000-0002-5875-4126)

**References**

1. Hallam HJ, Hallam S, Rodriguez SE, et al. Baseline mapping of Lassa fever virology, epidemiology and vaccine research and development. NPJ Vaccines 2018;3:11.

2. Rowaiye AB, Oli AN, Onuh OA, et al. Rhamnetin is a better inhibitor of SARS-CoV-2 2’-O-methyltransferase than dolutegravir: a computational prediction. Afr J Infect Dis 2022;16:80-96.

3. Ogbe O, Ajuoluchukwu E, Uneke CJ. Lassa fever in West African sub-region: an overview. J Vector Borne Dis 2007;44:1-11.

4. Merson L, Bourner J, Jalloh S, et al. Clinical characterization of Lassa fever: a systematic review of clinical reports and research to inform clinical trial design. PLoS Negl Trop Dis 2021;15:e0009788.

5. Oyekibbor AO. Lived experiences of Lassa fever survivors in Southsouth Nigeria [dissertation]. Minneapolis (MN): Walden University; 2021.

6. Fischer RJ, Pursushotham JN, van Doremalen N, et al. ChAdOx1-vectored Lassa fever vaccine elicits a robust cellular and humoral immune response and protects guinea pigs against lethal Lassa virus challenge. NPJ Vaccines 2021;6:32.

7. Shanmugam A, Rajoria S, George AL, Mittelman A, Suriano R, Tiwari RK. Synthetic Toll like receptor-4 (TLR-4) agonist peptides as a novel class of adjuvants. PLoS One 2012;7:e30839.

8. Herrera LR. In silico approach in designing a novel multi-epitope vaccine candidate against non-small cell lung cancer with overexpressed G protein-coupled receptor 56. Asian Pac J Cancer Prev 2020;21:2297-306.

9. Dorosti H, Eslami M, Negahdaripour M, et al. Vaccinomics approach for developing multi-epitope peptide pneumococcal vaccine. J Biomol Struct Dyn 2019;37:3524-35.

10. Nain Z, Abdulla F, Rahman MM, et al. Proteome-wide screening for designing a multi-epitope vaccine against emerging pathogen Elizabethkingia anophelis using immunoinformatic approaches. J Biomol Struct Dyn 2020;38:4850-67.

11. Abdellrazeq GS, Fry LM, Elnaggar MM, et al. Simultaneous cognate epitope recognition by bovine CD4 and CD8
T cells is essential for primary expansion of antigen-specific cytotoxic T-cells following ex vivo stimulation with a candidate Mycobacterium avium subsp. paratuberculosis peptide vaccine. Vaccine 2020;38:2016-25.

12. Martin WR, Cheng F. A rational design of a multi-epitope vaccine against SARS-CoV-2 which accounts for the glycan shield of the spike glycoprotein. J Biomol Struct Dyn 2022;40:7099-113.

13. Heo L, Park H, Seok C. GalaxyRefine: protein structure refinement driven by side-chain repacking. Nucleic Acids Res 2013;41(Web Server issue):W384-8.

14. Schiffrin B, Radford SE, Brockwell DJ, Calabrese AN. PyXlinkViewer: a flexible tool for visualization of protein chemical crosslinking data within the PyMOL molecular graphics system. Protein Sci 2020;29:1851-7.

15. Yan Y, Zhang D, Zhou P, Li B, Huang SY. HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. Nucleic Acids Res 2017;45(W1):W365-73.

16. Sharp PM, Li WH. The codon Adaptation Index: a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 1987;15:1281-95.

17. Stokes WA, Glick BS. MICA: desktop software for comprehensive searching of DNA databases. BMC Bioinformatics 2006;7:427.

18. Weisblum Y, Schmidt F, Zhang F, et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. Elife 2020;9:e61312.

19. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics 2007;8:4.

20. Kumar J, Qureshi R, Sagurthi SR, Qureshi IA. Designing of nucleocapsid protein based novel multi-epitope vaccine against SARS-CoV-2 using immunoinformatics approach. Int J Pept Res Ther 2021;27:941-56.

21. Scheiblhofer S, Laimer J, Machado Y, Weiss R, Thalhamer J. Influence of protein fold stability on immunogenicity and its implications for vaccine design. Expert Rev Vaccines 2017;16:479-89.

22. Gu H, Liao Y, Zhang J, et al. Rational design and evaluation of an artificial Escherichia coli K1 protein vaccine candidate based on the structure of OmpA. Front Cell Infect Microbiol 2018;8:172.

23. Haimov B, Srebnik S. A closer look into the α-helix basin. Sci Rep 2016;6:38341.

24. Aldakheel FM, Abrar A, Munir S, et al. Proteome-wide mapping and reverse vaccinology approaches to design a multi-epitope vaccine against Clostridium perfringens. Vaccines (Basel) 2021;9:1079.

25. Bonam SR, Partidos CD, Halmuthur SK, Muller S. An overview of novel adjuvants designed for improving vaccine efficacy. Trends Pharmacol Sci 2017;38:771-93.

26. Greenbaum JA, Andersen PH, Blythe M, et al. Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. J Mol Recognit 2007;20:75-82.

27. Erkes DA, Selvan SR. Hapten-induced contact hypersensitivity, autoimmune reactions, and tumor regression: plausibility of mediating antitumor immunity. J Immunol Res 2014;2014:175265.

28. Islinger M, Wildgruber R, Volkl A. Preparative free-flow electrophoresis, a versatile technology complementing gradient centrifugation in the isolation of highly purified cell organelles. Electrophoresis 2018;39:2288-99.

29. Oladipo EK, Ajayi AF, Ariyo OE, et al. Exploration of surface glycoprotein to design multi-epitope vaccine for the prevention of COVID-19. Inform Med Unlocked 2020;21:100438.

30. Habibi N, Mohd Hashim SZ, Norouzi A, Samian MR. A review of machine learning methods to predict the solubility of overexpressed recombinant proteins in Escherichia coli. BMC Bioinformatics 2014;15:134.

31. Williams CJ, Headd JJ, Moriarty NW, et al. MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci 2018;27:293-315.

32. Navone L, Vogl T, Luangthongkam P, et al. Disulfide bond engineering of AppA phytase for increased thermostability requires co-expression of protein disulfide isomerase in Pichia pastoris. Biotechnol Biofuels 2021;14:80.

33. Smith DK, Radivojac P, Obradovic Z, Dunker AK, Zhu G. Improved amino acid flexibility parameters. Protein Sci 2003;12:1060-72.

34. Lopez-Blanco JR, Aliaga JI, Quintana-Orti ES, Chacon P. iMODS: internal coordinates normal mode analysis server. Nucleic Acids Res 2014;42(Web Server issue):W271-6.

35. Xing Y, Gong R, Xu Y, Liu K, Zhou M. Codon usage bias affects α-amylase mRNA level by altering RNA stability and cytosine methylation patterns in Escherichia coli. Can J Microbiol 2020;66:521-8.

36. Lee S, Weon S, Lee S, Kang C. Relative codon adaptation index, a sensitive measure of codon usage bias. Evol Bioinform Online 2010;6:47-55.
37. Newman ZR, Young JM, Ingolia NT, Barton GM. Differences in codon bias and GC content contribute to the balanced expression of TLR7 and TLR9. Proc Natl Acad Sci USA 2016;113:E1362-71.

38. Grote A, Hiller K, Scheer M, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res 2005;33(Web Server issue): W526-31.

39. Biocompare. Bench Tips: Primers, by design: tips for optimal DNA primer design [Internet]. San Francisco (CA): Biocompare; 2013 [cited 2021 Mar 8]. Available from: https://www.biocompare.com/Bench-Tips/133581-Primers-by-Design-Tips-for-Optimal-DNA-Primer-Design/.

40. Kayama K, Kanno M, Chisaki N, et al. Prediction of PCR amplification from primer and template sequences using recurrent neural network. Sci Rep 2021;11:7493.