Infection, Genetics and Evolution 85 (2020) 104464

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research Paper

Computational design of a potential multi-epitope subunit vaccine using immunoinformatics to fight Ebola virus

Ajinkya Kadam[1], Santanu Sasidharan[1], Prakash Saudagar*

Department of Biotechnology, National Institute of Technology, Warangal 506004, Telangana, India

ARTICLE INFO

Keywords:
Immunoinformatics
MD simulation
Ebola virus
Multi-epitope vaccine
in-silico clone

ABSTRACT

Ebola virus (EBOV) is a rare but fatal disease that has been a burden to mankind for over 40 years. EBOV exhibits several symptoms including severe bleeding, organ failure and if left untreated causes death. It is assumed that fruit bats of the Pteropodidae family are natural hosts for the virus. Over the years, there has been no effective vaccine that can confer immunity to this virus. Considering the necessity of a vaccine against EBOV, this study to develop a multi-epitope subunit vaccine for the EBOV using the immunoinformatics approach was conducted. The construct was designed using structural and non-structural proteins of EBOV. Class I and Class II MHC epitopes were predicted and linked along with β defensin and compatible linkers. B-cell linear epitopes were also assessed and the physiological parameters of the vaccine were determined. The vaccine was capable of administration to humans and also is capable of an immune response. The vaccine was modeled further and affinity towards the TLR4 receptor was studied by docking and simulation for 20 ns. The trajectory analysis high affinity between the vaccine and the construct with an average hydrogen bond of 18. For ease of purification, the vaccine construct was ligated into pET28a(+) vector with His-tag. Concluding from the results, the vaccine construct has the potentiality to help develop immunity against the Ebola virus. Furthermore, experimental and immunological investigations will be required to verify the feasibility of the multi-epitope subunit construct as a commercial vaccine.

1. Introduction

Ebola virus (EBOV) is a serious and fatal disease that has survived on the planet for almost four decades. EBOV belongs to the Filoviridae family and is generally composed of single-stranded negative-sense RNA genome. They are enveloped viruses capable of causing fatal hemorrhagic fever, severe bleeding in human and primates. The virus has been recorded to have a fatality rate of 90%. The disease cycles in an instant with flu-like illness, followed by an initial incubation period (2–21 days). Common signs and symptoms of EBOV include nausea, cough, edema, organ failure, anorexia, postural hypotension, chest pain, vomiting, neurological complications and mucosal hemorrhage (Genton et al., 2012). EBOV spreads through contact with bodily fluids or blood from an infected living entity (Bausch et al., 2007). Wild-type Filovirus is the most deadly virus with a mortality rate ranging from 83% - 90%, in human cases only (Feldmann et al., 2013). Five strains of EBOV have been recorded to date in Sudan (SEBOV), Ivory Coast, Zaire (ZEBOV), Reston and Bundibugyo (Towner et al., 2008). Among these strains, SEBOV and ZEBOV have recorded the highest number of deaths to date and also are the most commonly occurring strains. The mode of transmission of these strains is mostly caused by the human-human transmission of bodily fluids. The high mortality rates, ineffective drugs and non-availability of potent vaccine have pushed the researchers for control and prevention of this disease.

An ideal vaccine is one that elicits both cell-mediated immune response and humoral immune response and prevents relapse of the infection. Conventional vaccines contain either attenuated or inactivated whole pathogen while the subunit vaccine does not contain live components of the pathogen. Conventional vaccines have a number of limitations, but safety is the fundamental problem as the pathogens employed for vaccination may revive its pathogenic form, potential enough to cause infection again. Researchers have been working to develop potential vaccines against EBOV. In December 2016, a study on the rVSV-EBOV vaccine found it to be effective by 70–100% against the Ebola virus, thereby, being the first vaccine against Ebola Virus Disease (Henao-Restrepo et al., 2017). The vaccine is currently in use with half of the administered people exhibiting adverse effects and several other vaccine candidates in clinical trials. The subunit vaccines contain

* Corresponding author.
E-mail address: ps@nitw.ac.in (P. Saudagar).
[1] The authors contributed equally.

https://doi.org/10.1016/j.meegid.2020.104464
Received 6 May 2020; Received in revised form 23 June 2020; Accepted 12 July 2020
Available online 15 July 2020

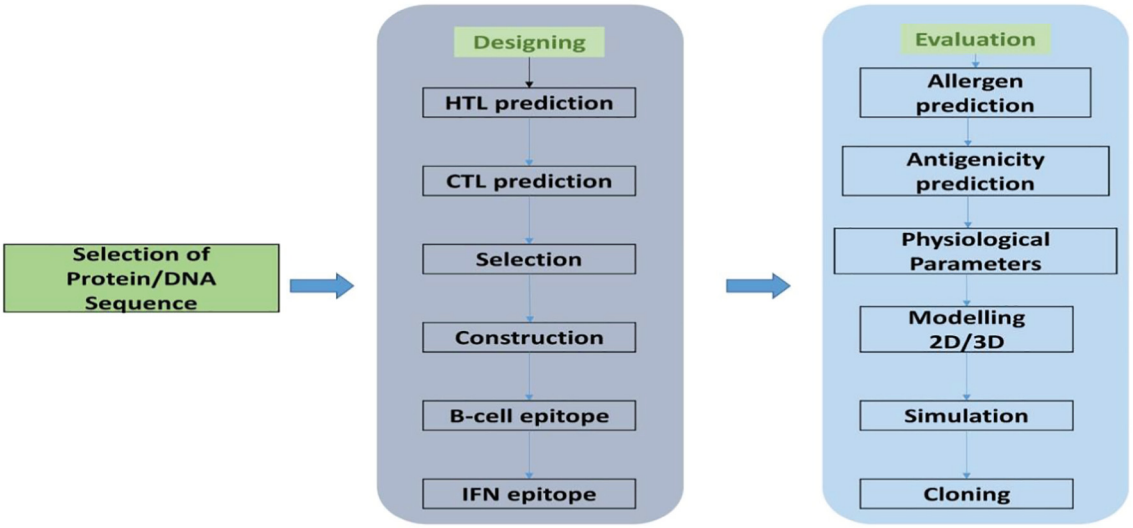

Fig. 1. Workflow of the study: Two phases i.e. designing and evaluation, of the study are illustrated in the figure where the experiments performed under each work phase are described.

antigenic parts of the pathogen that have the capability to evoke a protective, specific, long-lasting and also strong immune response within the human body with less cost and minimal side effects (Schiller and Lowy, 2014). Methods of vaccine design based on the T cell epitope can be defined as the detection of immune-dominant epitopes of the structural and non-structural proteins and synthesizing the protein to be used as vaccines to induce an immune response that is effective (Patronov and Doytchinova, 2013). Computer-based prediction tools have been known to greatly reduce the validation experiments performed and also reduce the time for epitope prediction (Lin et al., 2008). The potentiality of a T cell epitope-based vaccine is much more compared to B cell epitopes because it is composed of MHC-1 and MHC-2 restricted epitopes recognized by various clonal cells thereby enhancing their ability to induce strong cellular and humoral immune response (Klavinskis et al., 1989).

In this study, the authors designed and built a multi-epitope vaccine by utilizing immunoinformatic tools. The work was divided into two phases as represented in Fig. 1 i.e. design and validation. The design part involved the use of key proteins of EBOV that causes the disease for vaccine construction. The protein sequences were retrieved from the UniProt database and the epitopes were predicted and linked together. The adjuvant β-defensin was attached to the N-terminal of the constructs to improve the vaccine immunogenicity. Overlapping CTLs and HTLs were considered, followed by testing for B-cell epitopes in the vaccine. The physiological parameters of the vaccine were also determined and allergenicity/antigenicity properties of the vaccine construct were checked for. The vaccine construct was modeled, docked and dynamic simulations with TLR4 receptor were studied. The results of the study indicated an increased binding affinity between the multi-epitope vaccine and the human host TLR4 receptor and therefore, the vaccine construct can be a potent vaccine candidate for further real-time studies.

2. Materials and methods

2.1. Sequence selection

Ebola virus is a unique virus with restricted genes and protein in the viral genome. The single-stranded RNA virus has a nucleoprotein NP, viral protein VP35, VP30, VP 24 and VP60, surface glycoprotein GP and polymerase L in the 3′-5′ direction respectively. VP proteins are involved in inhibiting interactions, transcription, maturation of cells and signaling while GP is responsible for the virus attachment and entry. NP protein is a key component in the ribonucleoprotein complex and also protects the RNA genome whereas L is involved in transcription and regulation of the viral genome. Therefore, the amino acid sequences of these proteins were taken from the UniProt Knowledge Base (UniProtKB) database while eliminating the variable region for seven proteins of the Mayinga strain of Zaire EBOV. Protein sequenced were provided as input to the VaxiJen v2.0 server which predicts the subunit vaccine regions and also the antigen regions in a protein using defaults parameter (Doytchinova and Flower, 2007). Four proteins were selected based on the VaxiJen 2.0 score out of the seven proteins of the Mayinga strain. All parameters used for different servers in this study were maintained as default unless otherwise specified.

2.2. MHC class I cytotoxic T-lymphocytes epitope prediction

Cytotoxic T-lymphocyte or CTL have been known to be involved in infections and also in responses involving malfunctioning of cells. They are also known to perform effector function in the adaptive immunity part of the human body. CD8+ T cells and their receptors are capable of perceiving the peptides that are exhibited by the MHC-I complex. Self-peptides cannot elicit any reactions in the CTL's whereas the remote antigens evoke the CTLs to produce a cytotoxic reaction. Taking the role of CTL peptides into consideration for their role in the immunization, NetCTL-1.2 server was used for CTL prediction (Larsen et al., 2007). The web server (http://www.cbs.dtu.dk/services/NetCTL/) is designed for the identification of human CTL epitopes in target proteins. A2, A3 and B7 supertypes were selected for prediction as it covers almost 83–88.8% of the people worldwide and also from different ethnicity. A threshold score of 0.75 was set for the prediction in all the peptides (Lund et al., 2004). All 9mer peptides of the input protein were provided a score depending upon the cleavage of proteasomes, proficiency in the transport of TAP and mostly HLA-I restriction.

2.3. MHC class II helper T-lymphocytes epitope prediction

Elicitation of an immune response after antigenic peptide binding is conferred by major histocompatibility complex (MHC) and they expose the peptides to the respective T cells. MHC has two different and unique properties that make the evasion of pathogens difficult i.e. the MHC complexes are polygenic and polymorphic. They assist the activation of

CTL's and macrophages. Therefore, MHC based epitopes are preferably selected in the vaccine design study. Epitope specific for Helper T-lymphocyte cell or HTL receptor was checked in the selected four EBOV proteins by using the Immune Epitope Database (IEDB) MHC-II epitope server (Wang et al., 2008). The complete HLA reference set was utilized in the online tool IEDB server (http://tools.iedb.org/mhcii/) to predict the HTL epitopes. The method of prediction chosen was the recommended IEDB parameters where a complete set of 27 human alleles that includes 99% of the population was used (Vita et al., 2014). The epitopes of HTL selected depended on the IC_{50} values where the epitopes with low IC_{50} values have high affinity. The percentile rank of the predicted epitopes was based on the comparison to randomly generated SWISS-PRO database peptides numbering 5 million. The peptides that had higher affinity were provided the least rank.

2.4. Designing of multi-epitope vaccine

The epitope sequences were arranged and CTL and HTL epitopes that overlapped were given the first preference of choice. The final design was constructed by linking the epitopes using appropriate linkers. GPGPG linkers and AAY linkers were employed to link the HTL epitopes and CTL epitopes, respectively (Hajighahramani et al., 2017; Mittal et al., 2020). The vaccine sequence was provided with a β-defensin peptide at the N-terminal to improve the immunogenicity and immune response of the vaccine. β-defensin has been known to be effective against HIV and HCV contamination and therefore the peptide found its place in the N-terminal of the vaccine. This was followed by the top CTL epitopes corresponding to each supertype and later by two overlapped HTL epitopes of all four proteins. The adjuvant and CTL epitopes were linked together by the EAAAK linker (Ikram et al., 2018).

2.5. Prediction of B-cell epitope

B-cells have been known to get actively involved in antibody generation, humoral immunity, and also result in memory cell formation towards encountering future pathogens. The humoral immunity responds to linear B-cell epitopes that are specific epitopes and therefore, the epitopes for the vaccine construct were predicted by utilizing the BCPred Server (http://ailab.ist.psu.edu/bcpred/predict.html). The prediction of B-cell epitopes is carried out by an artificial neural network in this server. The anticipation of the linear B-cell epitopes in the server was set at 75% for 20-mer length linear epitopes.

2.6. Antigenicity, allergenicity and physiological parameters evaluation

The major problem with vaccines is the occurrence of allergy when administered to the human body and therefore, it is crucial to check for allergenicity. Considering the need, the vaccine construct was assessed for allergenicity in the AlgPred server (http://www.imtech.res.in/raghava/algpred/) (Saha and Raghava, 2006). VaxiJenv2.0 was used to predict the antigenicity of the vaccine construct (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) (Doytchinova and Flower, 2007). For the prediction of the physiological properties with respect to the designed vaccine construct, ProtParam Server (http://web.expasy.org/protparam/) was put into use. Properties that were predicted were the molecular weight, theoretical pI, instability index, half-life, aliphatic index and Grand Average of Hydropathy (GRAVY). These values have a critical role, in determining the pH characteristics of a protein, in the extinction coefficient for illustration of the amount of light a protein absorbed at a particular wavelength, estimated half-life of the protein in vitro and in vivo conditions, stability of the protein in a test tube is given by the instability index, the relative volume occupied by the amino acids is given by the aliphatic index where alanine, valine, isoleucine and leucine have an aliphatic side chain in their structure and Grand average of hydropathicity for a check on hydrophobicity of the construct (Gasteiger et al., 2005).

2.7. Optimization of codon and in silico cloning for bacterial expression:

For the optimal expression of any protein inside a bacterium, the gene sequence should have optimal codons for the bacterial systems. Java Codon Adaptation Tool (JCAT) was used for codon optimization in this study where the vaccine construct's protein sequence was utilized. The tool uses the sequence and optimizes it based on various organisms chosen including prokaryotic organisms. The server also gives information on the CAI-values and GC content of the codon-optimized sequence. The CAI values are determined as defined in Carbone et al. (2003). These properties are vital for obtaining high levels of purification in the bacterial organism. In this study, the sequence was optimized for expression of the construct in *Escherichia coli* K12 bacterial strain. *Escherichia coli* K12 was selected owing to its doubling time and survivability under specific conditions. The codon-optimized gene sequence was then restricted and cloned into the vector pET28a(+) which has been used widely for expression of large quantities of protein and also tags the vaccine construct with a His-tag for affinity purification (Mittal et al., 2020). For in silico cloning, restriction sites BamHI and *Hin*dIII of the vector were exploited and the vaccine design was cloned into it. The BamHI and HindIII restriction enzyme were chosen considering the presence of the sites in pET28a(+) and the use of these enzyme widely.

2.8. Assessment of TLR4 affinity by modelling, docking and molecular dynamics simulation

The 3-dimensional structure of the vaccine protein sequence was obtained by using the I-TASSER server. In brief, I-TASSSER is an automated server that predicts the 3D structure of a protein by running iterative simulations of threading assembly (Yang et al., 2015). Generated models are then evaluated for better quality using the C-score and are ranked based on it. The confidence of model quality is given by C-score. The model which has the highest C-score among the models was selected and the Galaxyrefine server tool was used to improve the model further (Shin et al., 2014). The server uses mild methods and aggressive methods of relaxation of the structural entities to improve the model structure. The final model was then assessed for quality using Verify-3D, ERRAT plot and Ramachandran plot provided by the SAVES server (https://servicesn.mbi.ucla.edu/SAVES/). The errors were rectified accordingly in the structure. The modeled vaccine was docked to the TLR4 receptor (PDB id: 3FXI) by the ClusPro server used for protein-protein docking (Sievers et al., 2011). The final results are given in the form of models and scores. The model which has the lowest binding energy score was then taken forward for molecular dynamics simulations using the open-source software GROMACS v5.1.4 package. The simulation constraints and NVT/NPT thermostat and barostat parameters were set as described in our study previously optimized in our lab (Raj et al., 2019; Sasidharan and Saudagar, 2019). In brief, the system was neutralized using 10 Na^+ ions. NPT and NVT equilibration was carried out with the help of Parinello-Rahman barostat and Modified Berendsen thermostat at 1 bar and 300 K respectively. The simulation was carried out for 20 ns and the trajectory was analyzed for hydrogen bonds between the receptor-vaccine complex to determine the affinity towards each other.

3. Results

3.1. Sequence retrieval of Ebola protein

The amino acid sequences of VP40, VP35, GP, VP30, NP, VP24 and L were taken from the UniProt database in the FASTA format. These proteins that were retrieved have been known to play a critical role in EBOV survival. The VaxiJen scores for the EBOV protein are listed in Table 1. The score showed that among the EBOV protein, VP35, GP, VP30 and VP40 have better scores when compared to NP, VP24 and L

Table 1
VaxiJen score of EBOV protein listed in the decreasing order of their scores. VP35 was predicted to have a high chance of eliciting an antigen response among the 7 proteins on the EBOV genome.

Protein	UniProt ID	Score
VP35	Q05127	0.5025
GP	Q05320	0.4897
VP30	Q05323	0.4683
VP40	Q05128	0.4586
NP	P18272	0.4476
VP24	Q05322	0.4469
L	Q05318	0.4101

proteins. These four proteins were then chosen for the design of CTL and HTL epitopes.

3.2. Prediction of CTL and HTL epitopes

Representative A2, A3 and B7 supertypes (which represent 80% of the population) for the input proteins were obtained by employing the NetCTL 1.2 server. The resultant epitope sequences with a default minimum threshold score were taken up for the screening, where a higher score represents low sensitivity and high specificity for the receptors of adaptive immunity. The CTL epitope with the highest score for all the selected EBOV protein respective, towards each MHC-I supertype, was selected and used for the design of the vaccine (Table 2). Therefore, a total of 12 CTL epitopes for 4 EBOV proteins was chosen to be added to the designed vaccine. Helper T-lymphocyte cells form an important part of the immune response's adaptive limb. By using the reference allele sets which cover more than 99% of the population for almost all proteins, a varied number of epitopes were predicted. The top epitopes were chosen from the IEDB server by taking the percentile cut-off score as 0.40. This indicated that any predicted epitope of more than 0.40 has a higher binding to the receptor. If the IC50 (half-maximal Inhibitory Concentration) value is found to be lower than 50 nM then the epitope has high affinity whereas if the IC 50 value is less than 500 nM then it is considered to be intermediate. Any epitope with an IC50 value greater than 500 nM has a low affinity to the receptor. The overlapping regions of the epitopes were also considered and the final HTL epitopes are listed in Table 3.

3.3. Design and B-cell epitope prediction of the vaccine

All the 12 CTLs and 8 HTLs that were predicted were stitched together with the help of GPGPG and AAY linkers respectively. The linkers have a role in increasing the efficacy of the vaccine by presenting the epitopes in a separate and proper fashion. In addition to this, β-defensin adjuvant consisting of 45 amino acid residues (GIINTLQKYY-CRVRGGRCAVLS CLPKEEQIGKCSTRGRKCCRRKK) was added to the N-terminal to elicit a response from TLRs naïve T cells, dendritic cells and CCR6 immune receptors. The finally designed multi-epitope vaccine was of 392 amino acids that made the final EBOV vaccine (Fig. 2). Designed vaccine included the adjuvant which is coupled with the CTL epitopes with EAAK linker. The presence of a B-cell epitope will solidify the immunization process of the vaccine and therefore BCPred server was chosen for the prediction. 8 B-cell epitopes were predicted in the vaccine construct with score 1 showed that the vaccine has a high affinity towards B-cell receptors (Table 4). A score higher than 0.8 is considered to have a high affinity towards B-cell receptors.

Table 2
CTL epitopes of EBOV proteins VP35, GP, VP30 and VP40 that were chosen after assessment and taken for vaccine design.

Protein	A2	A3	B7
VP35	MVAKYDLLV	RVCAEMVAK	IPRACQKSL
GP	FLYDRLAST	RTFSILNRK	MASENSSAM
VP30	LIMFITAFL	TLCAVMTRK	LPCESSAVV
VP40	FILEAMVNV	FILEAMVNV	TPTGSNGAL

3.4. Antigenicity, allergenicity and physicochemical characterization of vaccine protein

Antigenicity is an important factor that predicts any vaccine construct's success. A designed vaccine should be capable of being immunogenic to be able to stimulate the host cell-mediated immune cells and humoral immune cells and therefore, the event will result in the development of memory cells against EBOV. According to the VaxiJen 2.0 server, the designed vaccine construct had an antigenicity score of 0.4828 with the threshold value capping being set as 0.4. AllerPred results for the allergenicity determination proved the non-allergenic nature of the designed vaccine when administered in humans. The non-allergenic nature and antigenicity of the multi-epitope subunit vaccine was an important eligibility criterion that allowed the designed vaccine to proceed forward for physiological parameter assessment. ProtParam server was used to determine the physiological properties of the designed vaccine. The molecular weight of vaccine protein was found to be 42.22 kDa which fell in the ideal molecular weight range of a vaccine i.e. 40–50 kDa. The theoretical pI was determined to be basic in nature with a value of 9.78 which will be useful in isoelectric focusing and also chromatography based purification. 30 h in mammalian reticulocytes in vitro and more than 20 h and 10 h in yeast and *E. coli* in vivo, respectively was predicted as the estimated half-life of the designed vaccine construct. The instability score of less than 40 is considered to be very stable and in the designed vaccine, the instability index was 38. Hence, the stability of the vaccine protein was confirmed. The aliphatic index score of 84.54 displayed the thermostable nature of the vaccine construct and the Grand average of hydropathicity (GRAVY) value was 0.277 exhibiting the hydrophilic nature.

3.5. Modelling the EBOV vaccine construct

JCAT tool was used to codon optimize the vaccine construct for optimized expression in the *Escherichia coli* K12 bacterial strain. The codon-optimized sequence had a CAI value of 1 and the percentage of GC content which plays a role in the stability of the genetic sequence was 54.25%. The gene sequence was provided with BamHI and *Hin*dIII sites at 5′ and 3′ end respectively. The sequence was restricted sequentially and followed by ligation into the pET28a(+) vector for bacterial purification. The vector map of the vaccine construct is represented in Fig. 3A which was constructed with the help of the SnapGene viewer. His-tag in the vector will help in the purification of the vaccine construct and the bacterial promoter with lac operon will support the expression of the vaccine protein. The protein sequence of the vaccine construct was provided as input to the threading server I-TASSER and the highest C score model was selected for structural assessment. The top ten threading template utilized by I-TASSER were PDB id: 5hasA, 312aA, 5yfpE, 3s88, 3I29, 5yz0C, 3s88, 3fkeA, 3I29B and 5zmmA. The number of residues present in the favorable region before refinement was 79%. The model was further refined with the help of the server GalaxyRefine and the resulting best model was retrieved from the server. All the models were then validated using the SAVES server where the best model qualified the Errat quality and Verify3D threshold tests (Fig. 3B). Ramachandran plot analysis gave 90% of the residues present in the favored region, 7.7% in the allowed region and 2.3% in the outlier region (Fig. 3C). Ramachandran plot is a measure of the residues that are energetically favored. The energetically favorable residues are identified based on the torsional angles of the amino acids. The residues that are found to be in the range of permitted angles are assigned as favored, allowed and disallowed. This modeled construct was taken up for docking and simulation studies

Table 3
HTL epitopes predicted for EBOV proteins taken in two stretches depending on the scores from the IEDB server.

Protein	1st Stretch	Length of 1st Stretch	2nd Stretch	Length of 2nd Stretch
VP35	ITKRVPIFQDAAPPVIHIRR	280–300	AKTISSLNRVCAEMVAKY	125–142
GP	TKRWGFRSGVPPKVVNYE	83–100	QLPRDRFKRTSFFLWVIILFQ	8–28
VP40	DRQSLIMFITAFLNIALQLPC	231–251	TIEDSKLRALLTLCAVMTRKF	161–181
VP30	FDLTALKLITQPLPAATWT	174–192	ADQKTYSFDSTTAAIMLASYTI	101–122

with the TLR4 receptor.

3.6. Understanding the binding of multi-epitope vaccine with TLR4 receptor

Protein-protein docking was performed using the ClusPro server where the multi-epitope vaccine model was docked against the TLR4 receptor. The modeled construct showed good binding with the TLR4 receptor and low binding energies were observed in the docked cluster (Fig. 4A). Binding energy is inversely related to the affinity between the proteins. The vaccine-TLR4 conformation model was then carried forward for protein-protein molecular dynamic simulation for 20 ns. Interestingly, the RMSD analysis found the conformation to be stable for the 20 ns period with very insignificant backbone fluctuation (Fig. 4B). The average of the RMSD value for 20 ns was calculated as 0.9841 ± 0.005 nm. H-bond determination between the proteins will provide the affinity of the vaccine construct to the TLR4 receptor and the same was performed using the trajectory. The average number of H-bonds between them during the 20 ns simulation time was 18H-bonds (Fig. 4C). The electrostatic energy possessed by the receptor-vaccine complex was determined to be $3.2 \times 10^5 \pm 125$ kJ/mol whereas the Lennard Jones energy was calculated as $6.9 \times 10^5 \pm 315$ kJ/mol. The analysis of the trajectories of the receptor-vaccine proved the stability of the vaccine throughout the simulation time and also showed the capability of the vaccine construct to form H-bonds with the TLR4 receptor to elicit an immune response.

4. Discussion

EBOV has been a burden to mankind by affecting the lives of millions worldwide for decades. The vaccines produced currently have severe side-effects and also that the fear of relapse of the virus always looms among the administered patients. Considering the need for a new vaccine that is safe and side-effects free, the research was conducted to initially design and then formulate a multi-epitope subunit vaccine that can elicit an immune response and at the same time, be produced and purified. The experiments in the study were designed in two phases i.e. designing the vaccine and validating the vaccine (Fig. 1).

Table 4
Prediction of the B-cell epitope of the designed vaccine showing the presence of 8 epitopes at different positions of the vaccine sequence. The score of all the predicted epitopes was found to be > 0.8 resulting in high B-cell receptor affinity.

Position	B-Cell Epitope	Score
356	TQPLAATWTGPGPGADQKT	1
206	DAAPPVIHIRSRGPGPGAKT	1
255	VPPKVVNYEGPGPGQLPRDR	1
179	KAAYTPTGSNGALGPGPGIT	1
283	WVIILFQGPGPGDRQSLIMF	1
310	ALQLPCGPGPGTIEDSKLRA	1
332	TLCAVMTRKFGPGPGFDLTA	1
233	CAEMVAKYGPGPGTKRWGFR	1

The first phase i.e. the design phase, consisted of searching the EBOV genome for HTL and CTL epitopes that were ranked based on their ability to cause an immune response. The top hit epitopes were stitched together with AAU and GPGPG linkers wherein a β-defensin linker was also added with EAAK linker to the N-terminal for increased immunogenicity. The final vaccine was found to be of 392 amino acids long and the construct was non-allergenic in nature while capable of eliciting an antigenic response. Once the safety of the vaccine construct was confirmed, the physiological parameters of the vaccine were determined. Previous parameters of an ideal vaccine were compered and they were found to be in line with the results observed in this study (Bohra et al., 2020; Mittal et al., 2020). The molecular weight of the vaccine was calculated as 42.22 kDa and the theoretical pI value was 9.78 for the vaccine protein. The instability index of the protein was 38 which was well within the range of a stable protein. The instability index is a measure of the stability of an ideal vaccine wherein an instability index less than 40 is predicted to be stable and above 40 is considered to be unstable (Guruprasad et al., 1990). The ideal

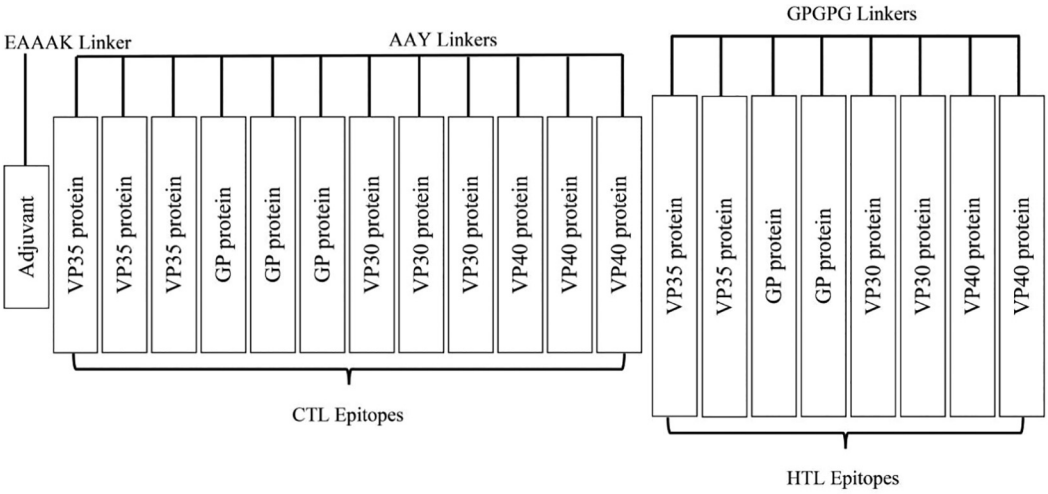

Fig. 2. Design of the multi-epitope vaccine construct: The order in which the HTL and CTL epitopes are arranged are represented in the figure. The N-terminal of the construct is provided with the adjuvant β-defensin followed by the attachment of HTL by GPGPG and CTL by AAY linkers.

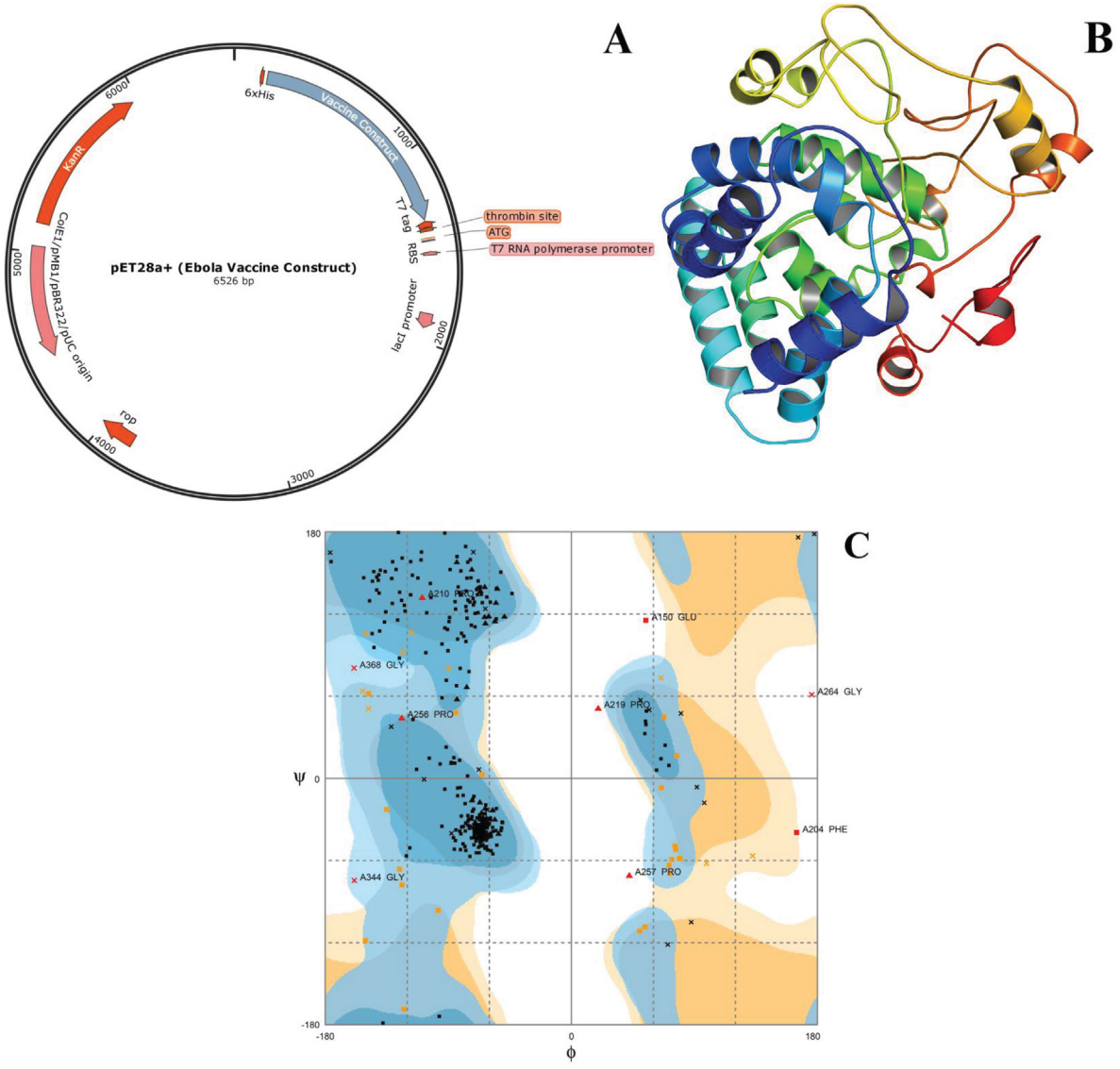

Fig. 3. Modelling and evaluation of designed vaccine: (A) The vector map of the vaccine construct which has been inserted into the pET28a(+) vector by cloning is illustrated in the figure. BamHI and *Hin*dIII restriction enzymes were utilized. The vaccine will be His-tagged at the N-terminal for ease of purification by affinity purification. (B) The modeled 3D secondary structure depiction of the vaccine construct showing a compact and refined model. (C) Ramachandran plot determination of favored, allowed and outlier residues. The determination showed 90% of the residues in the favored region, 7.7% of the residues in the allowed region and only 2.3% of the residues belonged to the outlier region.

parameters of the designed EBOV multi-epitope vaccine, when compared to other vaccine candidate proteins, allowed us to proceed to the validating the vaccine construct.

The validation of the vaccine construct was initiated by codon optimization for obtaining optimal bacterial expression in *E.coli* K12 strain. Consequently, the gene sequence of the vaccine was then restricted and inserted into the vector pET28a(+) by employing the BamHI and *Hin*dIII sites sequentially (Kalita et al., 2020; Kalita et al., 2019). A 3D model of the protein was built using the protein sequence of the vaccine followed by the refinement of the structure. The resultant structure was assessed for errors using the SAVES server and 90% of the amino acid residues were in the favored region in the final model. The vaccine model was then docked with the human TLR4 receptor and the model conformation which had the lowest binding energy was simulated for 20 ns. During the course of 20 ns, the receptor-vaccine complex was stable and did not show any structural instability as seen in RMSD analysis. The H-bond determination between the receptor-vaccine complex gave an average of 18H-bonds. This trajectory analysis showed that the vaccine constructed in this study is capable of binding to the TLR4 receptor of the immune system and elicit an immune response necessary which is a desired property of a vaccine. The study corroborated the docking analysis and therefore has an affinity towards the human TLR4 receptor.

Summing up the study, we have successfully constructed and validated a novel multi-epitope subunit EBOV vaccine that has been found to have no allergenic reaction and is safe for administration by in silico methods. For ease of production and purification, the vaccine has also been cloned into the pET28a(+) vector for expression in a bacterial system. The binding studies of the vaccine also displayed the strong affinity of the designed vaccine construct to the TLR4 receptor and also it's capacity to trigger an immune response. However, the vaccine has to be evaluated by real-time in vitro experiments before the vaccine can be implemented for public usage. In a nutshell, the EBOV multi-epitope sub-unit vaccine has the immense potentiality to be an alternative vaccine and to prevent the EBOV crisis from affecting us anymore.

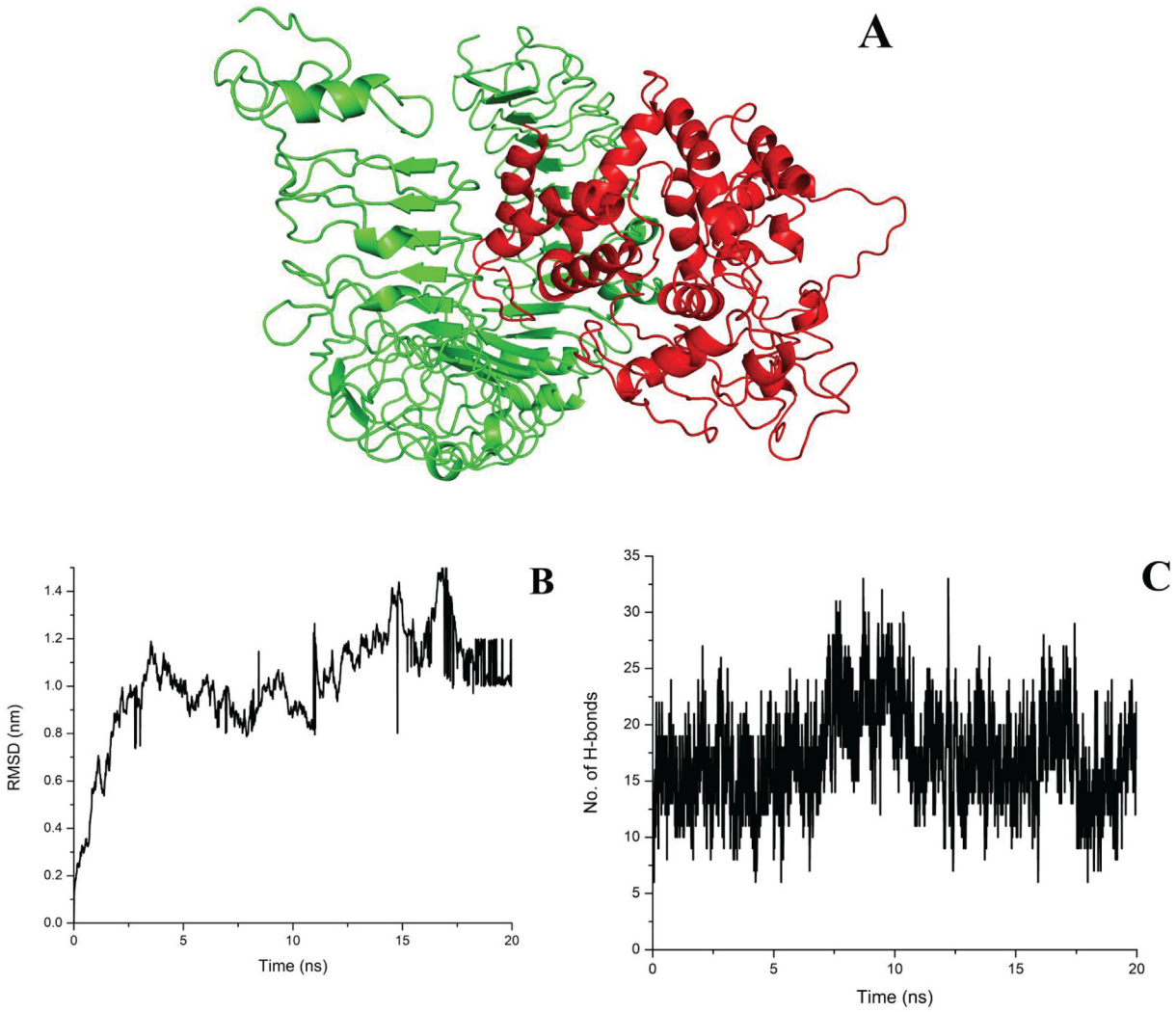

Fig. 4. Protein-Protein docking and simulation of the receptor-vaccine complex: (A) The docked conformation of the EBOV vaccine construct (in red) with the TLR4 receptor (in green) is shown in the fig. (B) The RMSD analysis of the receptor-vaccine complex for a period of 20 ns. 0.9841 ± 0.005 nm was calculated to be the average RMSD value. (C) The H-bond analysis is used to define the affinity of the vaccine to the receptor and in this study, an average of 18H-bonds was observed through the simulation time of 20 ns. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Funding

No funding was obtained in this study.

Declaration of Competing Interest

No potential conflict of interest was reported by the authors.

Acknowledgment

The authors thank Centre for Automation and Instrumentation (CAI), NIT Warangal for extending the computational power and facility required to carry out this work. The author SS acknowledges research fellowship from NIT Warangal.

References

Bausch, D.G., Towner, J.S., Dowell, S.F., Kaducu, F., Lukwiya, M., Sanchez, A., Nichol, S.T., Ksiazek, T.G., Rollin, P.E., 2007. Assessment of the risk of Ebola virus transmission from bodily fluids and fomites. J. Infect. Dis. 196, S142–S147.

Bohra, N., Sasidharan, S., Raj, S., Balaji, S., Saudagar, P., 2020. Utilising capsid proteins of poliovirus to design a multi-epitope based subunit vaccine by immunoinformatics approach. Mol. Simul. 46, 419–428.

Carbone, A., Zinovyev, A., Képes, F., 2003. Codon adaptation index as a measure of dominating codon bias. Bioinformatics 19, 2005–2015.

Doytchinova, I.A., Flower, D.R., 2007. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinform. 8, 4.

Feldmann, H., Sanchez, A., Geisbert, T., 2013. Filoviridae: Marburg and ebola viruses. In: Fields Virology: Sixth Edition. Kluwer Health Adis (ESP), Wolters.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy server. In: The Proteomics Protocols Handbook. Springer, Pp. pp. 571–607.

Genton, C., Cristescu, R., Gatti, S., Levrero, F., Bigot, E., Caillaud, D., Pierre, J.-S., Menard, N., 2012. Recovery potential of a western lowland gorilla population following a major Ebola outbreak: results from a ten year study. PLoS One 7, e37106.

Guruprasad, K., Reddy, B.B., Pandit, M.W., 1990. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng. Des. Sel. 4, 155–161.

Hajighahramani, N., Nezafat, N., Eslami, M., Negahdaripour, M., Rahmatabadi, S.S., Ghasemi, Y., 2017. Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against Staphylococcus aureus. Infect. Genet. Evol. 48, 83–94.

Henao-Restrepo, A.M., Camacho, A., Longini, I.M., Watson, C.H., Edmunds, W.J., Egger, M., Carroll, M.W., Dean, N.E., Diatta, I., Doumbia, M., 2017. Efficacy and effectiveness of an rVSV-vectored vaccine in preventing Ebola virus disease: final results from the Guinea ring vaccination, open-label, cluster-randomised trial (Ebola Ça Suffit!). Lancet 389, 505–518.

Ikram, A., Zaheer, T., Awan, F.M., Obaid, A., Naz, A., Hanif, R., Paracha, R.Z., Ali, A., Naveed, A.K., Janjua, H.A., 2018. Exploring NS3/4A, NS5A and NS5B proteins to design conserved subunit multi-epitope vaccine against HCV utilizing immunoinformatics approaches. Sci. Rep. 8, 16107.

Kalita, P., Lyngdoh, D.L., Padhi, A.K., Shukla, H., Tripathi, T., 2019. Development of multi-epitope driven subunit vaccine against Fasciola gigantica using

immunoinformatics approach. Int. J. Biol. Macromol. 138, 224–233.
Kalita, J., Padhi, A.K., Tripathi, T., 2020. Designing a vaccine for fascioliasis using immunogenic 24 kDa mu-class glutathione s-transferase. Infect. Genet. Evol. 83, 1–10 104352.
Klavinskis, L.S., Whitton, J., Oldstone, M., 1989. Molecularly engineered vaccine which expresses an immunodominant T-cell epitope induces cytotoxic T lymphocytes that confer protection from lethal virus infection. J. Virol. 63, 4311–4316.
Larsen, M.V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., Nielsen, M., 2007. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinform. 8, 424.
Lin, H.H., Ray, S., Tongchusak, S., Reinherz, E.L., Brusic, V., 2008. Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research. BMC Immunol. 9, 8.
Lund, O., Nielsen, M., Kesmir, C., Petersen, A.G., Lundegaard, C., Worning, P., Sylvester-Hvid, C., Lamberth, K., Røder, G., Justesen, S., 2004. Definition of supertypes for HLA molecules using clustering of specificity matrices. Immunogenetics 55, 797–810.
Mittal, A., Sasidharan, S., Raj, S., Balaji, S., Saudagar, P., 2020. Exploring the Zika genome to design a potential multiepitope vaccine using an immunoinformatics approach. Int. J. Pept. Res. Ther. 1–10.
Patronov, A., Doytchinova, I., 2013. T-cell epitope vaccine design by immunoinformatics. Open. Biol. 3, 120139.
Raj, S., Sasidharan, S., Dubey, V.K., Saudagar, P., 2019. Identification of lead molecules against potential drug target protein MAPK4 from L. donovani: an in-silico approach using docking, molecular dynamics and binding free energy calculation. PLoS One 14.
Saha, S., Raghava, G., 2006. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic Acids Res. 34, W202–W209.
Sasidharan, S., Saudagar, P., 2019. Biochemical and structural characterization of tyrosine aminotransferase suggests broad substrate specificity and a two-state folding mechanism in Leishmania donovani. FEBS Open. Bio. 9, 1769–1783.
Schiller, J.T., Lowy, D.R., 2014. Raising expectations for subunit vaccine. J. Infect. Dis. 211, 1373–1375.
Shin, W.-H., Lee, G.R., Heo, L., Lee, H., Seok, C., 2014. Prediction of protein structure and interaction by GALAXY protein modeling programs. Bio. Design. 2, 1–11.
Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. Mol. Syst. Biol. 7, 539.
Towner, J.S., Sealy, T.K., Khristova, M.L., Albariño, C.G., Conlan, S., Reeder, S.A., Quan, P.-L., Lipkin, W.I., Downing, R., Tappero, J.W., 2008. Newly discovered ebola virus associated with hemorrhagic fever outbreak in Uganda. PLoS Path. 4, e1000212.
Vita, R., Overton, J.A., Greenbaum, J.A., Ponomarenko, J., Clark, J.D., Cantrell, J.R., Wheeler, D.K., Gabbard, J.L., Hix, D., Sette, A., 2014. The immune epitope database (IEDB) 3.0. Nucleic Acids Res. 43, D405–D412.
Wang, P., Sidney, J., Dow, C., Mothe, B., Sette, A., Peters, B., 2008. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comp. Biol. 4, e1000048.
Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., 2015. The I-TASSER suite: protein structure and function prediction. Nat. Methods 12, 7.