



Research paper

Computational vaccinology guided design of multi-epitopes subunit vaccine designing against Hantaan virus and its validation through immune simulations



Dawood Ghafoor ^{a,b,*}, Ayesha Kousar ^c, Waqar Ahmed ^d, Soma Khan ^e, Zia Ullah ^f, Nasir Ullah ^g, Shahzeb Khan ^f, Sadia Ahmed ^h, Zafran Khan ^{b,i}, Rida Riaz ^j

^a CAS Key Laboratory of Special Pathogens, Wuhan Institute of Virology, Center for Biosafety Mega-Science, Chinese Academy of Sciences, Wuhan, Hubei, China

^b University of Chinese Academy of Sciences, Beijing 100049, China

^c Institute of Microbiology, University of Veterinary and Animal Sciences, Lahore, Pakistan

^d Department of Microbiology, Abdul Wali Khan University Mardan, KP, Pakistan

^e Department of Chemistry, Islamia College University Peshawar, KP, Pakistan

^f Center for Biotechnology and Microbiology, University of Swat, KP, Pakistan

^g Department of Chemistry, Government Post Graduate Jahangir College, Swat, KP, Pakistan

^h School of Material Science and Engineering, Beijing Institute of Technology, China

ⁱ State Key Laboratory of Respiratory Disease, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China

^j Department of Biomedical Engineering, School of Medical Instrument and Food Engineering, University of Shanghai for Science and Technology, Shanghai, China

ARTICLE INFO

ABSTRACT

Keywords:

Reverse vaccinology
Molecular docking
Hantaan virus
Multi-epitopes subunit vaccine
In-silico cloning
Immune simulations

The Hantaan virus belongs to Bunyaviridae family, an emerging virus that is responsible for hemorrhagic fevers. The virus is distributed worldwide and as of now there is no successful antiviral drug or vaccine developed to protect against the viral infections. Immunization or vaccination is an alternative approach for the protection against viral infections. A cost effective and thermodynamically stable vaccine should be developed to prevent a future possible pandemic. In this study a vaccine candidate was designed against the Hantaan virus, multiple immunoinformatics and reverse vaccinology tools were utilized for the prediction of both B and T cell epitopes for Nucleoprotein, RNA dependent RNA polymerase L and Envelope protein of the Hantaan virus. The individual epitopes were modeled for docking with respective HLAs and a multi-epitopes subunit vaccine candidate was constructed by joining together carefully evaluated B and T cell epitopes with suitable linkers. The vaccine model was evaluated for several physiochemical parameters i.e. Molecular weight, instability index and aliphatic index among the others, followed by 3D modeling of the vaccine for docking with TLR-4. Based on previous studies, Human beta-defensin was liked at the N-terminus of the vaccine sequence as an adjuvant to enhance immunogenicity. The docked complexes of vaccine-TLR-4 were then evaluated for residual interactions. Moreover, to validate final vaccine construct, immune simulations was carried out by C-IMMSIM server. A natural immune response was predicted by the immune simulation analysis. In-silico cloning was carried out using *E. coli* as host resulting in 0.93 CAI value, which suggests that the vaccine construct will attain maximal expression in *E. coli* host. The vaccine designed in this study needs experimental verification to confirm the immunogenicity and efficacy of the vaccine and ultimately used against Hantaan virus associated infections.

1. Introduction

Hantaviruses belong to the worldwide distributed Bunyaviridae family and sustain in environment by residing in their hosts (Jiang et al., 2014). More than 50 species of Hantaviruses have been reported

globally and are classified as endemic according to the geographic locations (Zuo et al., 2011). The pathogenesis of infection is divided into two distinct categories old world Hantaviruses and new world Hantaviruses, former comprises of Seoul (SEOV), Puumala (PUUV), and Dobrava (DOBV), which cause hemorrhagic fever with the renal

* Corresponding author at: CAS Key Laboratory of Special Pathogens, Wuhan Institute of Virology, Center for Biosafety Mega-Science, Chinese Academy of Sciences, Wuhan, Hubei, China.

E-mail address: dawoodghafoor500@gmail.com (D. Ghafoor).

syndrome with the mortality rate from 1% to 15% and the number of reported cases ranges between 100,000 and 150,000 yearly. While the new world Hantaviruses, including Andes Virus (ANDV) and Sin Nombre Virus (SNV) viruses, which causes the fatality rate of 40% but number of reported cases are not frequent, a few hundred cases every year (Kruger et al., 2015). The US National Institute for Allergy and Infectious Disease (NIAID) has categorized HPS-causing Hantaan virus, a Group A pathogens. In general, the transmission of viral particles is mainly through the urine, feces, saliva and aerosols of rodent animal's origin, and rarely from infected animals bites (Brocato and Hooper, 2019).

Recently, the HTNV infection rate has increased in China and Europe, (Dong et al., 2019) however, the nature of infections varies in different regions, depending on the prevalence of viral strains. The genome of Hantaan virus comprises of three segments, designated L (large), M (middle), and S (small). The L segment encodes an RNA-dependent RNA polymerase, and the M segment encodes a transmembrane glycoprotein precursor, which is further processed to create Gn and Gc, while S segment encodes a Nucleocapsid protein (Hall et al., 2010). The primary target site of HTNV and SEOV is kidneys, accompanied by hemorrhagic fever with renal syndrome (HFRS). In North America, infection with the Andes virus (ANDV) and the Sin Nombre virus (SNV) occurs predominantly in the lungs and contributes to Hantaan virus syndrome (HPS) or cardiopulmonary Hantaan virus syndrome (HCPS) with high mortality rates in Europe. Infections with the Puumala virus (PUUV) and Dobrava-Belgrade virus (DOBV) typically results in a milder version of HFRS and epidemic nephropathy (NE) (Echterdiek et al., 2019). It has been proposed that HFRS and HPS infections are almost similar because of the comparable underlying pathology and clinical manifestations such as vascular leakage and thrombocytopenia (Ahn et al., 2000; Clement et al., 2014; Rasmussen et al., 2011). However, in HFRS cases, renal involvement is more prevalent, leading to oliguria and renal failure. In HPS cases, pulmonary symptoms are more pronounced. Besides that many patients experience pulmonary edema (Echterdiek et al., 2019), increasing the endorsement that variations in the host factors may partly account for these differences in the pathogenesis.

Laboratory diagnosis of acute viral infection is dependent on serology and can also be confirmed using RT-PCR detection of viral genome in blood or serum samples (Ahn et al., 2000; Aitichou et al., 2005; Hjelle et al., 1994). Cell culture is also used for the virus isolation and identification of viral particles in tissue samples. Different serological tests, such as ELISA, Rapid Immunoblot Assay (RIBA), and Immune Fluorescent Antibody (IFA) etc. are used frequently however, RT-PCR is the most important method to identify infections (Ahn et al., 2000; Aitichou et al., 2005). Hantaan virus virions are spherical in nature with an average diameter of approximately 80–120 nm (Jonsson et al., 2010). No FDA-authenticated vaccines or therapeutics for HFRS and HPS are available, due to the lack of understanding of viral cycle. The challenge is to control this deadly virus by the availability of therapeutic modalities. Different approaches including antiviral agents, DNA-based vaccines, polyclonal and monoclonal antibodies are considered to neutralize viral particles. Although substantial vaccines have been available but there is no approved vaccines to control the infection so far. Nevertheless, inactivating Hantaan virus vaccines being approved in China and Korea but larger regulatory organizations such as FDA has yet authorized such a vaccine for human practice (Liu et al., 2020). However, several investigations are either in clinical or pre-clinical trials for various Hantaan virus vaccines such as Hantaan/Puumala virus DNA vaccine for prevention of hemorrhagic fever (in Phase2a clinical trials), Andes virus DNA vaccine for the prevention of Hantaan virus pulmonary syndrome, use of PharmaJetStratis (R) needle-free injection delivery device (Phase 1 trials) and combination HTNV and PUUV DNA vaccine (Phase 1 trials) (Liu et al., 2020). Alongside, immunotherapy is an effective and competent strategy for the prevention of infectious diseases. (Brennick et al., 2017; He et al., 2018; Lu

et al., 2017).

A B and T cell multi-epitopes vaccine composed of carefully selected immunogenic peptides is therefore a perfect strategy to prevent and treat tumors or viral infections. (Zhang, 2018). Vaccines stimulate the immune system to produce antibodies and memory cells, which can provide long-lasting protection from particular serotypes (Piccirillo et al., 2014). Immunoinformatics methods are developed to predict and screen suitable epitopes for designing an efficacious multi-epitope vaccine (Zhang, 2018). In-silico research based on immunoinformatics approaches was extended to pinpoint effective epitopes as a possible candidate for vaccine or drug designing (Khan et al., 2019).

In the present study, computational analysis was performed to design an effective vaccine based on epitopes (Sabah et al., 2018). For vaccine construction, CTL, B-cell, and HTL epitopes were predicted and analyzed using specified criteria for selecting potential immunogenic epitopes. The final vaccine construct ability to elicit immune response was verified by further implementation of molecular docking with TLR-4 receptors, immune simulations and codon optimization for maximal expression in *E. coli*. Therefore, this study offers an efficient way to establish a novel Hantaan virus vaccine candidate.

2. Material and methods

2.1. Retrieval of proteins

The complete set of Hantaan virus (Strain 76–18) proteins were downloaded from Uniport (<https://uniprot.org>) ("UniProt: the universal protein knowledgebase," 2016) in the standard FASTA format (UniProt, 2019). The viral proteins were subjected to epitopes prediction in next phase. Fig. 1 presents the overall work flow of the work.

2.2. Immunogenic epitopes prediction

NetCTL-1.2 server was used to predict CTL epitopes for the viral proteins (<http://www.cbs.dtu.dk/services/NetCTL/>) (Kaliampurthi et al., 2018) where the default parameters (0.75 epitope prediction threshold, 0.15C-terminus cleavage score and TAP efficiency score of 0.05) were retained for epitopes prediction (Larsen et al., 2007). These predictions were based on MHC-I binding peptide, proteasomal C-terminal cleavage score and transportation efficiency of antigen-related transporters (TAP). The peptide binding to MHC-I and Proteasomal C-terminal cleavage scores are calculated by an artificial neural network, while the weight matrix calculates TAP efficiency score. Furthermore, the prediction of HTL epitopes against the corresponding reference set of 7 HLAs were carried out using IEDB (Immune Epitope Database) web-server (<http://www.iedb.org/>) (Nielsen and Lund, 2009; Yang et al., 2017). Accuracy of the server is optimal for the reference set of seven HLAs in case of blind epitopes prediction. The IEDB The server allocates IC₅₀ values to the predicted epitopes, which are inversely related to the MHC-II binding affinity. IC₅₀ values below 50 nM display the highest binding affinity for MHC-II, IC₅₀ values below 500 nM suggest midrange affinity, while IC₅₀ values below 5000 nM are the lowest binding affinity.

2.3. Interferon gamma and vaccine

For the prediction of interferon-gamma inducing MHC-II (HTL) epitopes, IFN-epitope Web Server (<http://crdd.osdd.net/raghava/ifnepitope/>) (Kaliampurthi et al., 2019) was used. This server uses Motif and SVM hybrid algorithms to predict interferon inducing property of the epitopes and ssigns scores based on the SVM algorithm. The HTL epitopes capable of inducing IFN-gamma response were later used for vaccine construction.

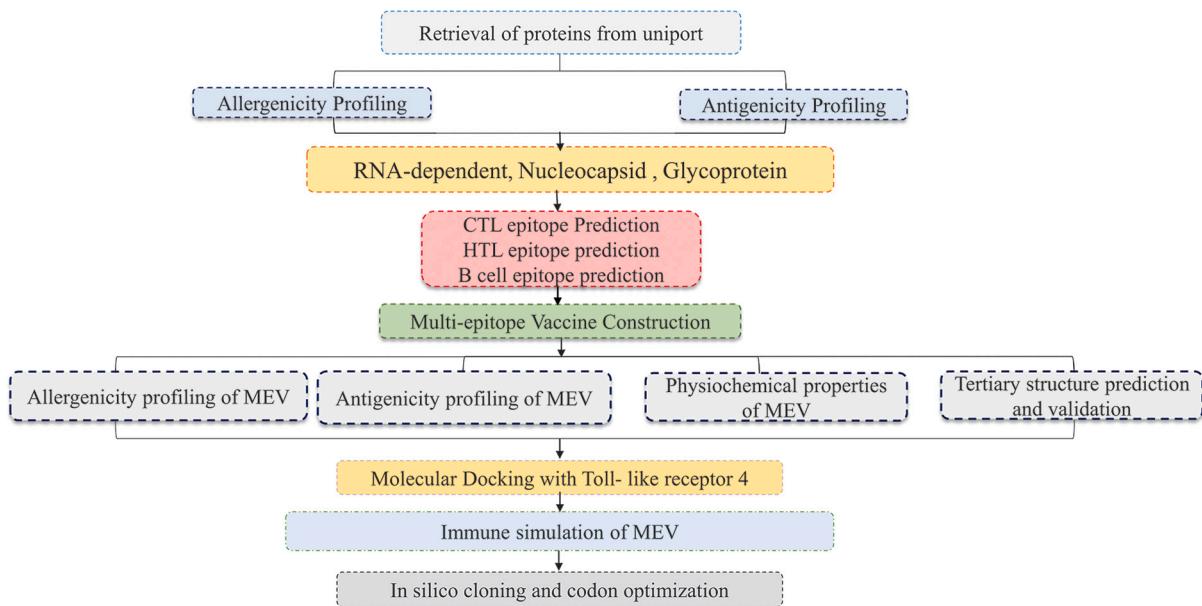


Fig. 1. Methodological workflow of the study, epitopes prediction for antigenic viral proteins followed by vaccine construction, modeling and immune simulations.

2.4. Prediction of linear B-cell epitopes

B-cells have been known to be actively involved in the production of antibody, humoral immunity and in the development of memory cells against future pathogens. Humoral immunity responds to linear B-cell epitopes that are specific epitopes. The B cell epitopes for vaccine construction was predicted using the BCpreds server (<http://ailab-projcts1.ist.psu.edu:8080/bcpred/>) (Chen et al., 2007). The server uses the kernel technique to predict 20-mer linear B-cell epitopes. BCpreds uses a support vector (SVM) algorithm for the prediction of B-cell (linear) epitopes. High scoring B-cell epitope per protein was selected for vaccine construction.

2.5. Designing of multi-epitope vaccine

The selected epitopes were joined together using multiple linkers namely, GPGPG linkers and AAY linkers for HTL epitopes and CTL epitopes, respectively (Hajighahramani et al., 2017; Mittal et al., 2020). The EAAAK linker was used to attach the adjuvant (Mammalian Beta Defensin) to the vaccine N terminal, which amplifies its immunogenicity. These linkers were used for constructing MEVC because they helps to facilitate the epitope display hence allowing effectual immune processes. Further, these linkers also helps to keep the epitopes separated and prevent them from folding (Ahmad et al., 2020; Gul et al., 2020; Khan et al., 2020).

2.6. Physicochemical properties evaluation

Algpred server (<http://crdd.osdd.net/raghava/algpred/>) (Saha and Raghava, 2006) was used for allergenicity prediction whereas Vaxijen server (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) (Doytchinova and Flower, 2007) was used to verify the antigenic potential of the MEVC. Furthermore, an online server ProtParam (<http://web.expasy.org/protparam/>) (Gasteiger et al., 2005) was used to evaluate the different physicochemical properties of the vaccine candidate. Amino acid composition, molecular weight, aliphatic index, in-vitro half-life, in-vivo half-life, index of instability, theoretical PI and GRAVY were determined by the server.

2.7. Secondary and tertiary structure prediction

The vaccine sequence's secondary structure was predicted by PSI-PREDV3.3 server (<http://bioinf.cs.ucl.ac.uk/psipred/>) (McGuffin et al., 2000) whereas Robetta server (<http://robetta.bakerlab.org>) (Kim et al., 2004) was used for tertiary structure modeing with high accuracy. Robetta uses a comparative modeling technique for structure generation to recognize, the template structure of given amino acid sequence is identified utilizing PSI-BLAST, BLAST, FFAS03 or 3D-Jury. In case there is no template identified with reasonable similarity, de novo Rosetta fragment insertion method is preferred. The 3D structure was further improved by using GalaxyRefine online server (<http://galaxy.seoklab.org/>) (Heo et al., 2013). The server uses CASP10 to refine the 3D structure of the query. The protein side chains are re-built by using CASP10 technique, followed by repackaging and using 3D-structure simulations for relaxation. Galaxyrefine improved 3D structure structural and global quality. YASARA software was used for energy reduction and structure correction (Krieger and Vriend, 2014).

2.8. Validation of tertiary structure

Validation of the tertiary structure is a very significant phase in the assessment of the tertiary structure. In this analysis the protein 3D structure was validated by three servers. ProSA-web-server, (Wiederstein and Sippl, 2007) calculates the Z-score (quality score) for a given 3D structure (<https://prosa.services.came.sbg.ac.at/prosa.php>). ProSA-web server uses 3D molecular viewer to highlight and encourage the identification of problematic portion display in a quality plot score. An ERRAT online server (<http://services.mbi.ucla.edu/ERRAT/>) (Colovos and Yeates, 1993) was used to calculate the non-bonded links present in the 3D structure. The PROCHECK server (<https://services.mbi.ucla.edu/PROCHECK/>) (Laskowski et al., 2006) was used for the analysis of the Ramachandran plot.

2.9. Molecular docking of peptides and vaccine

The 3D structure of epitopes were generated using PEP-FOLD3 (<https://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms:PEP-FOLD3>) (Lamiable et al., 2016), whereas the 3D models of HLAs were retrieved from Protein Databank, whereas the HLAs for which crystallographic information was absent were modeled using Phyre2

server (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/>) (Kelley et al., 2015). To evaluate the peptide-MHC binding, HLA-peptide docking of the selected epitopes was carried out using Hawkdock server (<http://cadd.zju.edu.cn/hawkdock/>) (Weng et al., 2019). Best docking poses were selected based on MM-GBSA scoring function of Hawkdock server. The MM-GBSA technique combines the molecular mechanical energies with the scale solvent approaches. The entropy term is eliminated because of convergence problems in some cases, and it cannot be calculated.

The immune receptors present in immune cells play an essential role in developing the correct immune response. Molecular docking of vaccine with the receptor [TLR-4 (PDB ID:3fxi)] was carried out using Cluspro server (<https://cluspro.bu.edu/login.php>) (Kozakov et al., 2017) to determine the tendency of the vaccine to induce an immunological response. Since the TLR-4 acts as a dimer therefore interactions with both the chains was evaluated. The residual interaction of MEV-TLR-4 was evaluated using PDBsum sever (<http://www.ebi.ac.uk/theronton-srv/databases/cgi-bin/pdbsum>) (Laskowski et al., 2018).

2.10. Immune simulation

C-ImmSim is an online server (Rapin et al., 2011) that uses an agent-based modeling approach to estimate a foreign antigen particle's immune system impact. The server uses the PSSM method to represent the immune response to the antigen. Upon the injection of vaccines, Antibodies, cytokines and interferon production were calculated. The recommended dose interval for majority of the vaccines that are currently in use is 4 weeks from dose 1 to dose 2 (Castiglione et al., 2012). All parameters were set at default with time steps set at 1, and 84, where each time step is equal to 8 h. Therefore, two injections were given four weeks apart. The Webserver also predicts Th1 and Th2 responses. Simpson Index or D were constructed by using default parameters.

2.11. Codon optimization and in-silico cloning of vaccine sequence

The vaccine was effectively exposed in a host (*E. coli* strain K12). JCAT (Grote et al., 2005) was used to reverse translation and optimize codon (Java Codon Adaption tool). The expression of the Hantaan virus genome is different from vector genome expression. Optimization of codon ensures optimal vaccine expression in the vector. Three additional options were chosen to attain the desired result, such as the prokaryote ribosome binding site, cleavage limit enzymes and the rho-independent transcript termination. JCat production contains an index of codon adaptation (CAI) and GC content is used to ensure high levels of protein expression (Khan et al., 2019). The desired gene is cloned in *E. coli* pET-28a(S) were applied to the sequence terminal C and N with two restriction sites *Nde*I and *Xho*I, respectively. Finally, a pET-28a (β) has been inserted for optimizing vaccine expression in the adapted sequence with the limitations site (Ahmad et al., 2020).

3. Results

3.1. Retrieval proteins

The Hantaan virus proteins, nucleoprotein (Uniprot ID: P05133), RNA-directed RNA polymerase L (Uniprot ID: P23456), envelope polyprotein (Uniprot ID: P08668) amino acid sequences were downloaded from UniprotKB in FASTA format. The retrieved sequences were used for the prediction of B and T cell epitopes for multi-epitope subunit vaccine construction.

3.2. Prediction of immunogenic epitopes

A set of 8, 73 and 26 CTL epitopes was predicted using the NetCTL1.2 server for the Nucleoprotein, RNA-directed RNA polymerase L, and Envelope polyprotein of Hantaan virus, respectively (Table S1).

Furthermore, HTL epitopes corresponding to a reference set of 7 human alleles were predicted by MHC-II module of IEDB MHC-II. HTL epitopes with low percentile rank and CTL epitopes having a high combined scores correspond to high binding affinity for respective MHC receptors were selected for vaccine construction. Moreover, non-overlapping, non-allergenic (score less than -0.4) and IFN inducing properties were considered before the selection of HTL epitopes. Algpred and IFNepitope servers were used for prediction of non-allergen and IFN inducing ability of HTL epitopes. B cell epitopes are an integral part of an efficient MEVC, BCpreds server predicted Linear B cell epitopes for the selected Hantaan virus proteins. Overall, 2 CTL, 3 HTL and 1 B cell epitopes per protein were selected for multi-epitopes vaccine construction.

3.3. Construction of multi-epitope subunit vaccine

The final vaccine sequence was constructed by joining together selected CTL, HTL and B cell epitopes using AAY, GPGPG and KK linkers, respectively. Addition of adjuvant enhance the immunogenicity of multi-epitope vaccine. The non-toxic human beta defensin-2 (hBD-2) was attached as an adjuvant with the help of EAAAK linker at the N-terminus of the vaccine sequence. The linkers enhances epitopes presentation on the receptors and prevent formation of junctional epitopes or neoepitopes. The final vaccine sequence was of 365 amino acids as depicted in Fig. 2, whereas the selected epitopes are shown in Table 1.

3.4. Prediction of physiochemical properties

AlgPred server predicted allergenicity score of -1.15 at the default threshold -0.4 and validated the vaccine's non-allergen nature. The antigenic potential was scored 0.52 by the Vaxijen server, confirming the antigenic nature of MEVC at 0.4 threshold with virus selected as a model organism. ProtParam server calculated nine physiochemical properties, molecular weight 41.9 KDa was within the ideal range of 30–70 KDa, theoretical protrusion index (PI) 9.92 (indicating the basic nature of vaccine), in-vivo half-life in *E. coli* > 10 h. The stability of vaccine was determined by instability index and this value was 36.28 (<40 is classified stable), whereas, the aliphatic index 83.61 identified MEVC thermostable nature. The grand average of hydropathy value of the vaccine was -0.716. The negative GRAVY values indicate that the protein is hydrophilic and hence can lead to improved interactions with the nearby water molecules.

3.5. Secondary and tertiary structure prediction

PSIPRED server was used to predict secondary structure for final vaccine construct. The results obtained from the server showing the presence of 62.73% alpha-helix, 26.02% coils, and 11.25% beta strands in the vaccine structure as shown in Fig. 3. Robetta predicted the 3D model of the candidate MEVC using the comparative modeling approach. In order to select the best model, ProSA-web, ERRAT and Ramachandran plot evaluation was performed. The Z-score of the preliminary input model calculated by ProSA-web was found to be -6.11, which endorses the range commonly found in the native proteins having similar size. The structure was further subjected to ERRAT validation which revealed the overall quality factor as 91.922. In addition, the Ramachandran plot analysis was carried out using Procheck server, which revealed that 95.1%, 3.6%, and 0.6% of the residues in the primary model were present in the favored, allowed, and outlier regions. The final analysis revealed that model 3 with a confidence score of 0.02 predicted by Robetta was the best model and was subjected to further analysis. Galaxyrefine server carried out the structural refinement of model 3. The server provided 5 models, model 2 was selected on several parameters namely, GDT-HA (0.9864), RMSD (0.307), MolProbity (1.762), Clash score (12.4), Poor rotamers (1.0) and Ramachandran plot (97.1), the 3D model is shown in Fig. 2b.

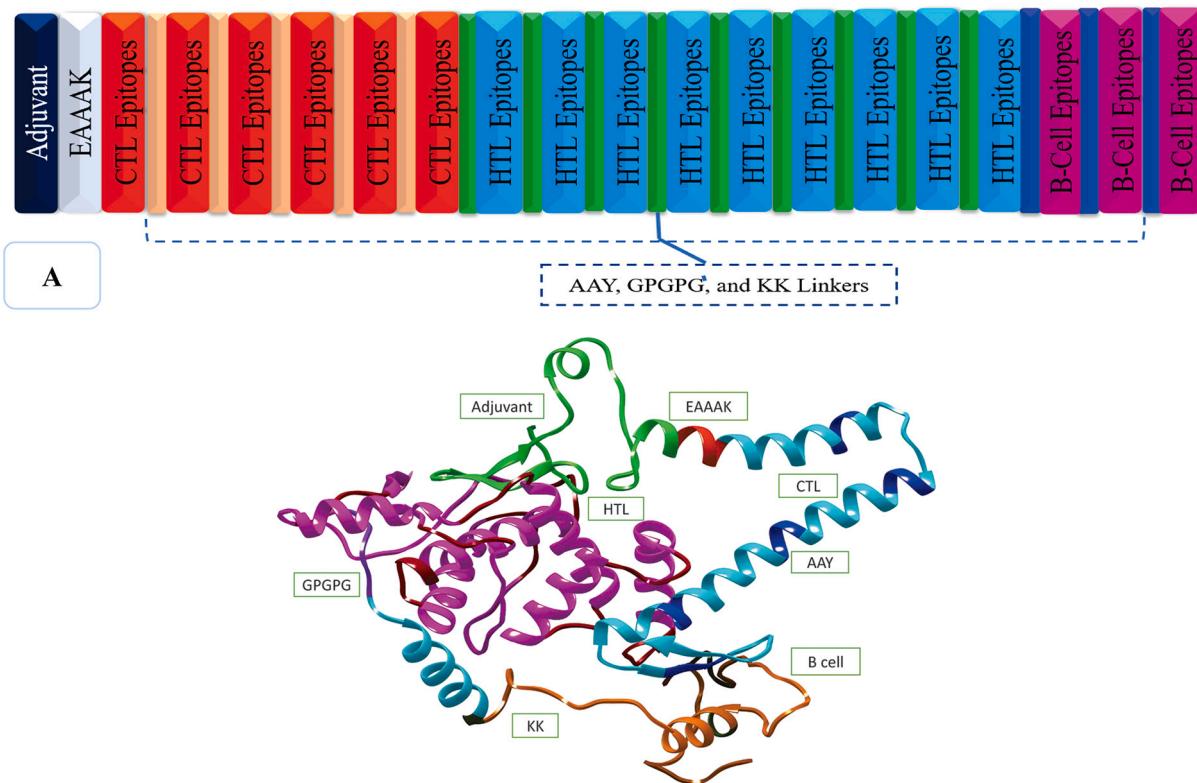


Fig. 2. Final MEVC (A) Graphical representation of epitopes arrangement in final vaccine construct (B) 3D model of MEVC with each component depicted in different color.

Table 1

Final epitopes for vaccine construct (A) Shortlisted CTL epitopes per protein with respective HLA (B) Final HTL epitopes with corresponding HLAs, Non-Allergen score and IFN inducing capacity (C) B cell epitopes.

Selected CTL Epitopes (A)						
Protein	ID	Peptide	Affinity	Cleavage	Tap score	Combine score
Nucleoprotein	HLA-A*01:01	LTDREGVAV	0.2776	0.4248	0.131	1.2491
	HLA-A*30:02	LKERSMLSY	0.2043	0.8132	2.811	1.1298
RDRP L	HLA-A*01:01	HSDDALFIY	0.7902	0.9351	2.675	3.629
	HLA-A*30:02	ATATWFQYY	0.7451	0.9772	3.073	3.4637
Envelope	HLA-B*57:01	LTSSSKYTY	0.666	0.9707	2.957	3.1213
	HLA-A*01:01	DTENKVQGY	0.6118	0.9515	2.584	2.8694

Selected HTL epitopes (B)					
Proteins	Allele	Peptide	Score	Allergenicity	IFN
Nucleoprotein	HLA-DRB5*01:01	PILLKALYMLTTRGR	1.4	-0.49	Yes
	HLA-DRB4*01:01	PELRTLAQSLIDVKV	1.5	-0.98	Yes
	HLA-DRB1*03:01	FFSILQDMRNTIMAS	2.5	-0.46	Yes
RDRP L	HLA-DRB1*15:01	LEPVRLIKSWVSRL	0.04	-0.91	Yes
	HLA-DRB1*03:01	HFKFIRMKRKLLMYVSA	0.35	-0.56	Yes
	HLA-DRB1*15:01	GNVILFILPSKSLEV	0.46	-0.53	Yes
Envelope	HLA-DRB4*01:01	YIFTMWIFLLVLESI	1.3	-0.4	Yes
	HLA-DRB1*15:01	VLCVFLLFSLVLLSI	1.8	-0.4	Yes
	HLA-DRB1*15:01	WVPVLTLRNVYDMKI	2.7	-0.4	Yes

Selected B cell epitopes (C)			
Proteins	Position	Peptide	Score
Nucleoprotein	182	PNAQSSMKAEEITPGRYRTA	0.997
RDRP L	214	KFNISTHKSQPYIPDYKG	0.996
Envelope	293	SIVGPANAKVPHSASSDTLS	0.996

3.6. Molecular docking of peptides and vaccine

A vaccine should express an excellent binding affinity with the host's

immune receptors to stimulate proper immune responses. The 3D structure of epitopes was generated using PEP-FOLD3, whereas the 3D models of HLAs were retrieved from Protein Databank (PDB), whereas

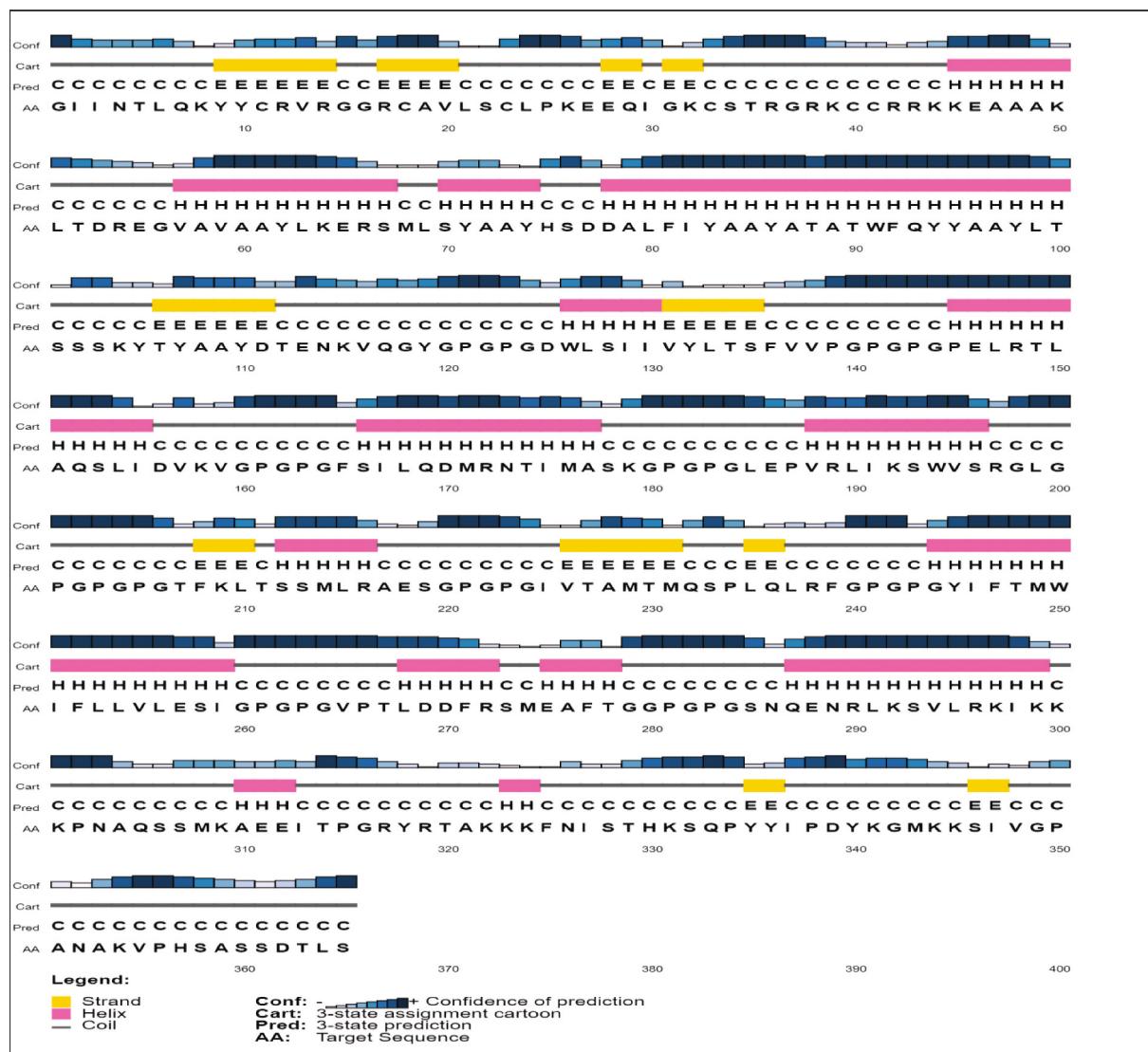


Fig. 3. Graphical depiction of the secondary structure procured for the multi-epitope subunit vaccine construct showing alpha-helix (62.73%), beta strands (11.25%) and coils (26.02%).

the HLAs for which crystallographic information was absent were modeled using Phyre2. To evaluate the peptide-MHC binding, HLA-peptide docking of the selected epitopes was carried out using Hawk-dock server. The MM-GBSA technique combines the molecular mechanical energies with the scale solvent approaches (Geheden and Ryde, 2015). The entropy term is eliminated because of convergence problems in some cases, and it cannot be calculated. The total binding free energy for LTDREGVAV-HLA-A*01:01 was reported to be -29.35 kcal/mol, -10.72 kcal/mol for LKERSMLSY-HLA-A*30:02, -41.55 kcal/mol for HSDDALFIY-HLA-A*01:01, -38.56 kcal/mol for ATATWFQYY-HLA-A*30:02, -34.58 kcal/mol for LTSSSKYTY-HLA-B*57:01, and -38.83 kcal/mol for DTENKVQGY-HLA-A*01:01. The peptide-HLA docking complexes are shown in Fig. 4B, whereas the binding free energy calculations are given in Fig. 5.

Furthermore, after construction and refinement, the best model of MEV was docked against the TLR-4 (Fig. 2C). Cluspro server was used for the MEV-TLR-4 docking. The residual interaction of MEV-TLR-4 complex revealed 8 salt bridges and 52 hydrogen bonds of MEV with the chain A of the TLR-4. The residual interaction was evaluated using PDBsum sever (Table S2). Graphical illustration is provided in Fig. 4A. Overall, these results show that the MEVC significantly interacts with the immune receptor TLR-4.

3.7. Immune simulations

The host immune simulation was performed to determine the immune response towards the vaccine upon the injection of dose 1 and dose 2, 4 weeks apart. The simulated immune system response was compatible with real time responses of the immune system (Fig. 6B). The secondary response (dose 2) was higher as compared to primary response. Elevated concentrations of IgM antibody was observed at primary immune response. In both the primary and secondary reactions, the increased concentrations of immunoglobulin (i.e., IgM, IgG + IgM, and IgG1 + IgG2, antibodies) was observed with concomitant reduction in antigen concentrations. Strong cytokine and interleukin response have also been observed. Higher level of IL-2 and IFN- γ was evident in the immune response to the vaccine. Besides the cytokine and interleukin response, a lower Simpson index (D) can be seen, indicating greater diversity (Fig. 6B). The immune simulation profile demonstrates memory development and an immune protection against Hantaan virus.

3.8. Codon optimization of final vaccine constructs

To optimize codons Java Codon Adaptation Tool was utilized to

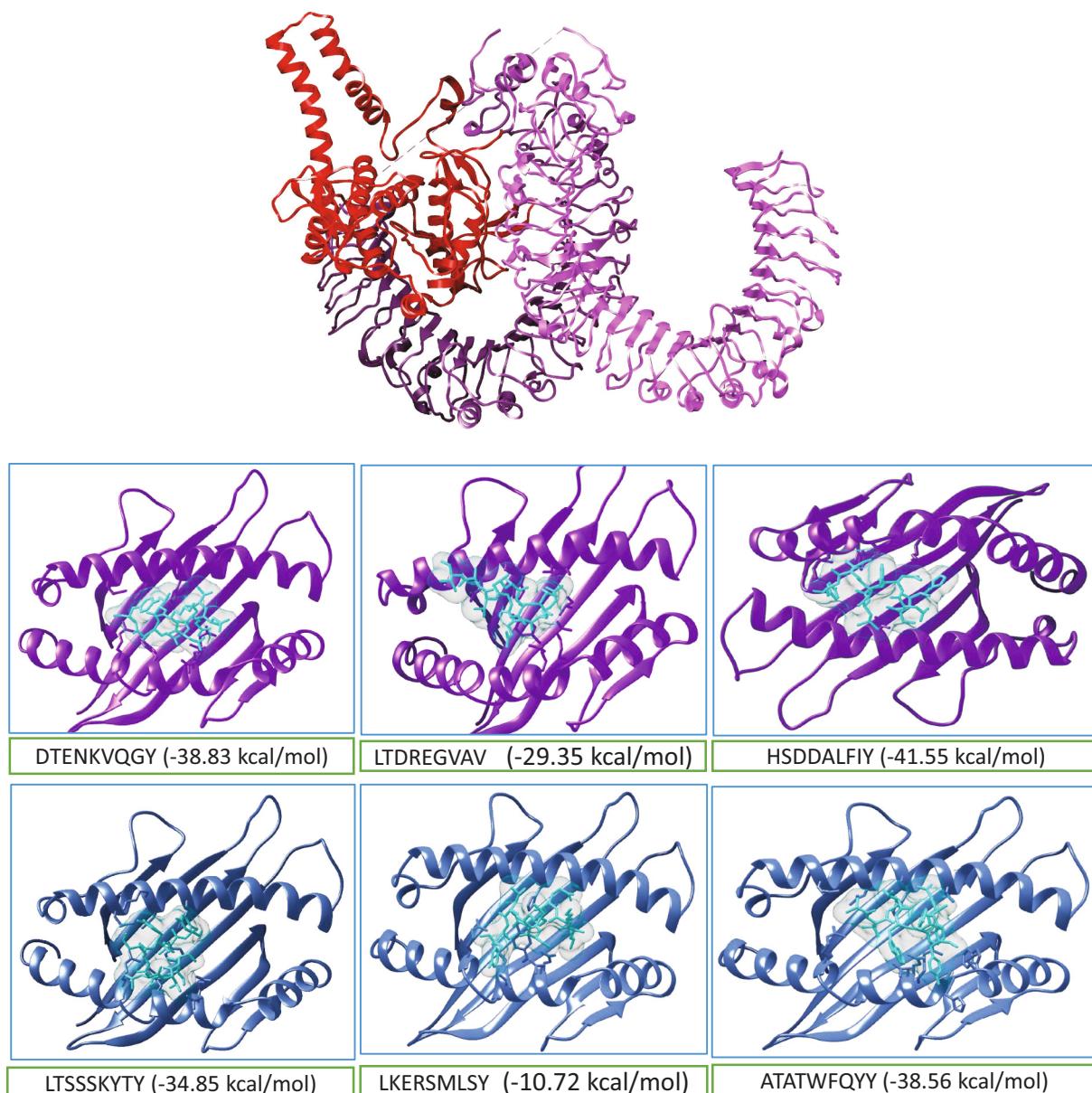


Fig. 4. TLR-4 (PDB ID: 3fxi)-vaccine docked complex. (A) The residual interaction of MEV-TLR-4 residues, where the Chain A is TLR-4 and Chain B is the MEV (B) HLA-peptide docking, where the HLAs are depicted in Purple and peptides in Cyan color (C) The MEV-TLR-4 docking complex, where TLR-4 (Receptor) is shown in Magenta while Red color represents the MEV as a ligand in the complex procured from Cluspro server.

optimize codons to be used in *E. coli* (strain K12) and attain maximum expression of vaccine protein. Optimized sequence was 1095 nucleotides where the Codon Adaptation Index (CAI) was calculated to be 0.93 and GC content was 52% which indicated that the vaccine protein in *E. coli* could be expressed well. The ideal GC content range should be between 30 and 70%. Finally, the restriction clone was formed by inserting adapted codon sequences into pET28a (+) vector (Fig. 6A).

4. Discussion

The Nucleoprotein, RDRP-L and Envelope proteins were selected based on their role in mediating the virus' entrance into the host cells and subsequently in packaging. The structural protein was shown to be the host immune response inducer. Immunization is one of the most successful and secure ways to improve public health efficiently, easily and cost-effectively, as well as the safest way to treat infectious diseases

in today's society. Modern research focuses primarily on sub-unit vaccines compared with whole-pathogen vaccines, as sub-unit vaccines contain unique immunogenic components. In recent years, genomic and proteomic information on Hantaan virus and many other pathogenic microorganisms is available online in huge quantities. The related data collected information helps to find new antigenic targets using computational strategies. Here, the assessment of the immune response of Hantaan virus protein inducing epitopes and also the simple understanding of their diverse associations with host MHC alleles and lymphocytes have enormous immunological values that can be used to establish a vaccine candidate against Hantavirus. In this scientific study, three Hantaan virus proteins were obtained for further analysis. Hantaan virus proteins were subjected to BLASTp against human proteomes to ensure MEV B does not cause an autoimmune response. The viral proteins were predicted to have B and T cell epitopes. T cells effectively recognize both MHC molecules as the T cell receptors' surface protein's

Chain A Chain B

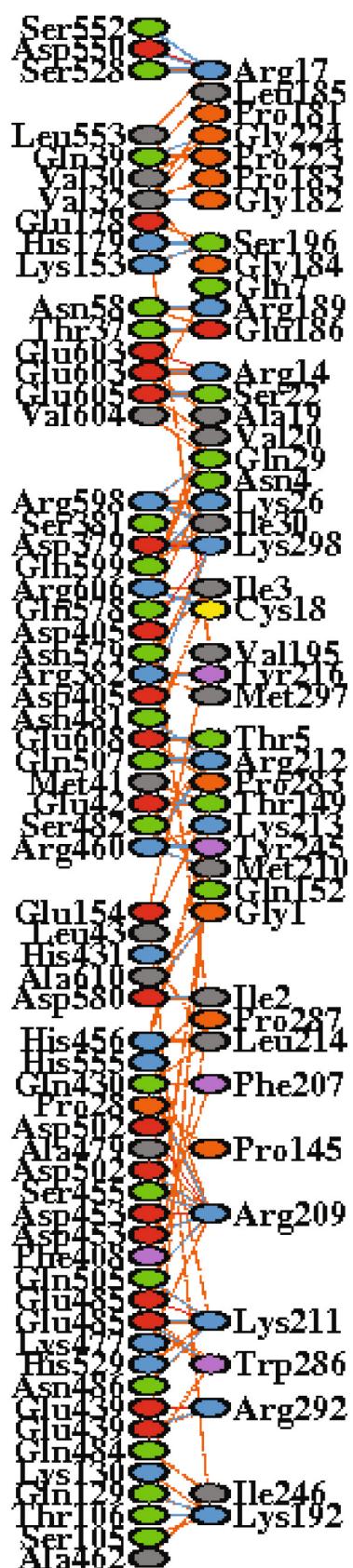


Fig. 4. (continued).

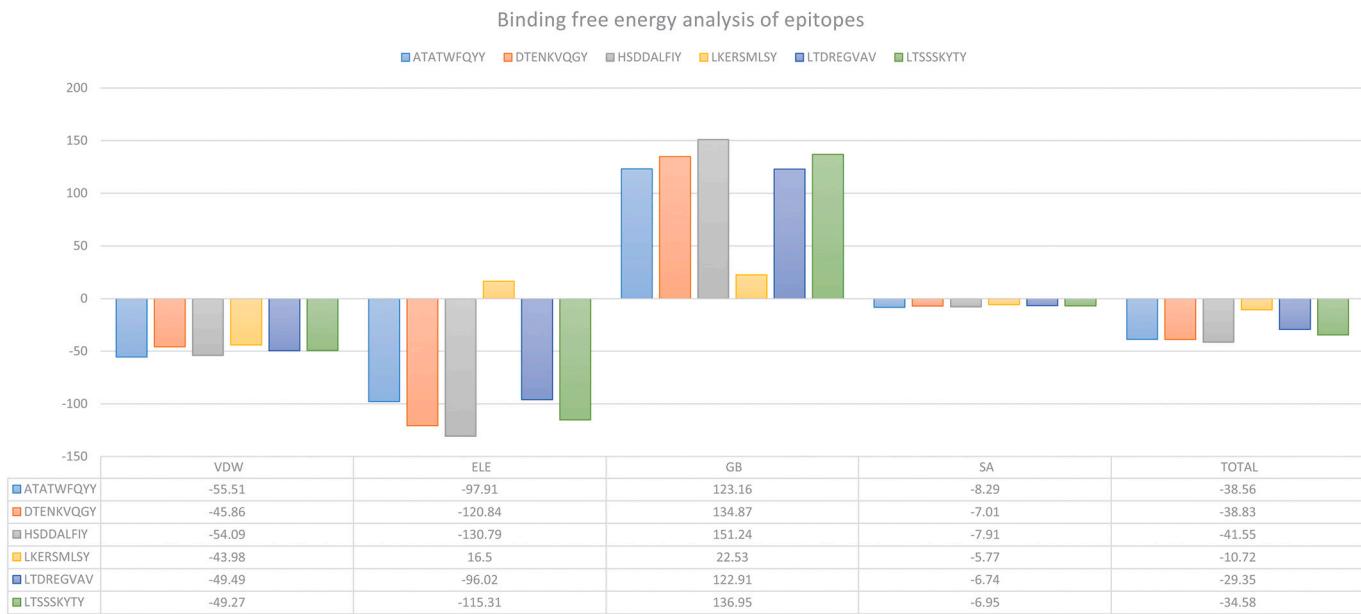


Fig. 5. The total binding free energy of HLA-peptide complexes, ATATWFQYY-HLA-A*30:02 (Light Blue), DTENKVQGY-HLA-A*01:01 (Red), HSDDALFIY-HLA-A*01:01 (Grey), LKERSMSLY-HLA-A*30:02 (Yellow), LTDREGVAV- HLA-A*01:01 (Blue), and LTSSSKYTY-HLA-B*57:01 (Green).

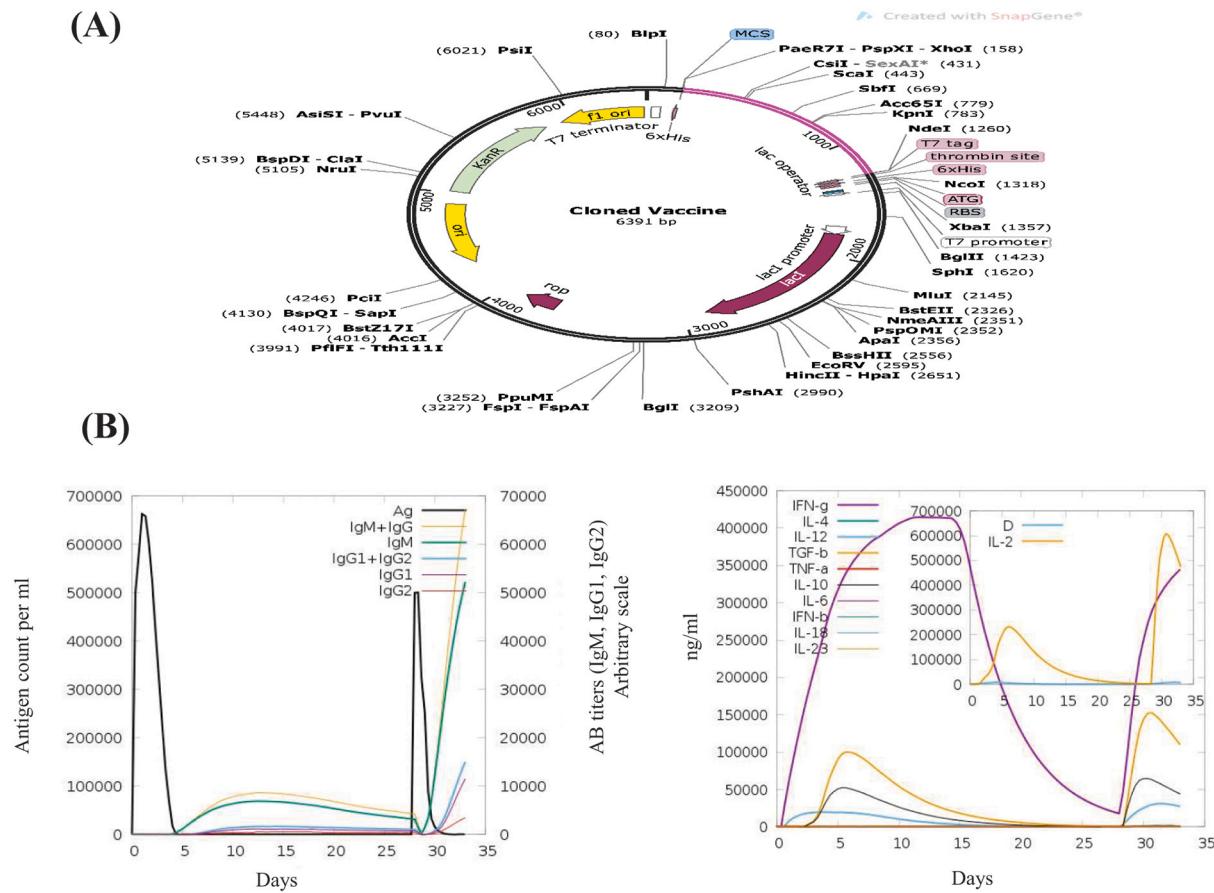


Fig. 6. (A) In silico restriction cloning of MEV into pET28a (+) expression vector, the Magenta line represents the MEV in the black circle (vector) (B) Show the immune response and different different host immune factors secretion upon the injection of 2 doses 4 weeks apart.

known antigenic cells (TCR). These antigen-presenting molecules (MHCs) are classes I or II. MHC class I molecules are present in almost all nucleated cells, which either accurately contain endogenous or

cytosolically scanned proteins or antigens which represent cytotoxic T lymphocytes T (CTLs). MHC class II molecules typically include exogenous antigens, usually surface proteins of the pathogen, processed

through endocytic pathways and often presented to help lymphocytes T or CD4 + T cells. Our final vaccine protein comprises several MHC class I, MHC class II, and B-cell linear epitopes in the immunoinformatics analysis based on physicochemical characteristics and structural properties in particular. The final multi-epitope subunit vaccine protein's antigenic and non-allergenic property makes it a potent vaccine. The molecular weight for the vaccine protein was 41.9 kDa. Simply demonstrating an average value to construct MEVC, theoretical pI was 9.92 suggesting basicity of vaccine protein, the aliphatic index shows that aliphatic side chains actually inhabit the protein and the instability index approves that the vaccine protein is quite stable. All these properties suggest that it can be a thermally stable protein.

For the evaluation of the secondary vaccine structure, PSIPRED V3.3 is a common and highly effective technique used in this study. Furthermore, the 3D structure acquired by homology modeling work provides sufficient knowledge on the spatial arrangement of certain important protein residues, as well as pretty good guidance in the study of normal protein activity, dynamics, ligand interaction, and other proteins. In the final vaccine model, structural validation methods were used to identify errors. The key Ramachandran plot shows that many of the residues are tightly clustered in the most desired region with only few residues in the prohibited region, suggesting that the overall reliability of the global model is satisfactory. Therefore, to know the response of the immune system, MEV-TLR-4 molecular docking was carried out. The immune system response as predicted by immune simulation analysis is consistent with a typical immune response, given the overall increase in immune response to subsequent exposures after 2 doses 4 weeks apart. The results revealed the release of high concentrations of IgM + IgG as a primary response, followed by IgM, IgG1 + IgG2 and IgG1 at the primary and secondary stages with concomitant reduction of antigen. Another interesting conclusion is that after administering the first dose, concentrations of interferon-gamma (IFN- γ) and interleukin-2 (IL-2) were elevated and were sustained at peaks after recurrent antigen contact. These findings depict an elevated concentration of TH cell and therefore effective antibodies production supporting a proper humoral response. Moreover, the Simpson Index (D) proposes a likelihood of diverse immune response for clonal specificity analysis. The overall immune simulation analysis validates the vaccine potential to elicit a proper immune response offering protection against Hantavirus. To achieve a high level of recombinant vaccine protein expression in *E. coli* (strain K12), codon optimization was performed for improving the effectiveness of transcription and translation. This was assessed by analysing the index of codon adaptation (CAI) and the DNA sequence's total GC content. A recombinant protein solubility over-expressed in *E. coli* host is one of the basic criteria of several biochemical and functional experimental studies. In an over-expressed condition, our protein vaccine demonstrates an appropriate percentage of solubility, corresponds to improve strength of proteins. In the experimental animal model, previously engineered and tested vaccines showed a strong immunogenic response, but due to the complex nature of human immunopathology, none of them showed the same response when tested in humans. Therefore, we have implemented new immunoinformatic methods in this research to design a possible, safe and immunogenic subunit vaccine that can regulate the infection of Hantavirus.

5. Conclusion

In this study, in-silico approaches were applied to design an effective multi-epitope vaccine against Hantavirus. For the designing of a vaccine, B and T cell epitopes were predicted and fused by suitable linkers. Tertiary vaccine structure was predicted and validated to check the functionality of the vaccine. The physicochemical characterization of the vaccine was performed to predict their allergenicity behavior, antigenicity, stability, and molecular weight. The vaccine has been docked with the TLR-4 to verify the receptor affinity of the vaccine. Finally, the protein was inversely used for codon optimization, translated to ensure

maximum vaccine expression in the host (*E. coli*). vivo trials are required to ensure the effectiveness of constructed vaccine. This study can be helpful in designing of an effective vaccine against Hantavirus.

Funding information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author's contribution

Dawood Ghafoor, Ayesha Kousar, Waqar Ahmed, Irfan Ahmad, and Shah Zeb Khan conceptualized the methodology. Dawood Ghafoor, Ayesha Kousar, Waqar Ahmed, Soma Khan, Zia Ullah, Nasir Ullah, and Shah Zeb Khan did the analysis. Dawood Ghafoor, Sadia Ahmed, Zafran Khan, Rida Riaz wrote the manuscript. Dawood Ghafoor and Shah Zeb Khan supervised the study. All the authors approved the manuscript.

Declaration of competing interest

The authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.meegid.2021.104950>.

References

- Ahmad, I., Ali, S.S., Zafar, B., Hashmi, H.F., Shah, I., Khan, S., Suleman, M., Khan, M., Ullah, S., Ali, S., 2020. Development of multi-epitope subunit vaccine for protection against the norovirus' infections based on computational vaccinology. *J. Biomol. Struct. Dyn.* 1–12.
- Ahn, C., Cho, J.T., Lee, J., Lim, C.S., Kim, Y.Y., Han, J., Kim, S., Lee, J.S., 2000. Detection of Hantaan and Seoul viruses by reverse transcriptase-polymerase chain reaction (RT-PCR) and restriction fragment length polymorphism (RFLP) in renal syndrome patients with hemorrhagic fever, 53, pp. 79–89.
- Atitechou, M., Saleh, S.S., McElroy, A.K., Schmaljohn, C., Ibrahim, M.S., 2005. Identification of Dobrava, Hantaan, Seoul, and Puumala Viruses by One-Step Real-Time RT-PCR, 124, pp. 21–26.
- Brennick, C.A., George, M.M., Corwin, W.L., Srivastava, P.K., Ebrahimi-Nik, H.J.I., 2017. Neoepitopes as cancer immunotherapy targets: key challenges and opportunities, 9, pp. 361–371.
- Brocato, R.L., Hooper, J.W.J.V., 2019. Progress on the prevention and treatment of hantavirus disease, 11, p. 610.
- Castiglione, F., Mantile, F., De Berardinis, P., Prisco, A., 2012. How the interval between prime and boost injection affects the immune response in a computational model of the immune system. *Comput. Math. Meth. Med.* 2012.
- Chen, J., Liu, H., Yang, J., Chou, K.-C., 2007. Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. *Amino Acids* 33, 423–428.
- Clement, J., Maes, P., Van Ranst, M.J.V.R., 2014. Hemorrhagic fever with renal syndrome in the new, and hantavirus pulmonary syndrome in the old world: paradigm (se) gm lost or regained?, 187, pp. 55–58.
- Colovos, C., Yeates, T.O., 1993. Verification of protein structures: patterns of nonbonded atomic interactions, 2, pp. 1511–1519.
- Dong, Y., Ma, T., Zhang, X., Ying, Q., Han, M., Zhang, M., Yang, R., Li, Y., Wang, F., Liu, R., Wu, Xingan, 2019. Incorporation of CD40 ligand or granulocyte-macrophage colony stimulating factor into Hantaan virus (HTNV) virus-like particles significantly enhances the long-term immunity potency against HTNV infection, 68, pp. 480–492.
- Doytchinova, I.A., Flower, D.R., 2007. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* 8, 1–7.
- Echterdiek, F., Kitterer, D., Alschner, M.D., Schwenger, V., Ruckenbrod, B., Bald, M., Latus, J., 2019. Clinical course of hantavirus-induced nephropathia epidemica in children compared to adults in Germany—analysis of 317 patients, 34, pp. 1247–1252.
- 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, pp. 571–607.
- Genheden, S., Ryde, U., 2015. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discovery* 10, 449–461.
- Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D.C., Jahn, D., 2005. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* 33, W526–W531.
- Gul, H., Ali, S.S., Saleem, S., Khan, S., Khan, J., Wadood, A., Rehman, A.U., Ullah, Z., Ali, S., Khan, H., 2020. Subtractive proteomics and immunoinformatics approaches to explore *Bartonella bacilliformis* proteome (virulence factors) to design B and T cell multi-epitope subunit vaccine. *Infect. Genet. Evol.* 85, 104551.

- 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.
- Hall, P.R., Leitão, A., Ye, C., Kilpatrick, K., Hjelle, B., Oprea, T.I., Larson, R.S., 2010. Small molecule inhibitors of hantavirus infection. *Bioorg. Med. Chem. Lett.* 20, 7085–7091.
- He, R., Yang, X., Liu, C., Chen, X., Wang, L., Xiao, M., Ye, J., Wu, Y., Ye, L.J.C., 2018. Efficient control of chronic LCMV infection by a CD4 T cell epitope-based heterologous prime-boost vaccination in a murine model. *Cell. Mol. Immunol.* 15, 815–826.
- Heo, L., Park, H., Seok, C., 2013. GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.* 41, W384–W388.
- Hjelle, B., Spiropoulou, C.F., Torrez-Martinez, N., Morzunov, S., Peters, C.J., Nichol, S.T., 1994. Detection of Muerto Canyon virus RNA in peripheral blood mononuclear cells from patients with hantavirus pulmonary syndrome, 170, pp. 1013–1017.
- Jiang, W., Wang, P.-Z., Yu, H.-T., Zhang, Y., Zhao, K., Du, H., Bai, X.-F., 2014. Development of a SYBR Green I based one-step real-time PCR assay for the detection of Hantaan virus. *J. Virol. Methods* 196, 145–151.
- Jonsson, C.B., Figueiredo, L.T.M., Vapalahti, O.J., 2010. A global perspective on hantavirus ecology, epidemiology, and disease, 23, pp. 412–441.
- Kaliampurthi, S., Selvaraj, G., Kaushik, A.C., Gu, K.-R., Wei, D.-Q., 2018. Designing of CD8⁺ and CD8⁺-overlapped CD4⁺ epitope vaccine by targeting late and early proteins of human papillomavirus, 12, p. 107.
- Kaliampurthi, S., Selvaraj, G., Chinnasamy, S., Wang, Q., Nangraj, A.S., Cho, W., Gu, K., Wei, D.-Q.J.V., 2019. Exploring the papillomaviral proteome to identify potential candidates for a chimeric vaccine against cervix papilloma using immunomics and computational structural vaccinology, 11, p. 63.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858.
- Khan, S., Khan, A., Rehman, A.U., Ahmad, I., Ullah, S., Khan, A.A., Ali, S.S., Afridi, S.G., Wei, D.-Q.J.I., 2019. Immunoinformatics and structural vaccinology driven prediction of multi-epitope vaccine against Mayaro virus and validation through *in silico* expression, 73, pp. 390–400.
- Khan, S., Ali, S.S., Zaheer, I., Saleem, S., Ziaullah, Zaman N., Iqbal, A., Suleman, M., Wadood, A., Rehman, A.U., 2020. Proteome-wide mapping and reverse vaccinology-based B and T cell multi-epitope subunit vaccine designing for immune response reinforcement against *Porphyromonas gingivalis*. *J. Biomol. Struct. Dyn.* 1–15.
- Kim, D.E., Chivian, D., Baker, D.J., 2004. Protein structure prediction and analysis using the Robetta server, 32, pp. W526–W531.
- Kozakov, D., Hall, D.R., Xia, B., Porter, K.A., Padhorny, D., Yueh, C., Beglov, D., Vajda, S., 2017. The ClusPro web server for protein–protein docking. *Nat. Protoc.* 12, 255.
- Krieger, E., Vriend, G.J.B., 2014. YASARA view—molecular graphics for all devices—from smartphones to workstations, 30, pp. 2981–2982.
- Kruger, D.H., Figueiredo, L.T.M., Song, J.-W., Klempa, B.J., 2015. Hantaviruses—globally emerging pathogens, 64, pp. 128–136.
- Lamiable, A., Thévenet, P., Rey, J., Vavrusa, M., Derreumaux, P., Tufféry, P., 2016. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res.* 44, W449–W454.
- Larsen, M.V., Lundsgaard, C., Lamberth, K., Buus, S., Lund, O., Nielsen, M.J., 2007. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction, 8, p. 424.
- Laskowski, R., MacArthur, M., Thornton, J., 2006. PROCHECK: Validation of Protein-Structure Coordinates.
- Laskowski, R.A., Jabłońska, J., Pravda, L., Vařeková, R.S., Thornton, J.M., 2018. PDBsum: structural summaries of PDB entries. *Protein Sci.* 27, 129–134.
- Liu, R., Ma, H., Shu, J., Zhang, Q., Han, M., Liu, Z., Jin, X., Zhang, F., Wu, X.J., 2020. Vaccines and Therapeutics Against Hantaviruses, 10, p. 2989.
- Lu, I.-N., Farinelle, S., Sausy, A., Muller, C.P.J.C., 2017. Identification of a CD4 T-cell epitope in the hemagglutinin stalk domain of pandemic H1N1 influenza virus and its antigen-driven TCR usage signature in BALB/c mice. *Cell. Mol. Immunol.* 14, 511–520.
- McGuffin, L.J., Bryson, K., Jones, D.T.J.B., 2000. The PSIPRED protein structure prediction server, 16, pp. 404–405.
- Mittal, A., Sasidharan, S., Raj, S., Balaji, S., Saudagar, P.J., 2020. Exploring the Zika genome to design a potential multiepitope vaccine using an immunoinformatics approach. *Int. J. Pept. Res. Ther.* 1–10.
- Nielsen, M., Lund, O.J., 2009. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction, 10, p. 296.
- Piccirillo, C.A., Bjur, E., Topisirovic, I., Sonenberg, N., Larsson, O.J., 2014. Translational control of immune responses: from transcripts to translomes, 15, pp. 503–511.
- Rapin, N., Lund, O., Castiglione, F., 2011. Immune system simulation online. *Bioinformatics* 27, 2013–2014.
- Rasmussen, J., Andersson, C., Norrman, E., Haney, M., Evander, M., Ahlm, C., 2011. Time to revise the paradigm of hantavirus syndromes? Hantavirus pulmonary syndrome caused by European hantavirus. *Eur. J. Clin. Microbiol. Infect. Dis.* 30, 685–690.
- Sabah, S.N., Gazi, M.A., Sthity, R.A., Husain, A.B., Quyyum, S.A., Rahman, M., Islam, M. R., 2018. Designing of epitope-focused vaccine by targeting E6 and E7 conserved protein sequences: an immuno-informatics approach in human papillomavirus 58 isolates. *Interdiscip. Sci.* 10, 251–260.
- Saha, S., Raghava, G., 2006. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res.* 34, W202–W209.
- UniProt, C., 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515.
- Weng, G., Wang, E., Wang, Z., Liu, H., Zhu, F., Li, D., Hou, T., 2019. HawkDock: a web server to predict and analyze the protein–protein complex based on computational docking and MM/GBSA. *Nucleic Acids Res.* 47, W322–W330.
- Wiederstein, M., Sippl, M.J., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, 35, pp. W407–W410.
- Yang, D., Shao, J., Hu, R., Chen, H., Xie, P., Liu, C.J., 2017. Angiotensin II promotes the anticoagulant effects of rivaroxaban via angiotensin type 2 receptor signaling in mice, 7, pp. 1–11.
- Zhang, L., 2018. Multi-epitope vaccines: a promising strategy against tumors and viral infections. *Cell. Mol. Immunol.* 15, 182–184.
- Zuo, S.-Q., Fang, L.-Q., Zhan, L., Zhang, P.-H., Jiang, J.-F., Wang, L.-P., Ma, J.-Q., Wang, B.-C., Wang, R.-M., Wu, X.-M., 2011. Geo-spatial hotspots of hemorrhagic fever with renal syndrome and genetic characterization of Seoul variants in Beijing, China, 5, e945.