



# Identification of functional epitopes of structural proteins and in-silico designing of dual acting multiepitope anti-tick vaccine against emerging Crimean-Congo hemorrhagic fever virus



Neha Shrivastava<sup>a</sup>, Ankit Verma<sup>a</sup>, Paban Kumar Dash<sup>a,\*</sup>

<sup>a</sup> Division of Virology, Defence Research and Development Establishment, Jhansi Road, Gwalior, PIN-474002, India

## ARTICLE INFO

**Keywords:**  
CCHF  
Multiepitope vaccine  
Immunoinformatics  
Subolesin  
Anti tick vaccine

## ABSTRACT

Recurrent outbreaks of Crimean-Congo hemorrhagic fever (CCHF) virus infection in different parts of world are a major global health concern. The CCHF viral infection is associated with severe hemorrhagic fevers and mortality up to 40%. More than 30 countries in Asia, Europe and Africa are affected with CCHF infection. Prevention of infection through vaccine becomes more important when no effective antiviral and associated therapies are available. Further ticks play a crucial role in maintenance and transmission of CCHFV. Therefore, the control of transmission by ticks is warranted for ultimate prevention of outbreak. The study employed a series of immunoinformatics approaches to design novel multiepitope vaccine targeting highly immunodominant epitopes of major structural proteins (Nucleoprotein and Glycoprotein complex) of CCHFV. Vaccine was designed by incorporating linear and conformational B cell, helper and cytotoxic T cell epitopes from these crucial immunogenic proteins adjoined with appropriate linkers and adjuvant. This vaccine construct was also complemented with a highly immunogenic and conserved protective tick salivary antigen named subolesin to impart dual activity as a unique transmission blocking vaccine. The B-cell peptides were also experimentally validated. The designed vaccine was further *in silico* validated for its physiochemical properties, allergenicity and immunogenicity etc. The proposed candidate vaccine construct has the potential to function both as a vaccine against CCHF virus as well as a universal anti-tick vaccine.

## Abbreviation

CCHF:	Crimean-Congo hemorrhagic fever virus
NP:	Nucleoprotein
GPC:	Glycoprotein complex
CFR:	Case fatality rate
RdRp:	RNA dependent RNA polymerase
ACC:	Auto cross covariance
Th:	Helper T cell
Tc:	Cytotoxic T cell
IEDB:	Immune epitope database
GRAVY:	Grand average of hydropathicity
MDS:	molecular dynamic simulation
RMSD:	Root mean square deviation

## 1. Introduction

Crimean-Congo hemorrhagic fever (CCHF) is an emerging tick

borne viral disease caused by CCHF Orthenairovirus. CCHF is one of the most disseminated arboviral diseases in the world and is endemic in many parts of world (Al-Abri et al., 2017). According to World health organisation (WHO) and Global Infectious Diseases and Epidemiology Online Network (GIDEON) more than 50 cases of human CCHF infection are reported annually from endemic zones with case fatality rate of more than 80% in humans (Messina et al., 2015). Since the emergence of CCHFV from Crimea in 1944, it covered huge area in the southern Soviet Union, from western China across Southern Asia to Middle East, Bulgaria and Balkans and throughout most of the Africa (Bente et al., 2013). Off late, since 2011, CCHF has also been reported from many parts of India at regular interval.

The genome of CCHF virus is negative sense single stranded tri segmented RNA. CCHFV belongs to the genus *Orthenairovirus* of family *Nairoviridae* (Adams et al., 2017). It is classified as risk group 4 (RG4) pathogen requiring handling of the live virus in biosafety level 4 containment facility. Its 19.2 kb genome has three segments namely small (S), medium (M), and large (L) RNA segments encoding viral

\* Corresponding author. Tel: + 91-751-2390290  
E-mail address: [pabandash@drde.drdo.in](mailto:pabandash@drde.drdo.in) (P.K. Dash).

nucleocapsid protein (NP) of 1672 nucleotides, glycoprotein precursor complex (GPC) of 5364 nucleotides, and RNA dependent RNA polymerase (RdRp) of 12150 nucleotides respectively (Duh et al., 2008).

Serosurveillance studies showed high seropositivity in domestic and wild life animals, though the livestock are asymptomatic reservoir of the virus (Mourya et al., 2015). Animal handlers are at high risk of gaining infection due to close contacts with infected animals and their body fluids. Similarly, nosocomial infections are very common among health care workers (Aradaib et al., 2010).

Hard ticks belonging to the genus *Hyalomma* are considered as the principal vector of CCHF *Orthonairovirus* and are widely distributed across the endemic areas. The geographic distribution of *Hyalomma* ticks extends up to 50° north latitude encompassing the endemic areas (Jamshidi et al., 2017). Virus primarily transmits to animals and humans through these ticks.

In spite of widespread circulation of virus in the world, there is still a lack of effective treatment and vaccines for controlling infection among population. Efforts were being made to develop CCHF vaccine targeting nucleoprotein that seemed to be immunogenic but failed to confer protection against lethal disease (Dowall et al., 2015). So far, there is no human vaccine available for CCHFV. However, few commercial cattle tick vaccines are available that aims for prevention of transmission through ticks in animals. Although this could lead to reduction of infections through tick bites, however the risk of infection through other modes remained unaddressed. In a study, it was found that disease occur during March-July when ticks are most active and observed that majority of patients were infected via tick bite while few with contact of CCHF patients (Papa et al., 2004). Such scenario of vaccine give rise an urgent need for development of effective, safe and potent vaccine candidate that helps in management of disease by preventing infection. Therefore, the research focused to map and utilize highly immunodominant epitopes for the development of novel multi-epitope vaccine against CCHF orthonairovirus in order to intensify cellular as well as humoral immune responses against the virus. Vaccine is designed in such a way that serves a dual purpose of providing immunity to host (humans and animals) by neutralizing viral antigens and also serves as anti-tick vaccine that proportionally reduce the infection associated with tick population.

Major requisite for a potential vaccine candidate is an immunogen harboring adequate competent B cell and cytotoxic T cell lymphocyte (CTL) epitopes. The structural proteins of CCHFV i.e., NP and GPC are the two most immunogenic regions of the virus against which immune system primarily respond and thus selected for vaccine development. NP is one of the most abundant protein in viral particles and plays a key role in not only encapsulating viral RNA but also helps in viral replication along with RdRp (Zivcec et al., 2016). Envelope glycoprotein complex (GPC) precursor releases Gn and Gc by proteolytic cleavage and forms important structural components of CCHFV. Gn and Gc proteins of glycoprotein complex plays significant role in entry and fusion, immune invasion and virion formation (Zivcec et al., 2016). Initial binding of the CCHFV to cell surface membrane is facilitated by Gn and Gc. Both the proteins have shown to play significant role in virus attachment and recognition by host immune system. Nucleoprotein is highly conserved compared to glycoprotein; however, both the proteins are highly immunogenic in nature (Hewson et al., 2004). Thus, targeting NP and GPC for vaccine development is the preferred choice for virus neutralization and vaccine development.

CCHF is transmitted by hard ticks belonging to *Hyalomma* species. Moreover, as a tick borne viral infection, the saliva of tick plays a critical role in intensifying host defenses and also facilitates the blood flow for adequate feeding. Tick saliva manifest multiple activities including vasodilator, anticoagulant, cytolytic, immunosuppressive and anti-inflammatory activity. Both saliva and salivary glands play crucial role in transmission of pathogen to vertebrate hosts. Number of salivary protein molecules have been identified that are involved in modulation of host defense and transmission of tick borne pathogens (Šimo et al.,

2017). Subolesin (SUB) is reported to be one of the most potent and conserved salivary protein antigen explored for development of an anti-tick vaccine (Shakya et al., 2014). Subolesin showed 99.2–100% nucleotide conservancy across tick species that provides cross-protection against variety of tick strains and thus makes a perfect choice for vaccination (De la Fuente et al., 2011). Earlier studies on immunization of tick SUB have showed a protective efficacy against tick infestations, leading to reduction of vectorial capacity and fertility of ticks. Thus appending SUB in vaccine candidate provides an additional feature to a vaccine against tick borne pathogens. In addition to protection against pathogen, this vaccine can impart protection against tick infestations, thereby acting as a broad spectrum candidate against tick borne infections. As such, ticks harbor very high concentration of CCHFV and bite of viremic tick generally leads to establishment of infection (Dickson and Turell, 1992, Papa et al., 2017). Co-feeding of ticks in host and transstadial transmission plays crucial role in dissemination of virus. Further, the prevention of infection in animals is likely to break the chain of transmission from animals to human primarily in those involved in dairy farming, animal husbandry and animal care givers. Therefore, subolesin amino acid sequences are added to CCHF multi-epitope vaccine that provides immunity against both pathogen and tick infestation, akin to the concept of anti-tick vaccine.

Vaccinomics, a latest technology platform, is now widely explored for development of novel vaccine molecules. Using different computational programs, present study focused to propose epitopes as potential candidate for vaccine development against CCHFV using *in-silico* methods taking NP & GPC taken into consideration. Prospective B cell and T cell epitopes of the structural proteins of pathogens have been predicted. In the current study, the immunodominant epitopes of CCHF virus are predicted and adjoined with suitable adjuvant using appropriate linkers to construct a vaccine molecule. Designed vaccine candidate is a blend of highly immunodominant epitopes that neutralizes the virus along with the immunogenic salivary gland protein of tick that proportionally act on ticks and as serves as anti-tick vaccine. Vaccine construct was further analyzed for different parameters including antigenicity and allergenicity. We also report the simulation of the vaccine construct with prevalent MHC molecule to study the stability in *in-vivo* conditions.

## 2. Methodology

### 2.1. Retrieval of CCHFV Proteome

Amino acid sequences of NP and GPC were retrieved from NCBI database (<https://www.ncbi.nlm.nih.gov>) with Accession no. AEO72054 and AEO72050 respectively, corresponding to CCHFV strain NIV 112143 for further analysis. In addition, sequences of 84 CCHFV strains from diverse geographic areas were retrieved from GenBank and used for epitope conservancy analysis. All the protein sequences were retrieved in FASTA format. Amino acid sequence of subolesin was also retrieved from NCBI database with Accession no. AGI44624 for final addition in multiepitope vaccine. Nucleoprotein modeled structure was also retrieved from PDB with PDB ID: 4AQF for conformational epitope prediction.

### 2.2. Prediction of MHC-II defined helper T lymphocyte (HTL) epitope

Helper T cells that binds particularly to MHC-II molecule plays imperative part in activation of humoral immune reaction and also secretes cytokines that activates the cellular immune reaction. Helper T cell epitopes from NP and GPC were anticipated utilizing MHC-II binding prediction beneath IEDB server (<http://tools.iedb.org/mhcii/>). This determination strategy employments combining NN-align, Agreement approach, SMM-align and CombLib for the molecule to discover the finest fit for given MHC molecule and human gene. Peptides were sorted on the premise of percentile score.

### 2.3. Prediction of MHC-I defined Cytotoxic T lymphocyte (CTL) epitope

MHC-I inducing CTL epitopes that could generate effective immune response by triggering cytokines like IFN $\gamma$  and other interferons are crucial for vaccine effectiveness. Prediction of CTL epitopes was done on account of their combinatorial scores of specific MHC binding affinity, TAP transport affinity, proteasomal C-terminal cleavage and potential MHC binding ligand employing Net CTL 1.2 server (<http://www.cbs.dtu.dk/services/NetCTL/>).

### 2.4. Prediction of linear and conformational B cell epitope

B cell epitopes play significant role in humoral immune response that neutralizes many pathogens including viruses. Approaches to predict potential linear B cell epitopes involves different parameters such as hydrophilicity, flexibility, accessibility, turns, exposed surface, polarity and antigenic propensity of polypeptide chains that have been correlated with the location of linear epitopes. The linear B-cell epitopes are predicted using IEEDB web server (<http://tools.iedb.org/bcell/>). This has led to a search for empirical rules that would allow the position of linear epitopes to be predicted.

While ElliPro has been employed to predict the conformational B cell epitopes for which tertiary structure of protein is needed as input. Though the structure of NP was retrieved from PDB whereas GPC has not modeled yet. Tertiary structure of GPC has been modeled using I-Tasser B6 server followed by energy minimization with SPDBV 4.10 programme and validation with Ramachandran plot.

## 3. Validation of B cell epitopes

### 3.1. Serum Samples

Serum samples from sheep, goat and bovine were collected from hotspot zone of Gujarat, India during 2018–2019 outbreak and were screened and confirmed using an in-house CCHF multispecies indirect ELISA (Shrivastava et al., 2019).

### 3.2. Peptide based ELISA

Bioinformatics analysis revealed a list of potent epitopes on the basis of scores but out of these, only the highly reacting peptides were intended to append in the vaccine candidate. Predicted B-cell epitope peptide library was generated by custom synthesis of peptides with Fmoc chemistry (Imgenex, India). The peptide library comprising all 14 immunodominant BCE epitopic regions of structural proteins were screened using an indirect peptide ELISA against different species of CCHF infected host sera to find out the common epitopes reacting in multiple species. The reacting peptides were further confirmed with dot blot assay along with negative sera. Briefly, HPLC purified synthetic peptides (1 $\mu$ g) were dissolved in 200  $\mu$ l PBS and used to coat the microtiter wells of plate (Nunc, Denmark) in carbonate-bicarbonate buffer (pH 9.6). Following coating, plate was blocked with 5% BSA for 2 hrs at 37°C. The plate was washed with washing buffer (0.01% PBST) to remove blocking buffer from the wells. The heat inactivated test sera was diluted in serum dilution buffer (1% SMP with 0.1% PBST) in 1:1000 dilution and added 100  $\mu$ l/well, incubated for 1 hr at 37°C. The plate was washed with washing buffer (0.1% PBST) for 5 times (3 min each). Respective conjugates 100  $\mu$ L/well of HRP-conjugated anti-goat IgG, HRP-conjugated anti-sheep IgG and HRP-conjugated anti-bovine IgG were added in each wells with dilution of 1:10,000 in 0.05% PBST, incubated at 37°C for 60 min. Following incubation, plate was washed and developed with TMB (Sigma, USA) and the reaction was stopped with 1N HCl. The absorbance was taken at 450 nm using an ELISA plate reader (BioTek, USA). The cutoff was decided taking thrice of negative controls.

### 3.3. Epitope conservancy analysis

The predicted epitopes were checked for their conservancy among different CCHF viral strains in all 6 genotypes. All HTL, CTL and BCL epitopes of each of NP and GPC regions were analyzed with aforementioned retrieved amino acids sequences from NCBI dataset for calculating degree of conservancy through IEEDB Epitope conservancy analysis tool.

### 3.4. Construction of novel multiepitope vaccine

Predicted HTL, CTL and B cell epitopes of CCHF NP and GPC were joined together using appropriate linkers to form a final vaccine construct. In total eight HTL, eight CTL and eight BCL were shortlisted on the basis of higher scoring in different propensity scales in case of BCL epitopes, whereas CTL and HTL epitopes were shortlisted on the basis of higher IC<sub>50</sub> values with lower percentile score and merged using GGGGS, AAY and KK linkers respectively. To enhance the efficacy and antigenicity parameter of vaccine molecule, an amino acid sequence (KLKL)<sub>5</sub>KLK, a part of adjuvant IC31 was added in the N-terminal of construct. The sequence of Subolesin, a potent tick salivary protective antigen was added in C-terminal of vaccine construct that elicit an universal anti-tick vaccine effect. Adjuvant was adjoined to the vaccine molecule with EAKK linker to provide flexibility and efficient separation between adjacent epitopes (Argos, 1990). Whereas subolesin was added with KK linker in continuation of B cell epitopes (Rana and Akhter, 2016).

### 3.5. Prediction of allergenicity

The constructed novel multiepitope vaccine molecule was screened for its allergenicity, whether it is allergic or non-allergic in nature as allergic molecule might cause adverse effects. The allergenicity was checked with Allertop (<http://www.ddg-pharmfac.net/AllerTOP>) online available server which relies on the principle of auto cross covariance (ACC) transformation that is a sequence query of peptides of different length. Allertop perform BLAST with more than 4850 known allergen and non-allergen peptides from different species.

### 3.6. Prediction of antigenicity

The vaccine construct was also screened for its antigenic nature because antigenicity is one of major factor for vaccine molecule to be effective. Antigenicity was checked using online available server named Vaxijen 2.0 at 0.4 default threshold value (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>). Vaxijen works on the principle of auto cross covariance (ACC) transformation and is alignment free prediction platform. It allows antigen to be classified on the basis of physical-chemical properties of proteins.

### 3.7. Determination of physical and chemical properties

Dual purpose novel multiepitope vaccine molecule was further analyzed for its physicochemical properties. Length of amino acids, molecular weight, amino acid composition, pI index, extinction coefficient, estimated half life, instability index were calculated using Expasy tool named ProtParam (<https://web.expasy.org/protparam/>). The aliphatic index and grand average of hydropathy (GRAVY) was also estimated.

### 3.8. Modeling and validation of secondary and tertiary structure

Secondary and tertiary structure of multiepitope vaccine construct was predicted to understand the behavior of molecule in *in-vivo* conditions. Secondary and tertiary structure depict the information about its helical strands, sheets and coils formed to unreveal the mechanism

of interaction with different HLA allele molecules. Secondary structure was predicted using PSIPRED server (<http://bioinf.cs.ucl.ac.uk/psipred/>) from ExPasy. PSIPRED applies very stringent validation to evaluate the structure with accuracy with an average Q<sub>3</sub> score of more than 80% towards prediction. While the tertiary structure was designed with Phyre 2.0 (<http://www.sbg.bio.ic.ac.uk/phyre2/>) on the basis of homology modeling that states the quality of predicted 3D structure. Structure was then validated with Ramachandran plot using PROCHECK (<https://servicesn.mbi.ucla.edu/PROCHECK/>). PROCHECK assess the stereochemical quality of protein structure by comparing it with well-refined, high-resolution structures followed by energy minimization with SPDV 4.10.

### 3.9. Molecular docking and simulation with HLA molecules

The final vaccine construct was further analyzed for its interacting ability with HLA molecules, as the vaccine candidate must be able to efficiently interact with wide range of population having varying HLA alleles. Vaccine candidate was docked with highly circulating HLA gene in majority of European and Asian population (Jennifer M et al., 2000). Patchdock was used for docking of vaccine construct with HLA A02 allele which was further refined with FIREDOCK that generate top 10 structures on the basis of their scores, attractive energy, attractive and repulsive vanderwals energy (Vdw), atomic contact energy (ACE) and hydrogen bonds (HB) values. Top ranked structure was taken and studied for its molecular dynamic simulation (MDS). MDS was performed with HLA allele that is most prevalent in Asian population i.e. HLA 02 with iMODS online server (<http://imods.chaconlab.org/>). The stability and interactions between vaccine construct comprising multiple epitopes, anti tick salivary protein and HLA molecule was examined using various parameters following energy minimization. iMODS analyze the internal coordinates of protein with normal modes (NMA) (Lopez-Blanco et al., 2014). Simulation was run in advance mode of 20 and edNMA as elastic network model. Log files of the simulated model in given in supporting data. iMODS serve predicts the motions of docked complex in deformability, eigen values, B-factors and covariance parameters.

### 3.10. In silico cloning and optimization for expression

In order to express a multiepitope vaccine molecule, it has to adjoin in frame with an expression vector (pET sumo) under restriction enzyme sites. Further, codon optimization was carried out with acceptable GC content and codon adaptation index (CAI) to fortify the higher expression in *E. coli* as host employing Java codon adaptation tool (JCat).

## 4. Results and Discussion

### 4.1. Retrieval of CCHF Proteome and analysis

Amino acid sequences of CCHFV (accession no. AEO72054 and AEO72050) were retrieved for epitope mapping. All HTL, CTL and BCL are predicted from retrieved nucleoprotein and glycoprotein sequence of CCHFV genome. Amino acid sequences of CCHFV representing six different genotypes were also retrieved in Fasta format and aligned using ClustalW in MegaX software. Phylogenetic analysis was conducted with maximum likelihood statistical method at 500 bootstrap to cluster strains into representing genotypes (S1 Fig). One strain from each representing genotypes were included to study the conservancy of epitopes. The conservancy of epitopes will lead to the effectiveness of the candidate vaccine molecule against divergent strains.

### 4.2. Prediction of MHC-II defined Helper T lymphocyte (HTL) epitopes

A long lived T cell immune response was observed in survivors of many viral infections including CCHFV (Goedhals et al., 2017). This

suggests that the vaccine that includes helper T cell (Th) and cytotoxic T cell (Tc) will be effective in providing long term immunity against viral infection. Helper T cells epitopes of both the structural proteins were predicted (NP and GPC) with IEDB server for Human HLA genes DP, DQ and DR. Peptides which found to be common with higher IC<sub>50</sub> score and lower percentile rank in all three alleles were selected as potential HTL epitopes. The higher IC<sub>50</sub> score depicts the higher binding affinity of epitopes (S1 Table). A total eight epitopes four each from NP and GPC protein were finalized. (Table 2). Though QIDTAFFSS was scored first in percentile score but also selected in CTL, thus other top scoring in all three HLA allele were selected. The netmhpan<sub>IC50</sub> score for NP ranges from 87 to 1434 while for GPC ranges from 16 to 3918 (covering a wide range of immunogenicity) has been selected and combined with GGGGS linkers for final vaccine construct.

### 4.3. Prediction of MHC-I defined Cytotoxic T lymphocyte (CTL) epitope

CTL plays crucial role in effectiveness of a vaccine molecule by inducing large population of virus specific CD8+ T cells and related interferons which participates in clearance of pathogens. NetCTL 1.2 server has been employed to predict MHC-I inducing CTL epitopes. Peptides with higher predicted parameters score were selected from each of NP and GPC region (S2 Table). Total eight epitopes, four from each of NP and GPC were shortlisted on the basis of higher combined score of greater than 0.75 representing good binders are finally shortlisted and adjoined with AAY linkers for final vaccine construct (Table 2).

### 4.4. Prediction of linear and conformational B cell epitope

B cell epitope prediction of NP and GPC were done using immune epitope database (IEDB) with different algorithms. IEDB supports the prediction of epitopes with multiple algorithms representing their different properties (surface accessibility, hydrophilicity, antigenicity etc) of peptides for being most immunogenic among the rest. Parker hydrophilicity propensity scale has been employed for the prediction of hydrophilicity of peptides. Length of window size was selected to be 10 amino acids according to the putative mean epitope size. Peptides were selected on the basis of scores, with generated minimum and maximum score of 1.70 and 6.410 respectively. Chou and Fasman beta turn prediction algorithm has been used for predicting beta turns for antibody epitopes. The minimum and maximum scores were found to be 0.7 and 1.1 respectively. Emini surface accessibility scale was used to find the epitopes having probability for being found on the surface with minimum and maximum scores of 0.030 and 9.267 respectively. Karplus and Schulz flexibility scale was also used for prediction of epitopes based on mobility of protein segments on the basis of the known temperature or B factors of the  $\alpha$ -carbons of 31 proteins with minimum and maximum antigenic propensity scores of 0.888 and 1.110 respectively. Kolaskar and Tongaonkar antigenicity scale uses physicochemical properties of amino acid residues and their frequencies of occurrence in experimentally known segmental epitopes. This is one of the most important algorithms which can predict antigenic determinants with about 75% accuracy. The average propensity score of protein was 1.05, with maximum of 1.197 and minimum value of 0.879. Epitopes are finally selected based on the high scores in minimum of three algorithms (S3 Table). Nucleoprotein epitope fragments of NP\_BCL 3 (VEAL and NGYL) and 4 (EVDKA and DSMITN) share sequences that were also predicted and showed reactivity with positive sheep sera in western blot analysis in another study (Liu et al., 2014).

Conformational epitopes are predicted using Ellipro (<http://tools.iedb.org/ellipro/>). Ellipro require tertiary structure for conformational epitope prediction. Therefore, CCHF NP tertiary structure was retrieved from PDB with PDB ID 4AQF, modeled using X-ray diffraction with resolution of 3.1 Å (S4a Table). While CCHF GPC was modeled using I-Tasser B6 server using amino acid sequences of strain NIV 112143. The

**Table 1**

Screening of Fmoc synthesized B cell epitopes with different host sera using Indirect Peptide ELISA

S. no.	Peptide sequence	Goat sera	Sheep sera	Bovine sera
P1	NRGGDENPRGPVSRE	+	+	+
P2	NPPWDINKSGRS	+	+	+
P3	ELGKQPRGTKKMKKA	-	+	+
P4	FGTIPVANPDDAQSGHTKS	+	-	+
P5	QMASATDDAQKDSIYAS	-	+	-
P6	VGKQSPFQNAVNVKGN	+	+	-
P7	MENKIEVNNKDEMKNW	-	-	+
P8	PKPEETPTPSKSDKD	+	+	+
P9	DQGPTEENNHHNS	+	+	+
P10	TQDTHPSPTNRSKRN	+	+	+
P11	KIPLLGKMAIYICRMSNHP	-	+	-
P12	LHKKEWPHSRNWRNCPTWCW	-	-	+
P13	PKATCTGDCPERCGCTSSTC	+	-	+
P14	ERLNNKGKGNKLDRGERLADRK	-	+	+

predicted structure was validated with Ramachandran plot using RAMPAGE structure (S2 Fig). Tertiary structure of GPC was then utilized to predict conformational epitopes of GPC (S4b Table). The amino acid residues of conformational epitopes of NP and GPC, overlapping with linear predicted peptides are selected for vaccine designing. One of the predicted glycoprotein epitope (NVMLAVCKRM) has also been picked up in scanning peptides with positive pooled patient sera (Fritzen et al., 2018).

#### 4.5. Antigenicity of selected BCE by CCHFV antibody positive sera

The identified BCE were screened with CCHF goat, sheep and bovine IgG positive and negative sera via peptide ELISA to validate the antigenic nature of epitopes by different host species. Peptides were custom synthesized and tested which were not previously reported by any wet lab experiments. BCL\_1, BCL\_2, BCL\_3 and BCL\_8, predicted in this study are also found experimentally reactive in earlier wet lab experiments (Liu et al., 2014, Fritzen et al., 2018). The peptide library of 14 high scoring BCE were screened with peptide ELISA (Table 1). Five peptides found strongly reactive against sera from all the three host species were finally selected (BCL\_1, BCL\_2, BCL\_5, BCL\_6 and BCL\_7). The average absorbance of peptides against sera from different host species were plotted (Fig 1). These five peptides were also cross-checked with dot blot ELISA that revealed reactivity with confirmed sera of CCHF infection (S3 Fig). This validates the prediction methods and its potential utility for designing of a multi epitope vaccine.

**Table 2**

Selected BCL, HTL and CTL epitopes of respective target proteins

Target Protein	B cell epitopes	HTL epitopes	CTL epitopes
Nucleoprotein (NP)	NRGGDENPRGPVSRE (BCL_1)*	QQAALKWRKDGFPRV (HTL_1)	QIDTAFFSSY (CTL_1)
	NPPWDINKSGRS (BCL_2)*	RWGKKLYELFADDSE (HTL_2)	HIAKAQELY (CTL_2)
	AKKTVEALNGYLDKK (BCL_3)	DNMITNLKKHIQAQ (HTL_3)	AFSSYYWLY (CTL_3)
	DKHKEDEVVKASADNMITNLL (BCL_4)	SSYYWLYKAGVTPET (HTL_4)	DSFQQNRIY (CTL_4)
	PKPEETPTPSKSDKD (BCL_5)*	KSHVCVDYSLDTDGAI (HTL_5)	TAEVHGDNY (CTL_5)
	DQGPTEENNHHNS (BCL_6)*	VLLILFFMFGWRILF (HTL_6)	SVMTLSQMY (CTL_6)
	TQDTHPSPTNRSKRN (BCL_7)*	TMAFLFWFSFGYVIT (HTL_7)	LVDSVGNIF (CTL_7)
	NVMLAVCKRM (BCL_8)	KNMLSGIFGNVFLGI (HTL_8)	FTDYMVFVKW (CTL_8)

Selected BCL, HTL and CTL epitopes of respective target proteins

BCL: B cell epitope, HTL: Helper T lymphocyte, CTL: Cytotoxic T lymphocyte

\* Represents the peptide found reactive in wet lab experiments of this study

#### 4.6. Conservancy analysis of predicted epitopes

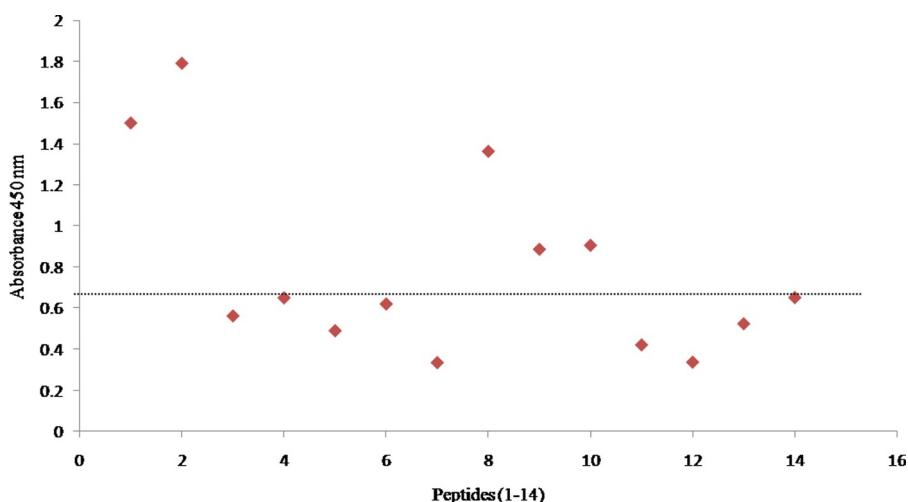
The success of vaccine demands its effectiveness against wide range of divergent CCHF viral strains circulating in different parts of the world. The study was focused to develop multi epitope vaccine against CCHF virus comprising highly immunodominant and highly conserved epitopes to impart broad range of protection. The conservancy of HTL and CTL epitopes were predicted with very high identity. Out of these, 5 epitopes were found 90–100% conserved across genotypes. Rest all epitopes revealed variation of only 2–4 residues (S5 Table). This higher conservancy indicates the potential effectiveness of the designed vaccine against divergent genotypes of CCHF virus circulating in different parts of the world.

#### 4.7. Strategy for selecting linkers

Linkers for the fusion proteins plays a vital role in biological activity and stability of the molecule. The selection of appropriate linker out of the three types of linkers (rigid, flexible and in-vivo cleavable linkers) depends on the utility of constructed molecule. Therefore, in this study, we have adopted flexible and rigid linkers. The flexible linkers like GGGGS allow the flexible interaction between domains and enhances the spatial separation between them which improve the folding and stability of the molecule (Argos, 1990). Whereas, rigid linker was selected between adjuvant and first epitope to maintain distance between the domain and for correct folding of the fusion protein. Alpha helix-forming rigid linker like EAKK is also known to improve the expression of fusion recombinant protein and form a tightly packed backbone with intra segment hydrogen bond (Chen et al., 2013). The dilysine (KK) linkers were selected because of its property to signal the protein for endoplasmic reticulum (ER) retention/retrieval (Jain et al., 2015).

#### 4.8. Construction of novel multi-epitope vaccine constructs

The final vaccine construct was designed by including the predicted B cell epitopes and T cell epitopes (both HTL and CTL). Construct comprises of all these epitopes joined with short linkers to provide sufficient spaces between epitopes in order to work efficiently. Arrangement of epitopes (BCL, CTL and HTL) were finalized by reshuffling the position of each epitope and checked for obtained antigenicity score with Vaxijen v2.0 to obtain highest antigenic score (S6 Table). In designing of construct, all NP and GPC HTLs were combined with GGGGS linkers followed by addition of all NP and GPC CTL linked with AAY linker. Finally, B cell epitopes of NP and GPC were linked



**Figure 1.** Indirect peptide ELISA illustrating average optical density of 14 peptides against serum samples from different species of hosts (sheep, goat and bovine). A threshold line (dotted line) determined the sera reacting against these peptides.

with KK linkers. To increase the immunogenicity score,  $(KLKL)_5$ KLK a part of IC31 adjuvant is added with EAKK linker to the designed molecule.

IC31 is a unique adjuvant comprising of antibacterial 11-mer peptide (KLKL(5)KLK) and immunostimulatory effect of synthetic oligodeoxynucleotide (ODN1a) which is a toll-like receptor 9 (TLR-9) agonist (Olafsdottir et al., 2009). KLKL5 is known to enhance the significant rise in IgG antibody titer and IL-4 and IL-5 cytokine response favoring Th-2 polarization (Fritz et al., 2004). It also enhances the coalition of antigens to antigen presenting cells that favors the depot at injection site (Schellack et al., 2006). Thus, a part of IC31 adjuvant was used to enhance the presentation of vaccine molecule and related immune response. Further, amino acid sequence of subolesin was added at C-terminal of vaccine construct to provide anti tick vaccine effect. The total length of novel multiepitope vaccine construct was 571 amino acids (Fig 2).

#### 4.9. Allergenicity and antigenicity scores

Any vaccine which will be used in humans must be sufficiently immunogenic. Further it should be non allergenic since vaccine allergy may cause adverse reactions that may lead to life threatening situations. The designed vaccine molecule is thus screened for its allergenicity and antigenicity with Allertop and Vaxijen 2.0 programmes respectively (Zaharieva et al., 2017). The vaccine construct of 571

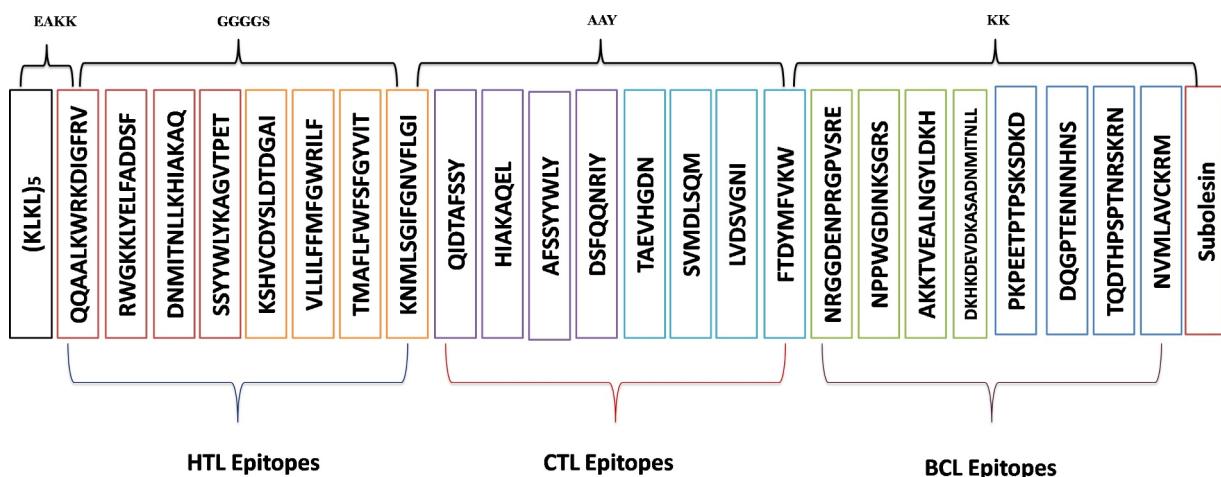
amino acids was found to be non allergen on screening against 2427 allergens and 2427 non-allergens from different species. However immunogenicity score of designed vaccine construct was found to be 0.5870 which depicts it to be an ideal immunogen. A threshold score above 0.4 is generally considered for good immunogen (Nezafat et al., 2017).

#### 4.10. Determination of physical-chemical properties

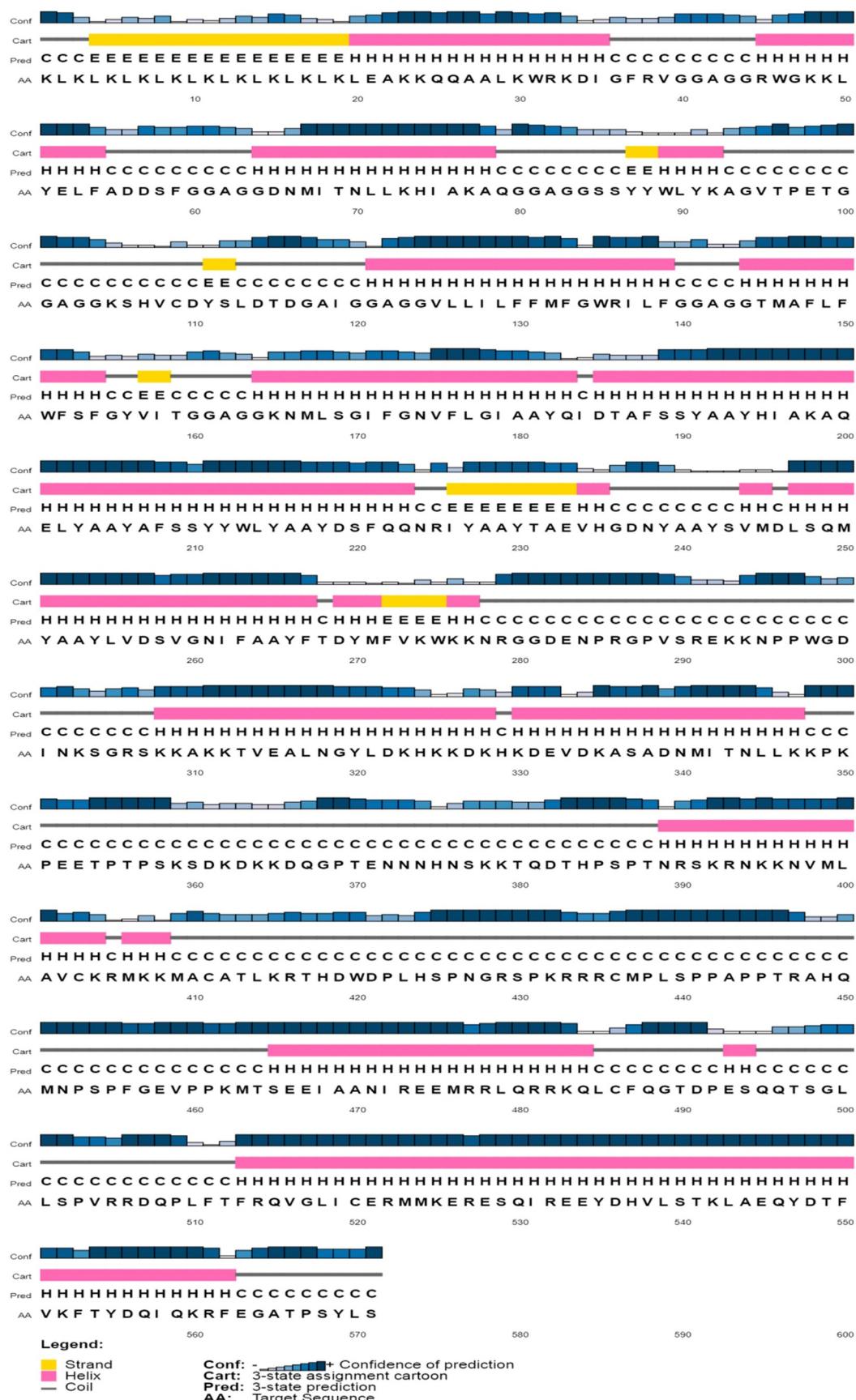
ProtParam tool (<https://web.expasy.org/protparam/>) was used to compute the various physical and chemical properties of designed multiepitope vaccine molecule. Molecular weight of protein was found to be 64 kilo daltons, comprising 571 amino acids that can easily elicit good immune response. Grand average of hydropathicity (GRAVY) was calculated as -0.644, that represent protein is hydrophilic in nature and can easily interact with water molecules. Computed aliphatic index of vaccine construct was 63.82 that depicts the high thermo stability of protein as greater the value, the protein will be highly thermo stable.

#### 4.11. Secondary and tertiary structure modeling and validation

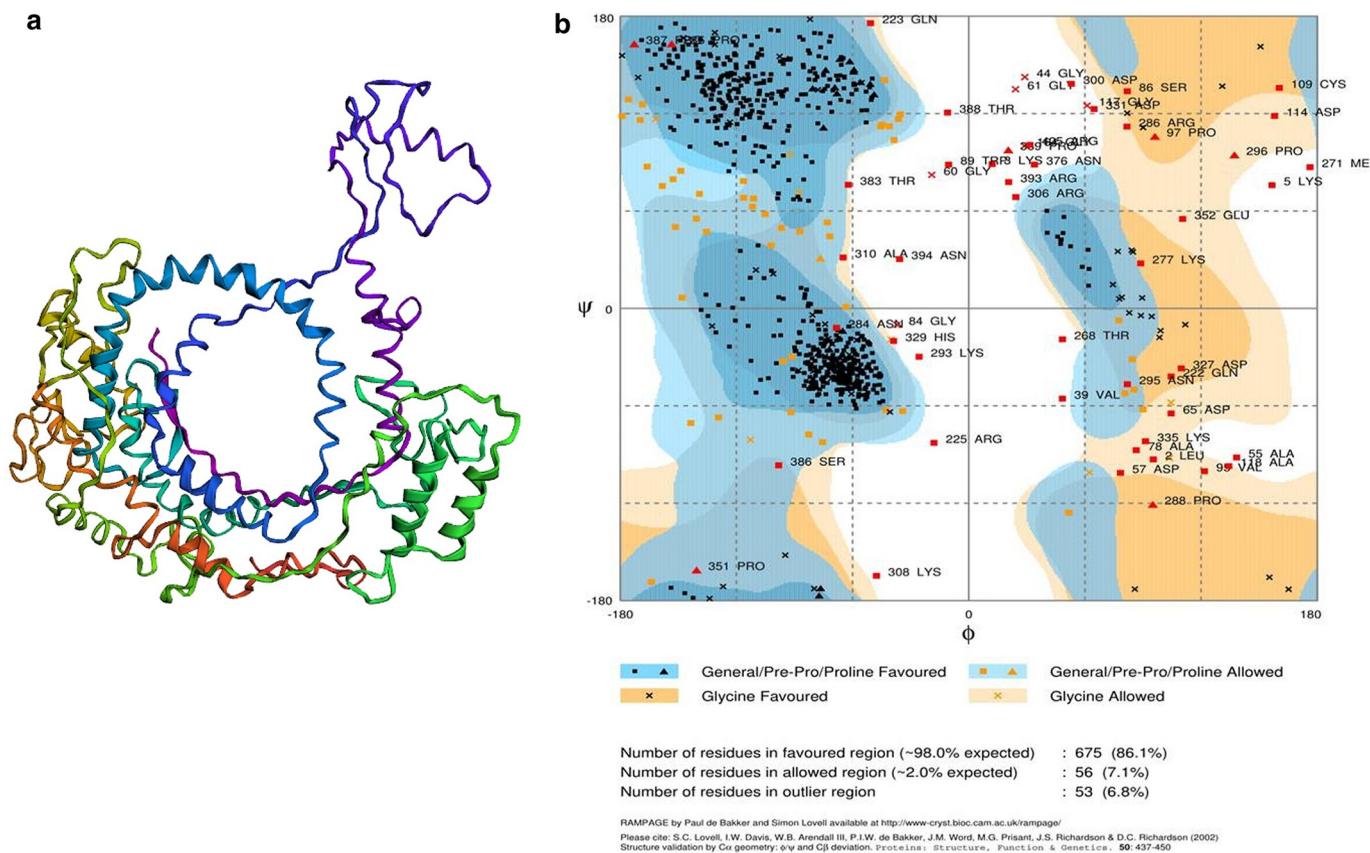
Online server PSIPRED was employed to predict the secondary structure of vaccine construct. Prediction was based on the composition and sequence of amino acids (Jones, 1999). Vaccine construct of 571 amino acids was examined for coils, beta turns, alpha helix and sheets



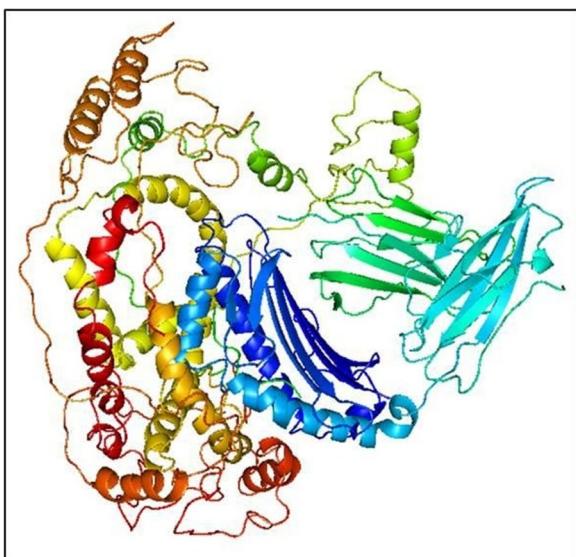
**Figure 2.** Representative arrangement of HTL, CTL and BCL epitopes of NP and GPC of CCHF virus with appropriate linkers in the multiepitope vaccine construct.



**Figure 3.** Secondary structure of multi epitope vaccine construct representing the arrangement of  $\alpha$ -helix (51.5%),  $\beta$ -strands (6%) and coils (42.5%).



**Figure 4.** Modeling and Validation of tertiary structure of Multiepitope vaccine construct a) Tertiary structure of vaccine construct with Phyre2.0 after refinement b) Tertiary structure validation with Ramachandran plot using PROCHECK with 94.2% residues in favored region and 5.8% in outlier region.



**Figure 5.** Protein-protein molecular docking between multiepitope vaccine and HLA allele HLA A02 using FireDock (Vaccine construct shown in Rainbow color and HLA molecule is shown in spring green and electric blue color).

which defines the efficacy and property of molecule. Structure was modeled with high confidence at net score of 51.098 and p-value 4e-04. A total of 243 amino acids (42.5%) are involved in coil formation, 294 amino acids (51.5%) are involved in alpha helix formation, while β-strands are formed by 34 amino acids (6%) (Fig 3).

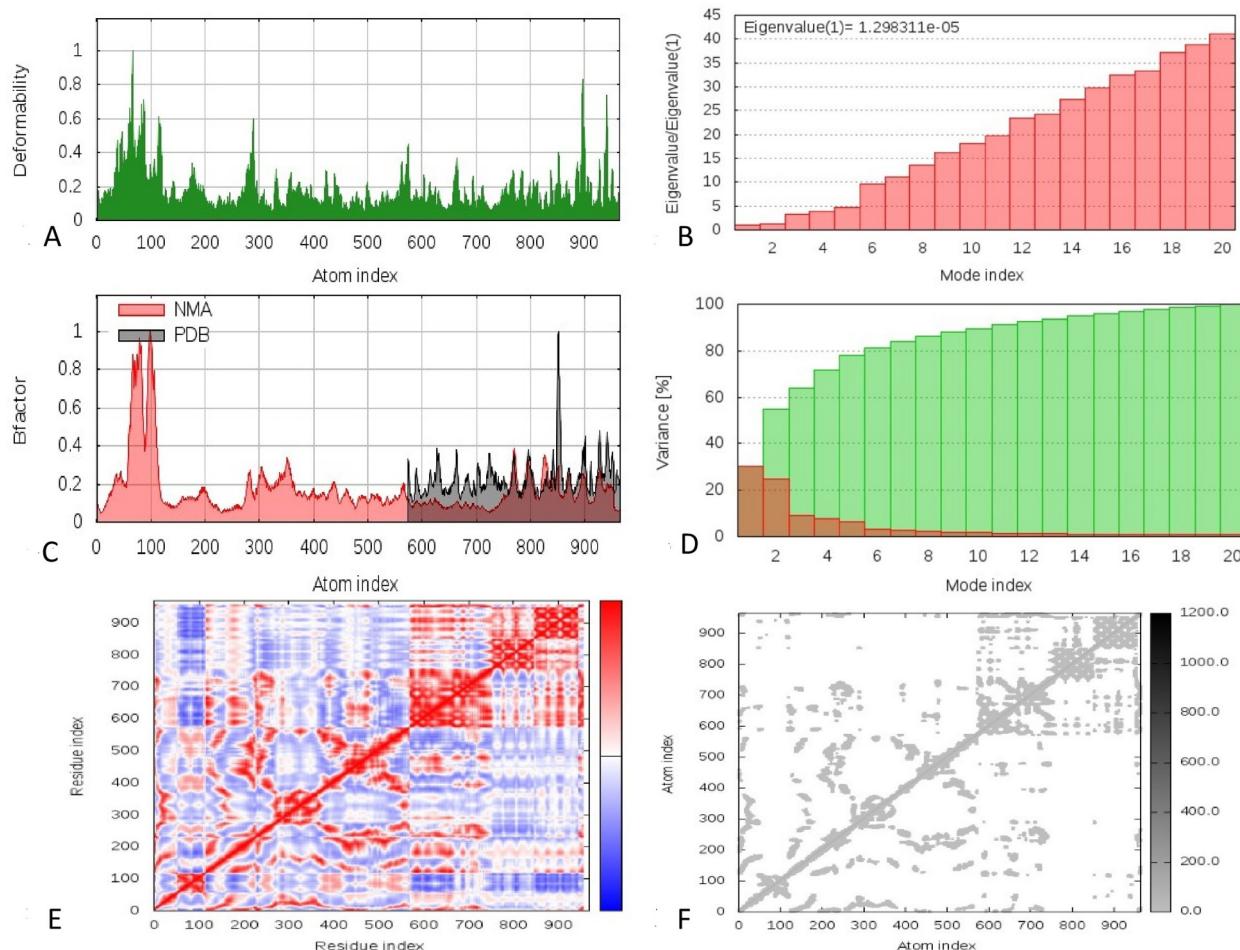
Tertiary structure was designed using Phyre 2.0 server (Fig 4a). Three templates were finalized to model the protein based on heuristics

to maximize percentage identity, confidence index and alignment coverage. Phyre2.0 used RNA binding protein from Crimean-Congo hemorrhagic fever virus protein (PDB ID: C3U3I) at 98% confidence score as best template for modeling.

The predicted tertiary structure was validated with Ramachandran plot using PROCHECK programme (<https://servicesn.mbi.ucla.edu/PROCHECK/>). The structure reveals the residues in most favored regions are 326 (94.2%), in disallowed regions are 20 (5.8%), whereas the number of glycine and proline residues are 46 and 12 respectively which reflects it to be a good quality model and confirmed the stability of designed multi epitope vaccine molecule (Fig 4b).

#### 4.12. Molecular docking with HLA molecules

Molecular docking was carried out in order to evaluate the interaction between HLA A02 and multi epitope vaccine construct using online server PatchDock 4.0 which was further refined with FireDock. The docking of multiepitope protein with HLA A02 allele complex was carried out using the HLA A02 structure retrieved from Protein Data Bank (PDB ID: 1AKJ) with 2.65 Å resolution [Pubmed Id: 9177355] at RMSD clustering value of 4.0. The structure consisting identical chain A and B each with 571 amino acids. Ligands and water molecule was removed from the structure before docking using CHIMERA software. Chain A and ligand HLA A02 molecule were selected for docking and molecular dynamics simulations for different structure trajectory files. FireDock refined output has shown the lowest global energy of 10.55 for solution number 3 and rank 1, attractive Vanderwal forces of -0.20 and ACE of 0.22. The lowest energies of this model favor the highest binding affinity between HLA A02 and vaccine construct (Fig 5).



**Figure 6.** Molecular dynamics simulation of multiepitope vaccine construct using iMOD. Protein structure stability was estimated via A) Deformities of atoms in torsional spaces B) Eigen value C) Variation of atoms due to temperature (B factor) D) Variance E) Covariance map F) Elasticity network.

#### 4.13. Molecular dynamic simulation

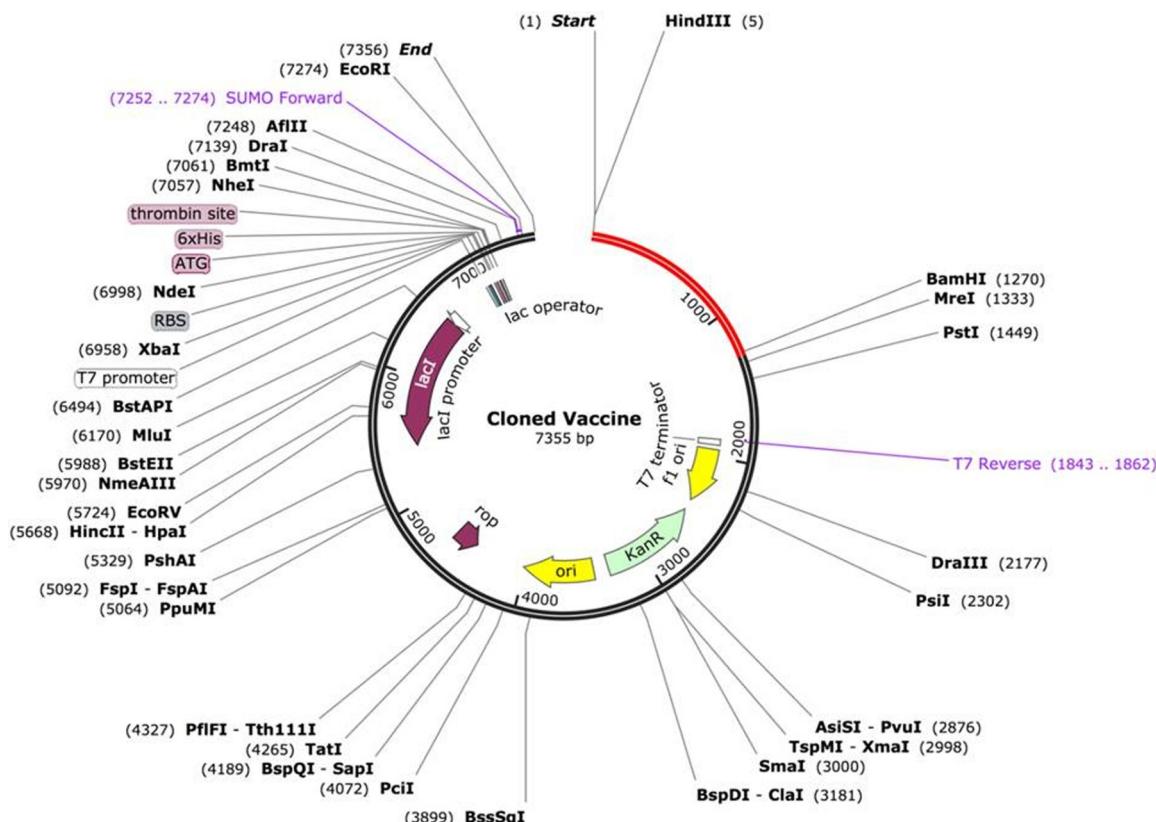
The docked complex of vaccine construct including novel multi-epitopes of CCHFV, adjuvant and tick salivary protein with HLA molecule were used for molecular dynamic simulation. MDS was performed to explore more detailed interaction paradigm followed by energy minimization. MD simulation was studied using iMODS online server with NMA to study the mobility of protein atoms in torsional space. Individual distortion of residue defines the deformability in complex denoted by hinges in structure, very few hinges are observed in the structure denotes the stability of complex rather than deformability (Fig 6A). The depicted eigen value of the complex was  $1.298311e^{-05}$  that defines the motion stiffness of molecule (Fig 6B). B factor defines the displacement of atom may be due to temperature dependent atomic vibrations in lattice from its mean position (Fig 6C). B factor also relates to the RMS fluctuations of the molecule that depicts the flexible regions of the protein. Ca atoms were used for the calculation. These residues fluctuated in stable region during simulation due to low RMSF value. Fig 6D shows the variance associated to equilibrium motion that is inversely related to eigen value. Covariance map was also generated that reflects the coupling between pair of residues, the colors red, white and blue depicts the correlated, uncorrelated and anti correlated fashion of residues (Fig 6E). Elastic network model was also generated that represent the dark grey dots as stiffer part of the molecule (Fig 6F), all these parameter represents the stability of simulated molecule with human HLA allele mimicking *in-vivo* conditions.

#### 4.14. In silico cloning and optimization for expression

Codon Usage was adapted to *Escherichia coli* (*strain ATCC 27325 / DSM 5911 / W3110 / K12*) using JCAT server and found that the value of codon adaptive index was 0.99 which depicts the higher protein expression in *E. coli* of the target gene. The GC-content of the improved sequence was 49.67% that should ideally lie between 30-70% representing good expression efficacy in *E. coli*. Simultaneously, *BamHI* and *HindIII* restriction sites were introduced at N and C terminal of the codon and the gene was cloned into the expression vector pET sumo. The target sequence is also tagged with 6X histidine residues for efficient protein purification. Finally, a cloned construct was generated with a sequence length of 7355 base pair (Fig 7).

#### 5. Conclusion

Crimean-Congo hemorrhagic fever virus posses a significant public health threat. It is known to cause severe hemorrhagic fever with higher fatality rates in humans. CCHFV is recently listed in WHO top priority agent under “WHO R&D Blueprint”, having potential to cause near outbreaks (WHO 2020). There is an urgent need to develop an effective vaccine especially in the absence of therapeutics and other prophylaxis. A recent study on development of multiepitope CCHF vaccine based on glycoprotein and nucleoprotein was reported (Nosrati et al., 2019). However, the candidate vaccine is limited by its composition, where only 7 epitopes were included. One of the major drawback associated with failure of subunit vaccine are attributed to presence of limited number of epitopes, in contrast to whole viral vaccine. In present study,



**Figure 7.** *In-silico* cloning map representing codon optimized multi epitope vaccine sequence (red) into pET Sumo vector.

in order to elicit strong immune response, 24 immuno dominant epitopes representing structural proteins of virus (i.e. nucleoprotein being highly conserved and glycoprotein) were incorporated in the vaccine candidate. The study also reports novel B cell epitopes, validated with CCHF positive sera. All the potential BCL, HTL and CTL epitopes were merged along with efficient adjuvant using appropriate linkers. Further, subolesin, a highly conserved and protective tick salivary protein was also added in multiepitope vaccine construct in order to provide protection against tick bite which proportionally target and kill the tick population regardless of host and virus. The final construct was examined for its antigenicity and allergenicity and found to be non allergen and highly immunogen. Later, the vaccine candidate was also docked and simulated with highly circulating MHC allele in majority of human population and studied for its stability mimicking *in-vivo* conditions. All these *in silico* analysis revealed that the vaccine construct can serve as a potential candidate against CCHF viral infection. In addition, this study also reports for the first time the potential utility of this dual acting multi epitope vaccine candidate in blocking the transmission of tick-borne viruses as well as tick infestations in vertebrate hosts. Efforts are needed to lower the disease burden with such safe vaccine alternative in humans as well as animals.

## Data Availability

All relevant data are available within the manuscript and its associated Supporting Information files.

## CRediT authorship contribution statement

**Neha Shrivastava:** Conceptualization, Methodology, Writing - original draft. **Ankit Verma:** Validation. **Paban Kumar Dash:** Conceptualization, Methodology, Writing - review & editing.

## Declaration of Competing Interest

The authors declare no competing interests.

## Acknowledgments

The authors are thankful to Dr D. K. Dubey, Director, DRDE, Gwalior, for his keen interest, constant support, and provision of necessary facilities for this study. This manuscript is assigned DRDE accession no. DRDE/VIRO/27/2019. The first author would also like to thank Department of Science and technology (DST), Govt. of India for providing INSPIRE fellowship.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.ejps.2020.105396](https://doi.org/10.1016/j.ejps.2020.105396).

## References

- Al-Abri, S., et al., 2017. Current status of Crimean-Congo hemorrhagic fever in the World Health Organization Eastern Mediterranean Region: issues, challenges, and future directions. *Int J Infect Dis* 58, 82–89.
- Messina, J.P., et al., 2015. The global distribution of Crimean-Congo hemorrhagic fever. *T Roy Soc Trop Med H* 109 (8), 503–513.
- Bente, D.A., et al., 2013. Crimean-Congo hemorrhagic fever: History, epidemiology, pathogenesis, clinical syndrome and genetic diversity. *Antivir Res* 100 (1), 159–189.
- Adams, M.J., et al., 2017. Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses. *Arch Virol* 162 (8), 2505–2538. <https://doi.org/10.1007/s00705-017-3358-5>.
- Duh, D., et al., 2008. The complete genome sequence of a Crimean-Congo Hemorrhagic Fever virus isolated from an endemic region in Kosovo. *Virol J* 5 (1), 7. <https://doi.org/10.1186/1743-422x-5-7>.
- Mourya, D.T., 2015. Cross-sectional Serosurvey of Crimean-Congo Hemorrhagic Fever Virus IgG in Livestock, India, 2013–2014. *Emerging Infectious Diseases* 21 (10), 1837–1839. <https://doi.org/10.3201/eid2110.141961>.

- Aradaib, I.E., et al., 2010. Nosocomial Outbreak of Crimean-Congo Hemorrhagic Fever. *Emerg Infect Dis* 16 (5), 837–839. <https://doi.org/10.3201/eid1605.091815>.
- Jamshidi1, M., et al., 2017. Crimean-Congo Hemorrhagic Fever: Report of Three Cases from Iran. *J Infect Dis Ther* 5, 1.
- Dowall, S., et al., 2015. A Crimean-Congo hemorrhagic fever (CCHF) viral vaccine expressing nucleoprotein is immunogenic but fails to confer protection against lethal disease. *Hum Vaccin Immunothe R*.2015 12 (2), 519–527. <https://doi.org/10.1080/21645515.2015.1078045>.
- Papa, A., et al., 2004. Crimean-Congo hemorrhagic fever in Bulgaria.”. *Emerg Infect Dis* 10 (8), 1465–1467.
- Zivcec, M., et al., 2016. Molecular Insights into Crimean-Congo Hemorrhagic Fever Virus. *Viruses* 8, 106.
- Hewson, R., et al., 2004. Crimean-Congo haemorrhagic fever virus: sequence analysis of the small RNA segments from a collection of viruses worldwide. *Virus Res* 102 (2), 185–189. <https://doi.org/10.1016/j.virusres.12.035>. PMID:15084400.
- Šimo, L., et al., 2017. The Essential Role of Tick Salivary Glands and Saliva in Tick Feeding and Pathogen Transmission. *Front Cell Infect Mi*. 7.
- Shakya, M., et al., 2014. Subolesin: A candidate vaccine antigen for the control of cattle tick infestations in Indian situation. *Vaccine*, 32 (28), 3488–3494.
- De la Fuente, J., et al., 2011. Targeting arthropod subolesin/akirin for the development of a universal vaccine for control of vector infestations and pathogen transmission. *Vet Parasitol*. 2011 181 (1), 17–22. <https://doi.org/10.1016/j.vetpar.04.018>.
- Dickson, D.L., Turell, M.J., 1992. Replication and tissue tropisms of Crimean-Congo hemorrhagic fever virus in experimentally infected adult *Hyalomma truncatum* (Acar: Ixodidae). *J. Med. Entomol.* 29, 767–773. <https://doi.org/10.1093/jmedent/29.5.767>.
- Papa, A., Tsengouli, K., Tsioka, K., Mirazimi, A., 2017. Crimean-Congo Hemorrhagic Fever: Tick-Host-Virus Interactions. *Front. Cell. Infect. Microbiol* 7, 213. <https://doi.org/10.3389/fcimb.2017.00213>.
- Shrivastava, N., et al., 2019. Development of Multispecies Recombinant Nucleoprotein-Based Indirect ELISA for High-Throughput Screening of Crimean-Congo Hemorrhagic Fever Virus-Specific Antibodies. *Front. Microbiol*. 10, 1822. <https://doi.org/10.3389/fmicb.2019.01822>.
- Argos, P., 1990. An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. *J Mol Biol* 211, 943–958 [PubMed: 2313701].
- Rana, A., Akhter, Y., 2016. A multi-subunit based, thermodynamically stable model vaccine using combined immunoinformatics and protein structure based approach. *Immunobiology* 221 (4), 544–557. <https://doi.org/10.1016/j.imbio.2015.12.004>.
- Jennifer M. E., et al., 2000. Frequencies of HLA-A2 alleles in five U.S. population groups: Predominance of A\*02011 and identification of HLA-A\*0231. *Int J Infect Dis*. 61 (3), 334–340.
- Lopez-Blanco, JR, et al., 2014. iMODS: internal coordinates normal mode analysis server. *Nucleic Acids Res*. 42, W271–W276.
- Goedhals, D., et al., 2017. Long-lived CD8+ T cell responses following Crimean-Congo haemorrhagic fever virus infection. *PLoS Negl Trop Dis* 11 (12), e0006149.
- Liu, D., et al., 2014. Fine Epitope Mapping of the Central Immunodominant Region of Nucleoprotein from Crimean-Congo Hemorrhagic Fever Virus (CCHFV). *PLoS ONE* 9 (11), e108419.
- Fritzen, A., et al., 2018. Epitope-mapping of the glycoprotein from Crimean-Congo hemorrhagic fever virus using a microarray approach. *PLoS Negl Trop Dis* 12 (7), e0006598.
- Chen, X., Zaro, J.L., Shen, W.-C., 2013. Fusion protein linkers: Property, design and functionality. *Advanced Drug Delivery Reviews* 65 (10), 1357–1369. <https://doi.org/10.1016/j.addr.2012.09.039>.
- Jain, N., Smith, S.W., Ghone, S., Tomczuk, B., 2015. Current ADC Linker Chemistry. *Pharmaceutical research* 32 (11), 3526–3540. <https://doi.org/10.1007/s11095-015-1657-7>.
- Olafsdottir, T.A., et al., 2009. IC31®, a Two-Component Novel Adjuvant Mixed with a Conjugate Vaccine Enhances Protective Immunity against Pneumococcal Disease in Neonatal Mice. *Scand J Immunol* 69 (3), 194–202.
- Fritz, J.H., et al., 2004. The artificial antimicrobial peptide KLKLLLLKLK induces predominantly a TH2-type immune response to co-injected antigens. *Vaccine* 22, 3274–3284.
- Schellack, C., et al., 2006. IC31, a novel adjuvant signaling via TLR9, induces potent cellular and humoral immune responses. *Vaccine*.2006 24, 5461–5472.
- Zaharieva, N., et al., 2017. Immunogenicity Prediction by VaxiJen: A Ten Year Overview. *J Proteomics Bioinform*. 10, 298–310.
- Nezafat, N., et al., 2017. Designing an efficient multi-epitope oral vaccine against *Helicobacter pylori* using immunoinformatics and structural vaccinology approaches. *Molecular BioSystems* 13 (4), 699–713. <https://doi.org/10.1039/c6mb00772d>.
- Jones, DT., 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292, 195–202.
- WHO: 2020 A research and development Blueprint for action to prevent epidemics for CCHF:<https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts>.
- Nosrati, M., et al., 2019. Towards the first multi-epitope recombinant vaccine against Crimean-Congo hemorrhagic fever virus: A computer-aided vaccine design approach. *Journal of Biomedical Informatics*. 103160.