Identification of epitopes for vaccine design against Crimean-Congo Haemorrhagic fever virus (CCHFV): An immuno-informatics approach

Tammanna R. Sahrawat[1] & Abhinav Kumar Checkervarty[2]
[1]Assistant Professor, [2]Post-graduate student
[1, 2]Centre for Systems Biology & Bioinformatics, UIEAST
Panjab University, Chandigarh, India

Received: December 07, 2018 Accepted: January 14, 2019

ABSTRACT: *Crimean-Congo haemorrhagic fever (CCHF) virus, is a tick and livestock animal borne viral disease that causes severe viral haemorrhagic fever. The present study was undertaken to identify candidate T-cell and B-cell epitopes for design of an effective vaccine against CCHFV using an immuno-informatics based approach. Proteome of CCHFV was retrieved from UniProtKB and analyzed using various in silico tools such as VaxiJen, NetCTL, AlgPred, AllerTOP and prediction tools of IEDB server for identification of T-cell and B-cell epitopes. The 9mer peptide in the region 1135-1144 was selected as a potential T-cell epitope as it interacted with highest number of HLA alleles having the highest population coverage in India. Structures of HLA alleles were retrieved from PDB, whereas epitope structure was predicted using PEP-FOLD followed by docking. Furthermore, B-cell epitope was predicted by analyzing and cross referencing various properties of the most antigenic protein sequence. It was found that peptide in the region 1121-1131 could elicit a significant B-cell immune response as this region was hydrophilic having constant beta turns and being surface accessible. Therefore, the region 1121-1144 of envelopment polyprotein of CCHFV contains potential candidates for T-cell and B-cell epitope based vaccine and may further be validated by in vitro and in vivo experiments. Thus immuno-informatics based approaches offer the possibility of design of potential vaccines for infectious pathogens in a timely manner.*

Key Words: *Immuno-informatics, T-cell epitope, B-cell epitope, vaccine, in silico.*

Introduction

The conventional approach for vaccine design is time consuming and involves tedious process of sub-culturing of a pathogen from a larger culture followed by its administration[1].This has led to the emergence of *in silico* approaches for identifying antigenic regions that can aid in vaccine design[2]. Novel bioinformatics approaches have been combined with cutting edge, high-dimensional assays to develop advanced vaccines for diseases such as malaria [3], sclerosis [4], tumors [5] and Ebola virus [6, 7].

Vaccinomics, integrates systems biology with fields of immunogentics/genomics and enhances our basic understanding and working of vaccine and its development. Immune profiling includes usage of many *in silico* tools and web servers for antigenicity analysis, epitope predictions, MHC-I binding analysis, docking studies and allergenicity assessment [8].

Crimean-Congo haemorrhagic fever virus (CCHFV) is a tick borne virus (*Nairovirus*) that belongs to *Bunyaviridae* family. It was first introduced in 1994 in Crimea and later identified in 1969 in Congo [9]. The virus can be transmitted from bites of ticks and interaction with the blood or tissues of infected animals during slaughter [10]. Host of CCHFV include animals (wild and domestic), for example goats, sheep and cattle which can get infected from ticks. Incubation period of virus after tick bites can take from one to the maximum of nine days. Onset of symptoms of CCHF are sudden like fever, myalgia, dizziness, photophobia and bleeding into skin and causes severe haemorrhagic fevers in humans, with a reported fatality of 10-40 %. CCHFV infection can be diagnosed using library tests like ELISA, antigen detection, RT-PCR, virus isolation and serum neutralization etc. [11].

Research on the pathology of CCHFV presents a major challenge, similar to Ebola Haemorrhagic Fever (EHF), because of its high virulence; it is needed to be studied under Bio-safety level 4 (BSL-4) containment. Furthermore, the lack of success in expressing the disease in laboratory animals except mice and its occurrence in medically unexplored regions makes the vaccine design even more strenuous. As a consequence, no effective vaccine is available for CCHFV, however after the onset of disease, an inadequate treatment with minimal success has been reported with the drug ribavirin [12, 13].In general, vaccines that are designed for various diseases are mostly based on B-cell immunity, however stronger immune response is shown in vaccines based on T-cell epitopes due to involvement of $CD8^+$ T-cell. Memory based humoral

immune response by B-cell weakens over time due to antigenic drift, whereas T-cell based response provides long lasting immunity [9].

In the absence of any licensed therapeutic vaccine, the present study was undertaken to design a vaccine against Crimean-Congo Haemorrhagic Fever Virus (CCHFV), based on both T-cell and B-cell epitopes.

Materials and methods

Proteome Retrieval

The proteome of CCHFV strain was retrieved from UniProtKB database [14] in FASTA format.

Allergenicity Analysis

Three protein sequences in the proteome were analyzed and scored according to their antigenicity using VaxiJen server 2.0 [15].

Prediction of T-Cell epitope

T-cell epitope prediction

NetCTL server 1.2 was used for the prediction of T-cell epitopes based on the combinatorial score. The epitopes were selected based on the combined score of TAP transport efficiency scores, major histocompatibility complex class I (MHC-I) binding and proteasomal C terminal cleavage [16].

T-cell epitope and MHC-I allele binding and population coverage analysis

MHC-I binding tool in IEDB [17] was used to analyze the T-cell and MHC-I interactions based on half maximal inhibitory concentrations (IC_{50}) of interactions calculated using stabilized matrix based method (SMM). To ascertain the population coverage of epitopes in India, the population coverage tool in IEDB was used.

Allergenicity assessment

To check whether an epitope is a probable allergen or non-allergen, AllerTOP and AlgPred servers were used. AllerTOP is an alignment free server, used for *in silico* based allergenicity prediction of a protein based on its physiochemical properties [18]. AlgPred was used to confirm the allergenicity of the epitopes based on similarity of known epitope with any region of protein [19].

Prediction of epitope structure

The structure of the epitope selected from binding analysis was predicted using PEP-FOLD-3 server [20]. Best model from the predicted top ten models from the server was selected for docking analysis.

Docking studies

The interaction between MHC-I alleles and T-cell epitopes was analyzed and confirmed using ZDOCK [21] and HDOCK [22] servers. The MHC-I class allele structure was retrieved from protein data bank (PDB). The docked complexes of epitope and allele were superimposed using SPDBV (Swiss-PDB Viewer) to compare and validate the interactions.

Prediction of B-cell epitope

B-cell epitope was predicted using antigenic sequence property tools of IEDB server [17]. To identify different properties of the fragments such as linear epitope, beta-turn, surface accessibility, flexibility, antigenicity and hydrophilicity of most antigenic protein sequence was done using tools: BepiPred linear epitope prediction, Chou and Fasman beta turn prediction, Emini surface accessibility prediction, Karplus and Schulz flexibility prediction, Kolskar and Tongaonkar antigenicity and Parker hydrophilicity prediction.

Results and Discussion

Proteome of CCHFV strain of Nigeria/IbAr10200/1970 (Proteome ID: UP000008767) was retrieved from UniProtKB. The reviewed sequences with Accession nos. Q8JSZ3, P89522 and Q6TQR were analyzed using VaxiJen server 2.0 to predict the sequence with highest antigenicity. Out of the three sequences, two sequences (Accession nos. Q6TQR6, Q8JSZ3) were identified as probable antigen and one of them (Accession no. P89522) was probable non-antigen. The sequence of envelopment polyprotein (Accession no. Q8JSZ3) with highest antigenicity score 0.5142 was selected for T-cell epitopes prediction using NetCTL server. Out of the 1676 epitopes (9-mer peptide sequences) obtained, top five epitopes (Table 1) were selected on the basis of combinatorial score for MHC-I binding analysis.

Table 1: Predicted epitopes From NetCTL server

Epitope	Combinatorial Score
FLSIDSGYY	2.3693
TAEIHGDNY	1.9652

SVMDLSQMY	1.5696
STDKEIHKL	1.5133
FTDYMFVKW	1.4137

MHC-I binding tool based on Stabilized Matrix Method (SMM) in IEDB analysis resource server was used for analysis of T-cell epitope interactions with MHC-I class alleles using reference alleles provided by IEDB server. Four epitopes (Table 2) that show interactions with alleles having affinity < 500nM (intermediate affinity) were selected for further analysis and STDKEIHKL was rejected as its affinity was greater than the cut-off. These four epitopes were then analyzed using Population coverage tool (IEDB server) and population coverage of 52.01% was obtained.

Table 2: Interactions of epitopes with corresponding MHC-I alleles using IEDB resource tools server

Epitope	Interacting MHC-I allele with an affinity <500 nM (with prediction score)
SVMDLSQMY	HLA-A*26:01(0.2) HLA-A*11:01(0.75) HLA-B*35:01(1.2) HLA-B*15:01(0.6) HLA-A*30:02(0.5) HLA-A*68:01(5.75)
FLSIDSGYY	HLA-A*30:02(0.4) HLA-A*01:01(0.5) HLA-B*15:01(1.1) HLA-B*35:01(2.2)
FTDYMFVKW	HLA-B*58:01(0.8) HLA-B*53:01(1.2)
TAEIHGDNY	HLA-B*35:01(3.4)

Amongst the four epitopes, the epitope SVMDLSQMY was found to have the highest number of interactions with MHC-I alleles *i.e.* six and highest population coverage in India *i.e.* 37.04%. Therefore, epitope SVMDLSQMY was selected as a probable T-cell epitope of CCHFV for further investigation. To predict the allergenicity of the protein sequence having Accession no. Q8JSZ3, containing epitope SVMDLSQMY, AllerTop and AlgPred servers were used as a vaccine may cause an allergic reaction in an individual which can be harmful or life threatening. From both the servers similar results were obtained that the protein was a probable non-allergen.

Based on the results of population coverage and the highest number of interaction with HLA alleles, the epitope SVMDLSQMY was chosen as an ideal candidate for T-cell based vaccine design. Protein-protein docking was performed to study the binding mode of epitope and HLA molecule. Structures of MHC-I class alleles HLA-A*11:01 (PDB: 5WJL) and HLA-B*15:01 (PDB: 1XR8) which showed significant interactions (Table 2) with the epitope SVMDLSQMY were retrieved from Protein Data Bank (PDB). Various groups bound to HLA molecules such as beta-2-microglobulin and GTS3 peptide and EBNA-3 nuclear protein were removed using Chimera software.

Due to non-availability of the structure of the epitope SVMDLSQMY in any structural database, it was predicted using PEP-FOLD server. The structure of the epitope thus obtained was docked with HLA-A*11:01 allele (Figure 1-a) and HLA-B*15:01 allele (Figure 1-b) using HDOCK and ZDOCK servers. These docked structures were superimposed using SPDBV and the RMSD value was 0.00Å, validating that the results of both HDOCK and ZDOCK servers were identical.

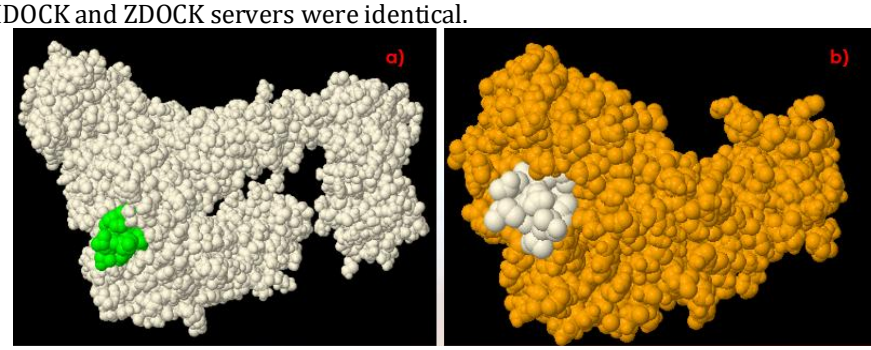

Figure 1: *Structure of a) HLA-A *11:01 b) HLA-B*15:01 with epitope SVMDLSQMY from HDOCK server*

For B-cell epitope prediction, the properties of different regions in sequence of Envelopment protein (Accession no. Q8JSZ3) of CCHFV proteome were analyzed using tools available in IEDB server[13]. T-Cell epitope vaccine candidate SVMDLSQMY was present in region 1135-1144 amino acid residues of envelopment protein sequence, therefore the graphs obtained from the results of tools of B-cell epitope prediction of IEDB server were analyzed in the region 1000-1500 residues (Figure 2).

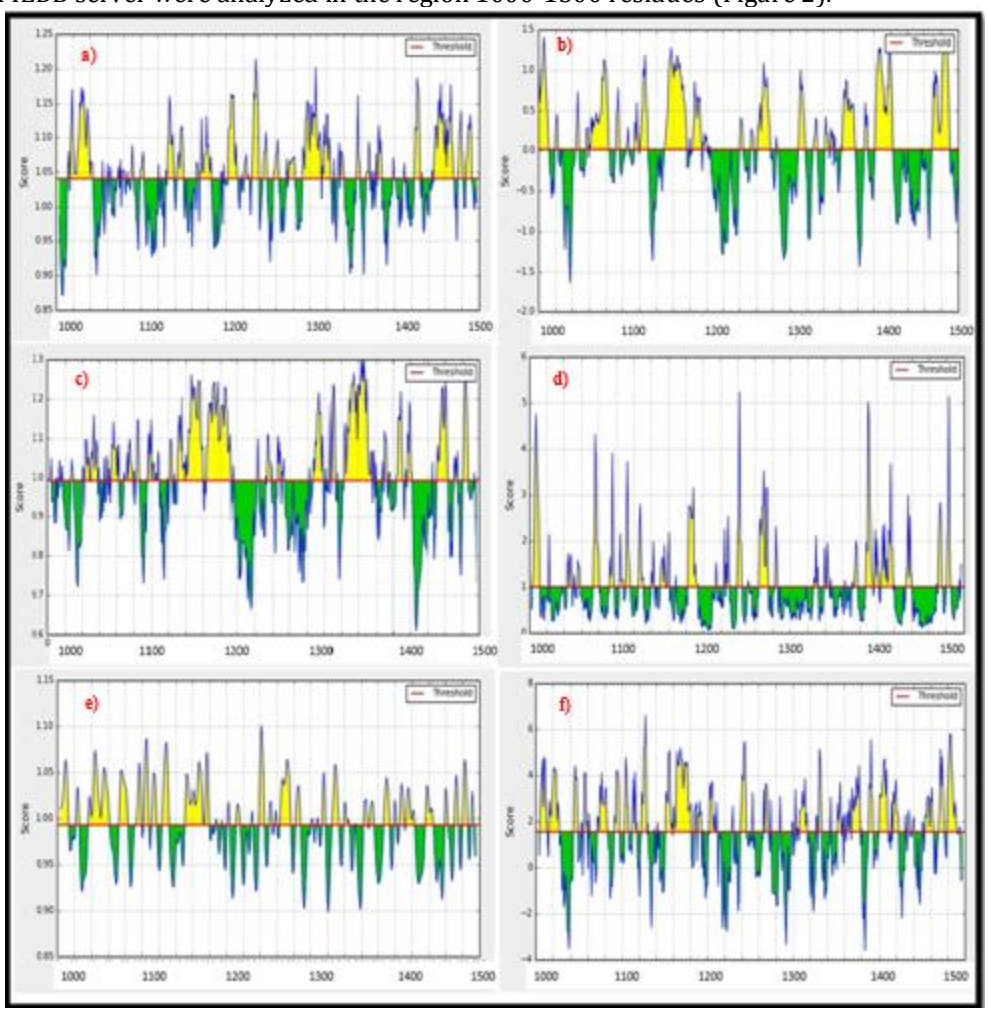

Figure 2: *Graphs of region 1001-1500 analyzed using B-cell prediction tools of IEDB. a) Kolaskar & Tongaonkar antigenicity b) Bepipred Linear epitope prediction c) Chou & Fasman Beta- Turn prediction d) Emini surface accessibility prediction e) Karplus & Schulz Flexibility prediction f) Parker Hydrophilicity prediction*

In Kolaskar & Tongaonkar antigenicity results, the maximum and minimum score obtained was 1.214 and 0.871 respectively with the region 1125-1140 having a score of 1.17 tagged with yellow in the graph indicating that the region has significant antigenicity. In BepiPred linear epitope prediction, the maximum score was 1.407 whereas the predicted peptide in the region 1120-1131 had a score of 1.2. In Chou & Fasman beta turn prediction, the maximum score and average scores were 1.299 and 0.994, respectively with the region from 1120 to 1130 having a score of 1.147. In Emini surface accessibility, the maximum score was 5.238 and the region 1126-1131 had a score of 2.794 which was shown as one of the predicted peptides. In Karplus & Schulz flexibility prediction, the maximum score was 1.101 with the region 1120-1132 having a score of 1.083, while in Parker hydrophilicity predictions, the region 1122-1131 had the highest score of 6.629.

Cross processing of the data obtained from IEDB server's B-cell epitope prediction tools was done and the region 1121-1131 residues i.e. LGVEDASESKL was found to be flexible, surface accessible, hydrophilic and contained constant beta turns. This indicates that the epitope LGVEDASESKL may be an effective B-cell epitope candidate that could initiate a significant B-cell immune response.

The results of T-cell and B-cell epitope prediction identified epitope vaccine candidates in the region 1135-1144 (SVMDLSQMY) and 1121-1131 (LGVEDASESKL), respectively. Both these candidate

epitopes are four residues apart and can be further explored using wet-lab experimentation to design an effective vaccine against CCHFV that is capable of inducing T-cell and B-cell immune responses, simultaneously.

Conclusion

Using conventional approaches, successful vaccine development against a disease can take decades due to cumbersome process of locating a region which can initiate a significant immune response. The immuno-informatics based approach can help achieve cost-efficient and timely development of vaccines due to narrowing down the number of probable effective epitopes to be validated by the wet lab researcher.

References

1. WHO, UNICEF, World Bank. (2009). State of the world's vaccines and immunization, Geneva (3rdedition).
2. Jackwood, M. W., Hickle, L., Kapil, S., Silva, R., Osterrieder, K., &Prideaux, C. (2008). Vaccine development using recombinant DNA technology. Council AgricSciTechnol, 38, 1-11.
3. López, J. A., Weilenman, C., Audran, R., Roggero, M. A., Bonelo, A., Tiercy, J. M&Corradin, G. (2001). A synthetic malaria vaccine elicits a potent CD8+ and CD4+ T lymphocyte immune response in humans. Implications for vaccination strategies. European journal of immunology, 31(7), 1989-1998.
4. Bourdette, D. N., Edmonds, E., Smith, C., Bowen, J. D., Guttmann, C. R., Nagy, Z. P., ... & Mass, M. (2005). A highly immunogenic trivalent T cell receptor peptide vaccine for multiple sclerosis. Multiple Sclerosis Journal, 11(5), 552-561.
5. Knutson, K. L., Schiffman, K., &Disis, M. L. (2001). Immunization with a HER-2/neu helper peptide vaccine generates HER-2/neu CD8 T-cell immunity in cancer patients. The Journal of Clinical Investigation, 107(4), 477-484.
6. Khan, M. A., Hossain, M. U., Rakib-Uz-Zaman, S. M., & Morshed, M. N. (2015). Epitope-based peptide vaccine design and target site depiction against Ebola viruses: an immunoinformatics study. Scandinavian journal of immunology, 82(1), 25-34.
7. Sahrawat, T. R. (2016). In-silico design of an Epitope-based peptide vaccine: A Computational Biology Approach. International Journal for Computational Biology (IJCB), 5(2), 1-5.
8. Poland, G. A., Ovsyannikova, I. G., Kennedy, R. B., Haralambieva, I. H., & Jacobson, R. M. (2011). Vaccinomics and a new paradigm for the development of preventive vaccines against viral infections. Omics: a journal of integrative biology, 15(9), 625-636.
9. Khan, A. S., Ksiazek, T. G., & Peters, C. J. (1997, January). Viral hemorrhagic fevers. In Seminars in Pediatric Infectious Diseases (Vol. 8, No. 1, pp. 64-73). WB Saunders.
10. Akbas, E. (2007). Crimean-Congo Hemorrhagic Fever-A Biological Weapon? NATO security through science Series E human and societal dynamics, 20, 89.
11. World Health Organization. (2013, January). Crimean-Congo haemorrhagic fever, Retrieved 2018, March, https://www.who.int/en/news-room/fact-sheets/detail/crimean-congo-haemorrhagic-fever.
12. Bray, M. (2007). Comparative pathogenesis of Crimean-Congo hemorrhagic fever and Ebola hemorrhagic fever. In Crimean-Congo Hemorrhagic Fever (pp. 221-231). Springer, Dordrecht.
13. Dowall, S. D., Carroll, M. W., & Hewson, R. (2017). Development of vaccines against Crimean-Congo haemorrhagic fever virus. Vaccine, 35(44), 6015-6023.
14. UniProt Consortium. (2016). UniProt: the universal protein knowledgebase. Nucleic acids research, 45(D1), D158-D169.
15. Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC bioinformatics, 8(1), 4.
16. Larsen, M. V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., & Nielsen, M. (2007). Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC bioinformatics, 8(1), 424.
17. Sette, A. (2004). The Immune Epitope Database and Analysis Resource. Genome Informatics, 15(2), 299-299.
18. Dimitrov, I., Flower, D. R., &Doytchinova, I. (2013, April). AllerTOP-a server for in silico prediction of allergens. In BMC bioinformatics (Vol. 14, No. 6, p. S4). BioMed Central.
19. Saha, S., &Raghava, G. P. S. (2006). AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic acids research, 34(suppl_2), W202-W209.
20. 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 research, 44(W1), W449-W454.
21. Pierce, B. G., Wiehe, K., Hwang, H., Kim, B. H., Vreven, T., &Weng, Z. (2014). ZDOCK server: interactive docking prediction of protein–protein complexes and symmetric multimers. Bioinformatics, 30(12), 1771-1773.
22. Yan, Y., Zhang, D., Zhou, P., Li, B., & Huang, S. Y. (2017). HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. Nucleic acids research, 45(W1), W365-W373.