

Identification of highly conserved regions in L-segment of Crimean–Congo hemorrhagic fever virus and immunoinformatic prediction about potential novel vaccine

Arafat Rahman Oany¹
Shah Adil Ishtiyag Ahmad¹
Mohammad Uzzal Hossain¹
Tahmina Pervin Jyoti²

¹Department of Biotechnology and Genetic Engineering, Faculty of Life Science, Mawlana Bhashani Science and Technology University, Santosh, Tangail, Bangladesh; ²Biotechnology and Genetic Engineering Discipline, Life Science School, Khulna University, Khulna, Bangladesh

Abstract: Crimean–Congo hemorrhagic fever (CCHF) is a tick-borne zoonotic viral disease with a disease fatality rate between 15% and 70%. Despite the wide range of distribution, the virus (CCHFV) is basically endemic in Africa, Asia, eastern Europe, and the Middle East. Acute febrile illness associated with petechiae, disseminated intravascular coagulation, and multiple-organ failure are the main symptoms of the disease. With all these fatal effects, CCHFV is considered a huge threat as no successful therapeutic approach is currently available for the treatment of this disease. In the present study, we have used the immunoinformatics approach to design a potential epitope-based vaccine against the RNA-dependent RNA polymerase-L of CCHFV. Both the T-cell and B-cell epitopes were assessed, and the epitope “DCSSTPPDR” was found to be the most potential one, with 100% conservancy among all the strains of CCHFV. The epitope was also found to interact with both type I and II major histocompatibility complex molecules and is considered nonallergenic as well. In vivo study of our proposed peptide is advised for novel universal vaccine production, which might be an effective path to prevent CCHF disease.

Keywords: single-stranded RNA, immunoinformatics, RNA-dependent RNA polymerase, epitope

Introduction

Crimean–Congo hemorrhagic fever virus (CCHFV) is the causative agent of the hemorrhagic fever, which was first described in Crimea in 1945.¹ The virus consists of a negative-sense single-stranded RNA, and its tripartite genome comprises the small (S), medium (M), and large (L) segments, which encode the viral nucleocapsid (N), glycoprotein precursor, and polymerase-L proteins, respectively.² Africa, Asia, eastern Europe, and the Middle East are the hotspots of this viral infection, with a high fatality rate between 15% and 70%.^{3,4} In case of human infection, the progression of this disease is very rapid and it causes acute febrile illness, associated with petechiae, ecchymosis, disseminated intravascular coagulation, and multiple-organ failure.⁵ A wide range of ticks that belong to the *Hyalomma* genus are considered the main vectors of this viral disease and the disease is spread through the bite of these ticks to humans or other domestic animals.⁶ Due to inadequate hospital care, the disease is also spread nosocomially and a case reported that, in South Africa, about 33% medical personnel are infected through needlestick injury.⁷ Alternate routes of infection, such as aerosol and droplet respiratory route of infection, were also suspected for several cases in Russia.⁸ With a broad geographic distribution,

Correspondence: Arafat Rahman Oany
Department of Biotechnology and Genetic Engineering, Faculty of Life Science, Mawlana Bhashani Science and Technology University, Santosh, Tangail 1902, Bangladesh
Tel +880 15 5881 9130
Email arafatr@outlook.com

this virus already covers >30 countries across Africa, southeastern Europe, the Middle East, western Asia, and more recently, some places of southwestern Europe, particularly Spain.^{9–11} The wide distribution of this virus indicates the ability of the tick hosts to adapt across different environmental and geographic regions.¹² Despite the outbreaks, there is no reliable vaccine or drug for the treatment of CCHFV infection in animals or humans.¹³ As a single-stranded-RNA-containing genome and due to the error-prone nature of its polymerase, the CCHFV incorporates random mutations into the genome. The rate of recombination in its RNA is also very high. As a consequence, the development of a vaccine or antiviral drug against CCHFV is very difficult.¹⁴

The design of epitope-based vaccines against some deadly viruses has gained much popularity due to its increased potency and safety.^{15,16} The application of bioinformatics in immunology, which is also termed immunoinformatics,

is now widely accepted. Immunoinformatics can assist in designing new vaccines through the identification of potential T-cell epitope, B-cell epitope, and human leukocyte antigen (HLA) ligands.^{17–19} This novel approach has proven its efficacy in the case of multiple sclerosis,²⁰ malaria,²¹ human immunodeficiency virus,²² and tuberculosis,²³ with desired results. In our present study, we have proposed the design of a potential conserved epitope candidate through the immunoinformatics approaches in order to minimize the fatal effects of the CCHFV, with the expectation of finding a novel candidate for the vaccine via wet laboratory validation.

Materials and methods

Sequence retrieval and conserved region identification

UniProtKB²⁴ database was used for the retrieval of the sequences of RNA-dependent RNA polymerase-L²⁵ and

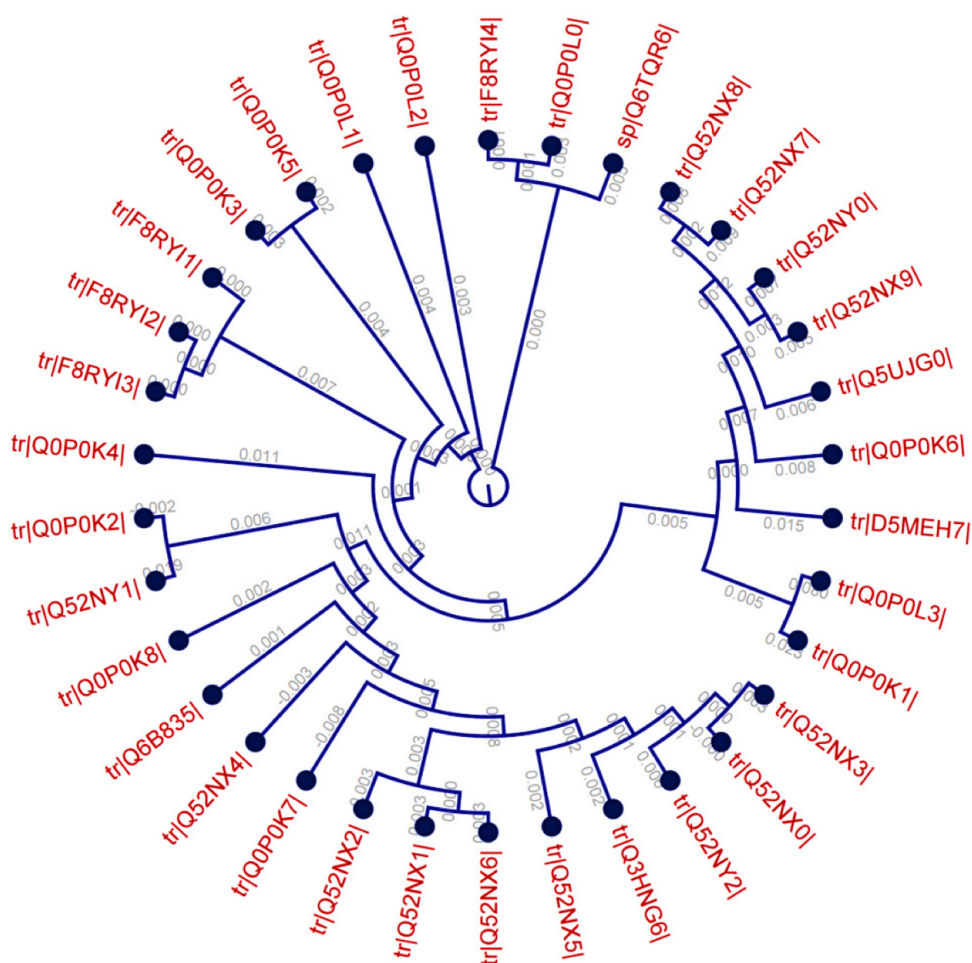


Figure 1 Phylogenetic tree showing the evolutionary divergence among the different RNA-dependent RNA polymerase-L molecules of the CCHFV.

Notes: Here, cladogram view is shown with appropriate distance among the different strains. The blue dotted view indicates the node of the tree.

Abbreviation: CCHFV, Crimean–Congo hemorrhagic fever virus.

Table 1 Antigenicity determination of the conserved peptide by Vaxijen server

Peptide number	Peptide	Region	Vaxijen score (threshold: 0.4)
1	LIQTLFPDKFEDFLDRTQLHPEFRDLTPDFSLTQKVHFKRNQIPSVEN VQISIDATLPESVEAVPVTERKMFPLPETPLSEVHSIERIMENFTRLM	3,563–3,658	0.6556
2	DYGERGIVEENHMKFSGEDQLETRQLLLVEVGFQTDIDGKIRT DHKKWKDILKLELLGIKCSFIACDCSSTPPDRWWI	3,694–3,773	0.4268
3	EDRVRVLKNSVSFLFNKLSRNSPTEVTDIVVGAISTQKVRYSY LKAGTATKTPVSTKDVLETWEK	3,775–3,838	0.3662
4	MKEHILNRPTGLTLPTSLEQAMRKGLEGVVISKEGSESCINMLK ENLDRITDEFERTKFKHELTQNITTSEKML	3,841–3,915	0.1044

the envelope glycoprotein²⁶ of the CCHFV in the FASTA protein format.

BioEdit v7.2.3 sequence alignment editor²⁷ was used for the identification of the conserved region among the sequences through multiple-sequence alignment (MSA) with ClustalW.²⁸ Finally, Jalview v2 tool²⁹ was used to retrieve the alignment and the CLC Sequence Viewer v7.0.2 (<http://www.clcbio.com>) was used for analysis of the divergence among the different strains of the CCHFV.

Antigenicity determination of the conserved peptides

VaxiJen v2. 0, a Web-based server,³⁰ was used for the determination of the antigenicity of the conserved sequences.

Herein, we used the default parameters for the prediction, with a threshold value of 0.4.

T-cell epitope prediction

Two online servers were used for the prediction of the T-cell epitope. Initially, the NetCTL v1.2 server was used for the identification of the potential T-cell epitope.³¹ We used the default approach to predict the epitopes, including major histocompatibility complex class I (MHC-I) binding, proteasomal C terminal cleavage, and transporter of antigenic peptide (TAP) transport efficiency. The epitope prediction was restricted to 12 MHC-I supertypes. MHC-I binding and proteasomal cleavage were performed through artificial neural networks and the weight matrix was used

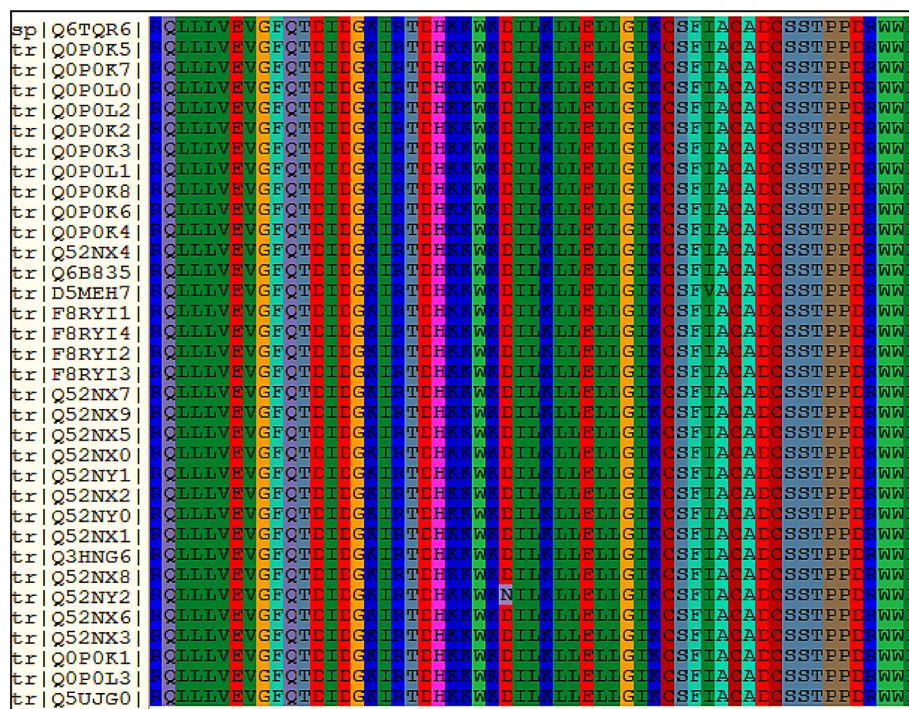


Figure 2 MSA of the conserved region of RNA-dependent RNA polymerase-L. Only the partial sequences containing the proposed epitope sequence are shown here.

Notes: Clustalx color is used here. Different colors indicate different amino acid residues.

Abbreviation: MSA, multiple-sequence alignment.

for estimating TAP transport efficiency. The threshold for epitope identification was set at 0.5 to maintain sensitivity and specificity of 0.89 and 0.94, respectively. Finally, CTLPred³² was implemented additionally for further confirmation about the prediction with default parameters.

MHC-I and MHC-II restriction analysis

T Cell Epitope Prediction Tools from Immune Epitope Database and Analysis Resource (IEDB-AR) was used for the prediction of MHC-I³³ and MHC-II^{34,35} binding of the peptide. The Stabilized Matrix Method³⁶ was used to calculate the half-maximal inhibitory concentration (IC₅₀) of peptide binding to MHC-I molecules from different prediction methods, with a preselected 9.0-mer epitope. In case of MHC-II binding analysis, the IEDB-recommended method was used for the specific HLA-DQ, HLA-DP, and HLA-DR loci. Here, we used specific peptides for the prediction of MHC-II interaction on the basis of the antigenic conservancy and MHC-I analysis.

B-cell epitope prediction

B-cell epitope initiates immunoresponse through the interaction with B lymphocytes and causes the differentiation of B lymphocytes into plasma and memory cells.³⁷ IEDB-AR hosts a number of Web-based tools for the prediction of B-cell epitope. Multiple tools, including the Kolaskar and Tongaonkar antigenicity scale,³⁸ Emimi surface accessibility prediction³⁹ and Bepipred linear epitope prediction analysis, were used for the B-cell epitope prediction with high accuracy.⁴⁰

Homology modeling and protein variability determination of the conserved region

Homology model of the conserved region was obtained by MODELLER v9,⁴¹ and the predicted model was assessed by PROCHECK⁴² and QMEAN⁴³ servers of the SWISS-MODEL Workspace.⁴⁴ For the disorder prediction

Table 2 Prediction of the T-cell epitope by NetCTL server on the basis of combined score

Number	Epitope	Combined score (nM)
1	CSSTPPDRW	1.7235
2	DVLETWEKM	1.6491
3	WEKMKHEIL	1.5089
4	ERTKFKHEL	1.4955
5	VPVTERKMF	1.4734

Table 3 Prediction of the T-cell epitope by CTLPred server

Epitope	Start position	Score (ANN/SVM)
LNRPTGLTL	246	1.00/-0.23718777
FSLTQKVHF	30	0.99/-0.23718777
ADCSSTPPD	164	0.99/-0.23718777
LSRNSPTEV	194	0.99/-0.23718777
TSLEQAMRK	256	0.99/-0.23718777

Abbreviations: ANN/SVM, artificial neural networks/support vector machines; CTLP, Cytotoxic T Lymphocyte Prediction (CTLPred).

among the amino acid sequences, DISOPRED v3⁴⁵ was used. The Protein variability server was used to calculate protein variability index using Shannon variability coefficient.⁴⁶

Allergenicity and epitope conservancy analysis

In order to predict the allergenicity of the proposed epitopes with high accuracy, a Web-based server AlgPred⁴⁷ was used. Herein, we used a hybrid prediction (SVMc + IgEepitope + ARPs BLAST + MAST) approach to predict the allergenicity with an accuracy of about 86% at a threshold value of -0.4. The prediction procedure follows the guidelines of the Food and Agriculture Organization/World Health Organization, 2003. A Web-based tool from IEDB-AR⁴⁸ was used in order to identify the specific conservancy of the proposed epitopes.

Table 4 MHC-I and MHC-II interaction of the proposed sequence by IEDB-AR

Epitope	MHC interaction
MHC-I interaction analysis	
DCSSTPPDR	HLA-C*12:03, HLA-C*03:03, HLA-C*14:02, HLA-B*58:01, HLA-C*12:03, HLA-C*03:03, HLA-C*15:02, HLA-B*57:01
MHC-II interaction analysis	
FIACDCSSTPPDRW	HLA-DRB1*04:01, HLA-DRB1*03:01, HLA-DRB1*08:02, HLA-DRB3*01:01, DQB1*03:01, HLA-DRB1*09:01, HLA-DRB5*01:01, HLA-DRB1*04:05, DQB1*04:02, HLA-DRB1*11:01, HLA-DRB1*07:01, DQB1*02:01, HLA-DRB1*13:02, DQB1*03:02, DQB1*05:01, HLA-DRB4*01:01, HLA-DRB1*01:01, HLA-DRB3*02:02, DPB1*04:01, DQB1*06:02, DPB1*14:01, HLA-DRB1*15:01, DPB1*02:01, HLA-DRB1*12:01, DPB1*04:02, DPB1*05:01, DPB1*01:01

Abbreviations: IEDB-AR, Immune Epitope Database and Analysis Resource; MHC, major histocompatibility complex.

Table 5 Kolaskar and Tongaonkar antigenicity analysis

Number	Start position	End position	Peptide	Peptide length
1	16	23	RTQLHPEF	8
2	28	39	PDFSLTQKVHFK	12
3	43	67	IPSVENVQISIDVTLPEVVEAVPVT	25
4	72	85	FPLPETPLSEVHSI	14
5	117	130	QSAVEHESPSISAF	14
6	154	163	TRQLLLVEVG	10
7	182	204	ILKLELLGIKCSFIACADCST	23
8	215	229	RVRVLKNSVSFLFNK	15
9	238	246	VTDIVVGAI	9
10	248	258	TQKVRSYLKAG	11
11	262	273	KTPVSTKDVLET	12
12	287	295	GLTLPASLE	9
13	302	310	LVEGVVISK	9
14	314	319	ESCINM	6

Results

Analysis of the retrieved sequences and their divergence

A total of 80 envelope glycoproteins and 34 RNA-dependent RNA polymerase-L molecules from different variants of the CCHFV were retrieved from the UniProt database. The MSA of two different types of proteins was retrieved from BioEdit tool through ClustalW with 1,000 bootstrap replicates (Figures S1 and S2). Conserved regions were grouped for the antigenic property analysis. CLC Sequence Viewer was used to construct phylograms for both proteins from the MSA obtained from BioEdit, in order to analyze the divergence

among the retrieved sequences. Phylogram of RNA-dependent RNA polymerase-L is depicted in Figure 1 and the phylogram for the envelope glycoprotein is provided in Figure S3.

Antigenic peptide identification

Initially, the conserved sequences (MSA number: 3,563–3,915) of RNA-dependent RNA polymerase-L were separated into four conserved peptides, according to their continuity in the MSA. Then the VaxiJen v2.0 server was used to predict the antigenicity of all the grouped conserved peptides (Table 1). On the basis of the VaxiJen score, the top two conserved peptides (MSA number: 3,563–3,658 and 3,694–3,773) were

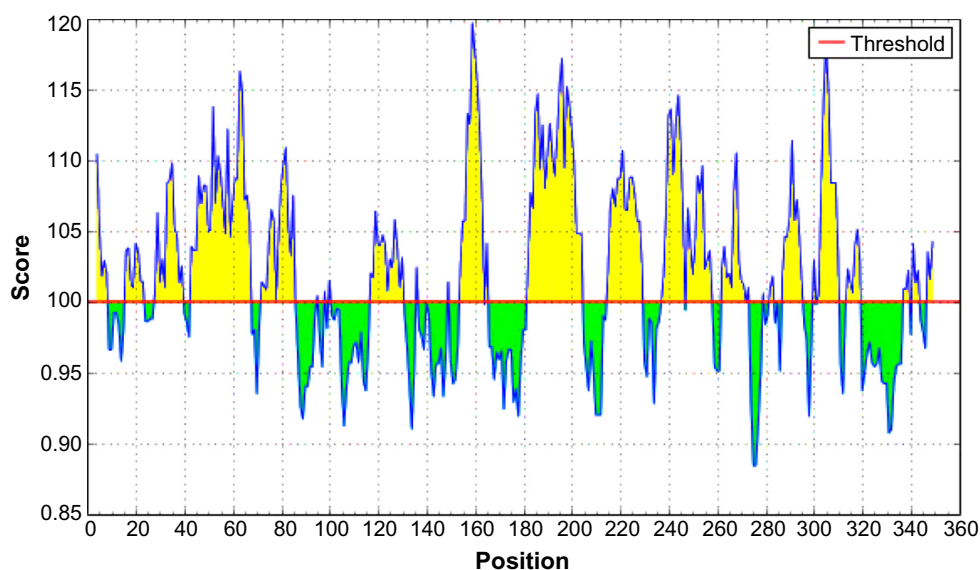


Figure 3 Kolaskar and Tongaonkar antigenicity prediction of the conserved peptide, ranging from 3,563 to 3,915 MSA number.

Notes: The region from 197 to 202 is the proposed epitope. The X- and Y-axes represent the sequence position and antigenic propensity score, respectively. The threshold value is 1.0. The regions above the threshold are antigenic, shown in yellow.

Abbreviation: MSA, multiple-sequence alignment.

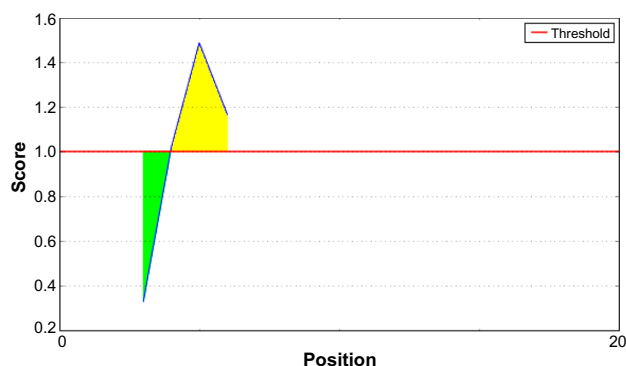


Figure 4 Emini surface accessibility prediction of the proposed epitope, with a minimum propensity score of 0.327 and maximum score of 1.488.

Notes: The X- and Y-axes represent the sequence position and surface probability, respectively. The threshold value is 1.0. The regions above the threshold are antigenic, shown in yellow.

selected for further analysis. The peptide with second-best VaxiJen score (MSA: 3,694–3,773) showed better results during further analysis of T-cell epitope identification, MHC interaction analysis, and B-cell epitope identification. The conservancy of this region among the different viral strains is partially depicted in Figure 2.

T-cell epitope identification and MHC interaction analysis

NetCTL v1.2 server predicted the T-cell epitopes through the combined approach for the 12 MHC-I supertypes. On the basis of the combined score, five epitopes with top scores (Table 2) were selected for further analysis. CTLPred server also predicted the T-cell epitopes based on an approach that combined artificial neural networks and support vector machines (Table 3). From the analysis, the common epitope-containing peptide, which was predicted by both servers, was selected and used for the MHC-binding analysis.



Figure 5 Bepipred linear epitope prediction of the proposed epitope with a minimum propensity score of 1.631 and maximum score of 2.094.

Notes: The X- and Y-axes represent the sequence position and propensity score, respectively. The threshold is 0.350. The regions having beta turns are shown in yellow. The highest peak region indicates the most potent B-cell epitope.

MHC-I-binding prediction, which was run through the Stabilized Matrix Method, predicted a wide range of MHC-I allele interactions for the proposed T-cell epitopes. The MHC-I alleles for which the epitope showed higher affinity ($IC_{50} < 200$ nM) are listed in Table 4. The output of the MHC-II interaction analysis is also shown in Table 4.

B-cell epitope identification

Amino acid-based methods were used for the prediction of potential B-cell epitope. Antigenic property of the peptides was assessed using the Kolaskar and Tongaonkar antigenicity scale. The average antigenic propensity score of the protein was 1.019, with a maximum of 1.197 and a minimum of 0.884. The threshold value for antigenic determination of the peptide was 1.0, where all epitopes with values > 1.0 were potential antigenic determinants. We found that 14 epitopes satisfied the threshold value set prior to the analysis, and they had the potential to express the B-cell response. The results are summarized in Table 5 and Figure 3. Peptides with the potential to function as B-cell epitope must be surface accessible. For this reason, Emini surface accessibility prediction was employed, with a maximum propensity score of 4.946 at threshold 1.0. Results are summarized in Figure 4. Finally, Bepipred linear epitope prediction tool predicted the linear B-cell epitope with the most reliable results of prediction. Results are shown in Figure 5.

Structure analysis and protein variability determination

Homology model of the conserved region was obtained by the MODELLER software, which is shown in Figure 6A. PROCHECK server validated the stereochemical quality of the model through Ramachandran Plot (Figure 6B), and Qmean server also assessed the tertiary structure, with a Qmean6 score of 0.268. DISOPRED v3 server predicted the disorder of the conserved peptide in order to get insight about the disorder among the conserved sequences, which is depicted in Figure 7. Protein variability server predicted the variability of the conserved region of the RNA-dependent RNA polymerase-L (Figure 8) to ensure that the proposed epitope is within the invariable region.

Epitope conservancy and allergenicity analysis

IEDB conservancy analysis tool analyzed the epitope conservancy of the proposed epitopes that are shown in Table 6. AlgPred predicted the allergenicity of the epitopes based on

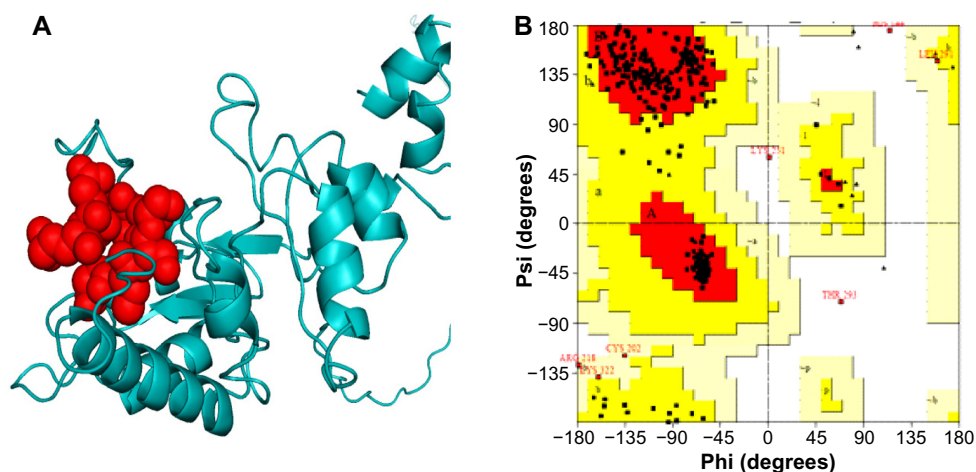


Figure 6 Three-dimensional structure prediction and validation.

Notes: (A) Three-dimensional model of the conserved region. Here, the epitope “DCSSTPPDR” is shown spherically. The outside location of the epitope indicates its surface accessibility. (B) Ramachandran plot of the predicted model shows that most of the residues are in the allowed region of the plot, proving the validity of the model.

amino acid composition. The prediction score of AlgPred for the two epitopes in combination was 0.49752311 at threshold of -0.4 .

Discussion

With a widely distributed endemically affected region and a randomly mutated genome, CCHFV imposes a great challenge to researchers in developing a successful therapeutic approach against it. The ability of an epitope-based vaccine to stimulate an effective specific immune response with a minute structure and without any unexpected side effects has made it a good choice for vaccine development.⁴⁹ In this instance, we started with the

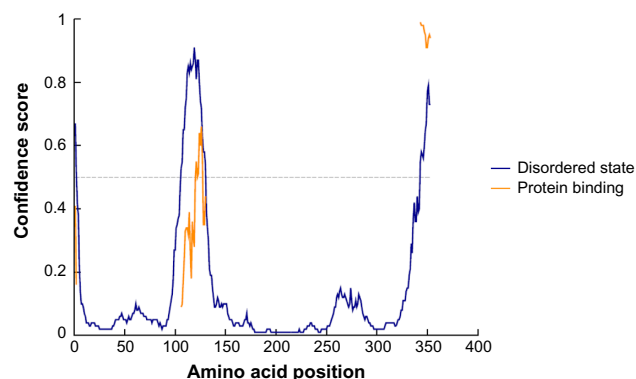


Figure 7 Disorder prediction of the conserved antigenic amino acid sequences. Here, our proposed epitope lies outside (197–202) of the disordered region to secure its potentiality as an effective epitope.

Notes: Amino acids in the input sequence are considered disordered when the blue line is above the gray dashed line, that is, when the confidence score is >0.5 . The orange line shows the confidence score of the disordered protein-binding residue predictions.

preferable target, namely, the envelope glycoprotein, but failed to identify any unique conserved region (Figure S1; multiple-sequence alignment of the envelope glycoprotein of CCHFV.) to design a peptide vaccine against the envelope glycoprotein. This was also revealed by phylogeny analysis, which is shown in Figure S3; phylogenetic tree showing the evolutionary divergence among the different envelope glycoproteins of CCHFV. RNA-dependent RNA polymerase, a product of the L-segment of the genome, comprises a unique conserved region among all the available strains of CCHFV (Figure S2; multiple-sequence alignment of the RNA-dependent RNA polymerase-L of CCHFV). This was the pedestal to think about a novel vaccine candidate. To ensure a firm immune response, we looked for the activation of both T-cell and B-cell immunity with a single epitope.⁵⁰ Antigenicity of the conserved peptides indicated their ability to provoke potential immune response and they were used for further analysis involving T-cell epitope prediction. Through the analysis of the output of both NetCTL and CTLPred, it was found that the epitope “DCSSTPPDR” would be the best candidate for the activation of T-cell immunity with potential antigenicity. Analysis of the MHC ligands for both type I and II revealed that the core epitope “DCSSTPPDR” would interact with the highest number of HLA molecules and that it would support the MHC molecules to present the epitope on the T-cell surface. The complete peptide for MHC-II restriction was FIACADCSSTPPDRW. The peptide DCSSTPPDR was also found to be the most potential candidate to raise B-cell immune response by amino acid-based B-cell epitope

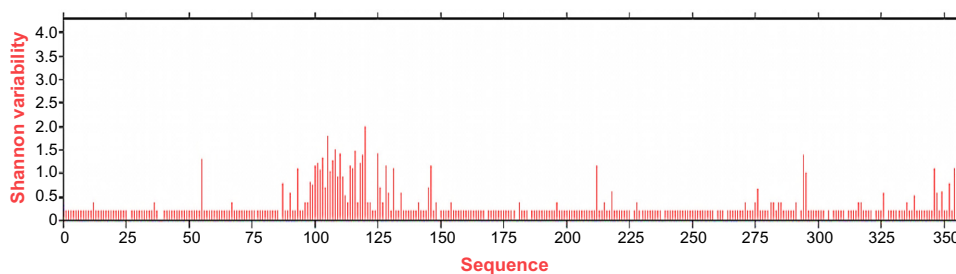


Figure 8 Protein variability index of the conserved peptides of all the sequences. The prediction suggests that our proposed epitope (197–202) falls in the invariable region.

Notes: The conservancy threshold was 1.0 in this analysis. The X-axis indicates the amino acid positions in the sequences and the Y-axis indicates the Shannon variability score.

prediction, including Kolaskar and Tongaonkar antigenicity scale, Emini surface accessibility prediction, and Bepipred linear epitope prediction.

In order to see the exact location of the proposed epitope of the protein (shown spherically in Figure 6A), the three-dimensional structure of the conserved peptide was modeled. This predicted model was validated with Ramachandran Plot (Figure 6B), whereby 89.8% amino acid residues were found within the favored region. The disorderliness of the peptide remains outside of the proposed epitope region, which would secure the functioning of our predicted epitope (Figure 7).

The most important feature of an epitope enabling its use as a vaccine is its conservancy. Conservancy analysis of the proposed epitope found 100% conservancy among all the available sequences. Another important criterion of the peptide vaccine is its allergenicity⁵¹; our proposed epitope was examined *in silico* and found to be nonallergenic in nature. A recent study in designing a vaccine against CCHFV targeting the envelope glycoprotein showed a high success rate in a mouse model.⁵² But because it is a structural protein, the rate of mutation is higher than that of nonstructural proteins such as RNA polymerase,⁵³ a phenomenon that has been evidenced in our study through sequence analysis. As our proposed epitope is shown to be 100% conserved among different CCHF strains, we suggest that it will be the best possible candidate for vaccine designing.

Table 6 Epitope conservancy analysis

Peptide sequence	Peptide length	Percentage of protein sequence match	Maximum identity
DCSSTPPDR	9	100% (34/34)	100%
FIACADCSTPPDRW	15	100% (34/34)	100%

Epitope-based vaccine designing is now becoming more popular and already has been established for rhinovirus,⁵⁴ dengue virus,⁵⁵ human corona virus,⁵⁶ and some others. This type of work has also been proven *in vitro*.⁵⁷ So, considering the above analysis, we predict that our proposed epitope would also trigger an immune response *in vitro*.

Conclusion

The findings from this study denote that integrated computational approaches are very much effective for designing vaccine candidates against some deadly viruses such as the CCHFV, with the formally delineated experimental procedure. Thus, computational studies save both time and cost for researchers and can lead the wet laboratory work with higher possibilities of getting the desired outcome.

Disclosure

The authors report no conflicts of interests in this work.

References

- Cunha BA. *Tickborne Infectious Diseases: Diagnosis and Management*. Vol 24. Boca Raton, FL: CRC Press; 2000.
- Schmaljohn C, Hooper J. Bunyaviridae: the viruses and their replication. *Fields Virol*. 2001;2(2):20.
- Lacy MD, Smego R. Viral hemorrhagic fevers. *Adv Pediatr Infect Dis*. 1995;12:21–53.
- Centers for Disease Control and Prevention (CDC). Management of patients with suspected viral hemorrhagic fever. *MMWR Morb Mortal Wkly Rep*. 1988;37:1.
- Ergönül Ö. Crimean–Congo haemorrhagic fever. *Lancet Infect Dis*. 2006;6(4):203–214.
- Ergonul O. Crimean–Congo hemorrhagic fever virus: new outbreaks, new discoveries. *Curr Opin Virol*. 2012;2(2):215–220.
- Swanepoel R, Struthers J, Shepherd A, McGillivray G, Nel M, Jupp P. Crimean–Congo hemorrhagic fever in South Africa. *Am J Trop Med Hyg*. 1983;32(6):1407–1415.
- Whitehouse CA. Crimean–Congo hemorrhagic fever. *Antiviral Res*. 2004;64(3):145–160.
- David-West TS, Cooke AR, David-West AS. Seroepidemiology of Congo virus (related to the virus of Crimean haemorrhagic fever) in Nigeria. *Bull World Health Organ*. 1974;51(5):543.

10. Gonzalez J-P, LeGuenno B, Guillaud M, Wilson ML. A fatal case of Crimean–Congo haemorrhagic fever in Mauritania: virological and serological evidence suggesting epidemic transmission. *Trans R Soc Trop Med Hyg.* 1990;84(4):573–576.
11. Wilson ML, LeGuenno B, Guillaud M, Desoutter D, Gonzalez J-P, Camicas J-L. Distribution of Crimean–Congo hemorrhagic fever viral antibody in Senegal: environmental and vectorial correlates. *Am J Trop Med Hyg.* 1990;43(5):557–566.
12. Estrada-Peña A, Ruiz-Fons F, Acevedo P, Gortazar C, la Fuente J. Factors driving the circulation and possible expansion of Crimean–Congo haemorrhagic fever virus in the western Palearctic. *J Appl Microbiol.* 2013;114(1):278–286.
13. Hoogstraal H. The epidemiology of tick-borne Crimean–Congo hemorrhagic fever in Asia, Europe, and Africa. *J Med Entomol.* 1979;15(4):307–417.
14. Kraus AA, Mirazimi A. Molecular biology and pathogenesis of Crimean–Congo hemorrhagic fever virus. *Future Virol.* 2010;5(4):469–479.
15. Holland J, Domingo E. Origin and evolution of viruses. *Virus Genes.* 1998;16(1):13–21.
16. Sette A, Newman M, Livingston B, et al. Optimizing vaccine design for cellular processing, MHC binding and TCR recognition. *Tissue Antigens.* 2002;59(6):443–451.
17. Sette A, Fikes J. Epitope-based vaccines: an update on epitope identification, vaccine design and delivery. *Curr Opin Immunol.* 2003;15(4):461–470.
18. Poland GA, Ovsyannikova IG, Jacobson RM. Application of pharmacogenomics to vaccines. *Pharmacogenomics.* 2009;10(5):837–852.
19. Petrovsky N, Brusica V. Computational immunology: the coming of age. *Immunol Cell Biol.* 2002;80(3):248–254.
20. Bourdette DN, Edmonds E, Smith C, et al. A highly immunogenic trivalent T cell receptor peptide vaccine for multiple sclerosis. *Mult Scler.* 2005;11(5):552–561.
21. López JA, Weilenman C, Audran R, et al. A synthetic malaria vaccine elicits a potent CD8+ and CD4+ T lymphocyte immune response in humans. Implications for vaccination strategies. *Eur J Immunol.* 2001;31(7):1989–1998.
22. Wilson CC, McKinney D, Anders M, et al. Development of a DNA vaccine designed to induce cytotoxic T lymphocyte responses to multiple conserved epitopes in HIV-1. *J Immunol.* 2003;171(10):5611–5623.
23. Robinson HL, Amara RR. T cell vaccines for microbial infections. *Nat Med.* 2005;11:S25–S32.
24. Apweiler R, Bairoch A, Wu CH, et al. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 2004;32(suppl 1):D115–D119.
25. Kinsella E, Martin SG, Grolla A, Czub M, Feldmann H, Flick R. Sequence determination of the Crimean–Congo hemorrhagic fever virus L segment. *Virology.* 2004;321(1):23–28.
26. Sanchez AJ, Vincent MJ, Nichol ST. Characterization of the glycoproteins of Crimean–Congo hemorrhagic fever virus. *J Virol.* 2002;76(14):7263–7275.
27. Hall TA. *BioEdit: A User-Friendly Biological Sequence Alignment Editor and analysis Program for Windows 95/98/NT.* Paper presented at: Nucleic Acids Symposium Series; 1999; London.
28. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994;22(22):4673–4680.
29. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview version 2 – a multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 2009;25(9):1189–1191.
30. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics.* 2007;8(1):4.
31. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics.* 2007;8(1):424.
32. Bhasin M, Raghava G. Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine.* 2004;22(23):3195–3204.
33. Buus S, Lauemøller SL, Wørring P, et al. Sensitive quantitative predictions of peptide–MHC binding by a ‘Query by Committee’ artificial neural network approach. *Tissue Antigens.* 2003;62(5):378–384.
34. Wang P, Sidney J, Kim Y, et al. Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics.* 2010;11(1):568.
35. Wang P, Sidney J, Dow C, Mothe B, Sette A, Peters B. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS Comput Biol.* 2008;4(4):e1000048.
36. Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics.* 2005;6(1):132.
37. Nair DT, Singh K, Siddiqui Z, Nayak BP, Rao KV, Salunke DM. Epitope recognition by diverse antibodies suggests conformational convergence in an antibody response. *J Immunol.* 2002;168(5):2371–2382.
38. Kolaskar A, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett.* 1990;276(1):172–174.
39. Emimi EA, Hughes JV, Perlow D, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol.* 1985;55(3):836–839.
40. Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Res.* 2006;2(1):2.
41. Šali A, Potterton L, Yuan F, van Vlijmen H, Karplus M. Evaluation of comparative protein modeling by MODELLER. *Proteins.* 1995;23(3):318–326.
42. Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR.* 1996;8(4):477–486.
43. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics.* 2011;27(3):343–350.
44. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics.* 2006;22(2):195–201.
45. Ward JJ, McGuffin LJ, Bryson K, Buxton BF, Jones DT. The DISOPRED server for the prediction of protein disorder. *Bioinformatics.* 2004;20(13):2138–2139.
46. Garcia-Boronat M, Diez-Rivero CM, Reinherz EL, Reche PA. PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. *Nucleic Acids Res.* 2008;36(Suppl 2):W35–W41.
47. Saha S, Raghava G. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res.* 2006;34(Suppl 2):W202–W209.
48. Bui H-H, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics.* 2007;8(1):361.
49. Shrestha B, Diamond MS. Role of CD8+ T cells in control of West Nile virus infection. *J Virol.* 2004;78(15):8312–8321.
50. Arnon R. A novel approach to vaccine design – epitope-based vaccines. *FEBS J.* 2006;273:33–34.
51. McKeever TM, Lewis SA, Smith C, Hubbard R. Vaccination and allergic disease: a birth cohort study. *Am J Public Health.* 2004;94(6):985.
52. Buttigieg KR, Dowall SD, Findlay-Wilson S, et al. A novel vaccine against Crimean–Congo haemorrhagic fever protects 100% of animals against lethal challenge in a mouse model. *PLoS One.* 2014;9(3):e91516.
53. Lu R, Yu X, Wang W, et al. Characterization of human coronavirus etiology in Chinese adults with acute upper respiratory tract infection by real-time RT-PCR assays. *PLoS One.* 2012;7(6):e38638.
54. Lapelosa M, Gallicchio E, Arnold GF, Arnold E, Levy RM. In silico vaccine design based on molecular simulations of rhinovirus chimeras presenting HIV-1 gp41 epitopes. *J Mol Biol.* 2009;385(2):675–691.

55. Chakraborty S, Chakravorty R, Ahmed M, et al. A computational approach for identification of epitopes in dengue virus envelope protein: a step towards designing a universal dengue vaccine targeting endemic regions. *In Silico Biol.* 2010;10(5):235–246.
56. Oany AR, Emran AA, Jyoti TP. Design of an epitope-based peptide vaccine against spike protein of human corona virus: an in silico approach. *Drug Des Devel Ther.* 2014;8:1139–1149.
57. Khan MK, Zaman S, Chakraborty S, et al. In silico predicted mycobacterial epitope elicits in vitro T-cell responses. *Mol Immunol.* 2014;61(1):16–22.

Advances and Applications in Bioinformatics and Chemistry

Dovepress

Publish your work in this journal

Advances and Applications in Bioinformatics and Chemistry is an international, peer-reviewed open-access journal that publishes articles in the following fields: Computational biomodeling; Bioinformatics; Computational genomics; Molecular modeling; Protein structure modeling and structural genomics; Systems Biology; Computational Biochemistry;

Computational Biophysics; Chemoinformatics and Drug Design; In silico ADME/Tox prediction. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/advances-and-applications-in-bioinformatics-and-chemistry-journal>