Published in final edited form as:

Zoonoses (Burlingt). 2022 ; 2(1): . doi:10.15212/zoonoses-2022-0029.

In silico Design of a Crimean-Congo Hemorrhagic Fever Virus Glycoprotein Multi-Epitope Antigen for Vaccine Development

Megan C. Mears[1,#], **Dennis A. Bente**[1,2,3]

[1]Department of Pathology, University of Texas Medical Branch, Galveston, TX, USA

[2]Department of Microbiology & Immunology, University of Texas Medical Branch, Galveston, TX, USA

[3]Galveston National Laboratory, Institute for Human Infection and Immunity, University of Texas Medical Branch, Galveston, TX, USA

Abstract

Objective: There is no licensed vaccine available to prevent the severe tick-borne disease Crimean-Congo hemorrhagic fever (CCHF), caused by the CCHF virus (CCHFV). This study sought to show that a combination of computational methods and data from published literature can inform the design of a multi-epitope antigen for CCHFV that has the potential to be immunogenic.

Methods: Cytotoxic and helper T-cell epitopes were evaluated on the CCHFV GPC using bioinformatic servers, and this data was combined with work from previous studies to identify potentially immunodominant regions of the GPC. Regions of the GPC were selected for generation of a model multi-epitope antigen *in silico*, and the percent residue identity and similarity of each region was compared across sequences representing the widespread geographical and ecological distribution of CCHFV.

Results: Eleven multi-epitope regions were joined together with flexible linkers *in silico* to generate a model multi-epitope antigen, termed *EPIC*, which included 812 (75.7%) of all predicted epitopes. *EPIC* was predicted to be antigenic by two independent bioinformatic servers, suggesting that multi-epitope antigens should be explored further for CCHFV vaccine development.

Conclusion: The results presented within this manuscript provide information for potential targets within the CCHFV GPC for guiding future vaccine development.

Keywords

Bioinformatics; epitope prediction; Crimean-Congo hemorrhagic fever virus; CCHFV glycoprotein precursor; multi-epitope antigen

[#]**Correspondent**: memears@utmb.edu, 301 University Blvd., Route 0610, Galveston, Texas 77550.

Potential conflicts of interest. All authors: no reported conflicts.

INTRODUCTION

Crimean-Congo hemorrhagic fever virus (CCHFV) has the most extensive geographic range of medically significant tick-borne viruses, being endemic to 30 countries in Western Asia, Southeast Europe, the Middle East, and Africa[1]. An estimated three billion people are at risk of CCHFV infection[2]. Human cases of Crimean-Congo hemorrhagic fever (CCHF) can result in severe disease, including hemorrhaging, multi-organ failure, shock, and death, with associated case fatality rates of up to 30%[1,3,4]. Consequently, CCHFV poses a high risk to public health and has been classified as a priority pathogen for research and development by the World Health Organization (WHO), and as a Biodefense Category A pathogen by the United States National Institutes of Health (NIH)[5]. Despite decades of vaccine development research for CCHFV, no approved vaccine is widely available for human use. The only vaccine available for human use is not licensed by the FDA and is only used in Bulgaria, as there are safety concerns associated with this inactivated mouse brain preparation[1].

CCHFV contains a tri-segmented, negative sense, and single-stranded RNA genome, and is classified within the family *Nairoviridae*, genus *Orthonairovirus*[6]. The three segments, known as the small (S), medium (M), and large (L) segments, encode the nucleoprotein (NP), glycoprotein precursor (GPC), and RNA-dependent RNA polymerase (RdRp), respectively. Vaccine development for CCHFV has primarily focused on the use of the GPC or the NP in various platform technologies, with the majority of vaccine candidates demonstrating protection from challenge when encoding the glycoproteins. CCHFV is the most genetically diverse arbovirus, and the M segment demonstrates the largest nucleotide diversity of the three virus segments at 31%[1]. Despite the significant divergence of this segment, the M segment has been the most explored for antibody and T-cell epitope mapping, epitope predictions, and vaccine development for CCHFV since it encodes the structural glycoproteins present on the surface of the virion that induce both cellular and humoral immune responses. The CCHFV GPC undergoes the most extensive cleavage and processing of viruses in the order *Bunyavirales* to form the two structural glycoproteins G_N and G_C, and the three nonstructural glycoproteins, the mucin-like domain (MLD), GP38, and NS_M[7]. The complex processing of the GPC includes N-linked glycosylation, many cysteine residues that may form disulfide bonds, and numerous O-linked glycosylations[8]. The complicated processing of the GPC, which requires the nonstructural proteins for the proper maturation of the structural glycoproteins, has led to the use of the whole GPC as a common vaccine antigen.

Correlates of protection for CCHFV have yet to be defined[9]. However, it is known that T cells play an important role in CCHFV immunity and are necessary for survival of CCHFV infection[10,11]. Depletion of CD4+ and/or CD8+ T-cells exacerbates morbidity and mortality during acute CCHFV infection, vaccination using the whole GPC can robustly activate T-cells, and human survivors of CCHF have long-lived CD8+ T-cell responses[10–16]. T-cell immunogenicity is not evenly distributed across the GPC, as certain regions generate stronger recall responses than others. These recall responses have also been shown to vary depending on the CCHFV strain of the stimulating peptide pool, where peptides from a strain heterologous to the immunizing strain show significantly reduced recall responses[15]. This pattern is also seen with antibody binding assays using GPC peptides[17]. This suggests

that specific humoral or cellular responses could be induced from vaccination with specific GPC regions.

The size, complex processing, and uneven distribution of immunogenicity across the CCHFV GPC make it a strong candidate for vaccine development utilizing a multi-epitope antigen rather than the whole GPC[18–21]. Previous research has demonstrated the feasibility of a multi-epitope vaccine development strategy for other members of *Bunyavirales* with complex glycoprotein processing; a multi-epitope DNA vaccine generated with conserved epitopes selected from an alignment of the Hantaan virus (HTNV), Seoul virus (SEOV), and Puumala virus (PUUV) glycoproteins induced both humoral and cellular immunity against all three viruses in mice[22]. Ideally, this strategy for the development of a CCHFV multi-epitope antigen would allow for protection from diverse CCHF viruses, which has been a shortcoming of previous CCHFV vaccine candidates.

Previous studies attempting to develop multi-epitope vaccines for CCHFV have been limited to a “string of beads” approach, that links a few individual epitopes from the NP, structural glycoproteins, or RdRp, that are recognized only by specific major histocompatibility complex (MHC) alleles. In contrast, we sought to show that immunoinformatic analyses to predict cytotoxic T-lymphocyte (CTL) and helper T-lymphocyte (HTL) epitopes in the CCHFV GPC can be combined with regions that generate T-cell recall responses in the literature to generate a multi-epitope antigen that is composed of large GPC regions that include numerous predicted epitopes recognized by a variety of MHC alleles. It is hypothesized that generation of this type of multi-epitope antigen may generate robust T- and B-cell responses in future studies that are required for protection from lethal CCHFV challenge[10,12,14–16,23]. Regions of the GPC with multiple predicted epitopes were selected for generation of a model multi-epitope antigen, and the residue homology of each region was compared to the homology of whole GPC proteins from M segment sequences representing the widespread geographical and ecological distribution of the virus. Finally, a multi-epitope antigen was constructed *in silico*, and subcellular localization and antigenicity of the protein was predicted using bioinformatic servers. This work provides new information for potential targets during CCHFV vaccine development.

METHODS

Evaluation of CCHFV GPC Residue Diversity Across Clades

The amino acid sequence of the CCHFV GPC from strain Turkey200406546, designated throughout this work as Turkey2004 (Accession # KY362519) was used as a reference sequence for alignments, calculations of percent identity and percent similarity, and for prediction of CTL and HTL epitopes. This strain of CCHFV has not previously been used for *in silico* analysis of epitope prediction, despite the high annual incidence of CCHFV cases in Turkey[1]. Fifty GPC sequences were selected as representative sequences for the widespread geographical and ecological distribution of the virus, with sequences from each CCHFV clade, and from ticks, animals, and clinical cases. A phylogenetic tree was constructed using Geneious Tree Builder with default parameters, and clades were assigned based on previous publications[1,24]. Each full-length GPC sequence was aligned to the Turkey2004 sequence, and individual alignments were made for each GPC protein

(MLD, GP38, G_N, NS_M, and G_C) using Geneious Prime software (version 2021.1.1). The percent identity and percent similarity of each selected sequence to the Turkey2004 sequence was determined using William Pearson's *lalign* program run through the Swiss Institute of Bioinformatics ExPASy Bioinformatics Resource Portal (now available through the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) https://www.ebi.ac.uk/Tools/psa/lalign/), where sequence identity considers only the residues that match strictly between two sequences, and sequence similarity considers residues that match exactly, and of the residues that differ, how similar the physicochemical properties of the residues are.

CTL Epitope Prediction

The translation of the GPC of CCHFV strain Turkey2004 was used for bioinformatic server predictions to identify epitopes likely to be presented by human MHC class I molecules to CD8+ cytotoxic T-lymphocytes. CTL epitopes were predicted using the NetCTL 1.2 Server (https://services.healthtech.dtu.dk/service.php?NetCTL-1.2) powered by the Department of Bio and Health Informatics at the Technical University of Denmark. The NetCTL 1.2 Server was used to predict binding of 9-mer CTL peptides to 12 MHC class I supertypes (A1, A2, A3, A24, A26, B7, B8, B27, B39, B44, B58, and B62) using neural networks. Predicted peptides were selected based on an inclusion criterion of a combined score of >1.0 given from prediction of MHC class I binding, proteasomal C-terminal cleavage, and TAP transport efficiency. A combined score of >1.0 yields 70% sensitivity and 98.5% specificity of CTL ligand prediction accuracy. Peptides were excluded if they fell across known cleavage sites on the GPC polyprotein.

Helper T-lymphocyte (HTL) Epitope Prediction

The translation of the GPC of CCHFV strain Turkey2004 was used for bioinformatic server predictions to identify epitopes likely to be presented by human MHC class II molecules to CD4+ helper T-lymphocytes. HTL epitopes were predicted using the NetMHCII 2.3 Server (https://services.healthtech.dtu.dk/service.php?NetMHCII-2.3) powered by the Department of Bio and Health Informatics at the Technical University of Denmark. The NetMHCII 2.3 Server was used to predict binding of 15-mer HTL peptides to 25 HLA-DR, 20 HLA-DQ, and 9 HLA-DP alleles using artificial neural networks. Predicted peptides were selected if they met both inclusion criteria of (1) a strong binder threshold of <2.00%-rank to a set of 1,000,000 random natural peptides and (2) a predicted IC50 value <50 nM. Peptides were excluded if they fell across known cleavage sites on the GPC polyprotein.

Alignment of Peptides to the CCHFV GPC and Selection of Multi-Epitope Regions

All predicted CTL and HTL peptides that met the above inclusion criteria were aligned to the Turkey2004 GPC sequence using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). Groups of peptides were assigned numerical values for the number of peptides that overlapped a given residue in the GPC alignment, and peptides were graphed in GraphPad Prism (Version 8) using these assigned values. The only two specific T-cell epitopes that have been experimentally demonstrated from convalescent CCHF cases[13] were included in a different graphing panel for evaluation. Regions with the highest number of

overlapping predicted peptides and regions including the two human T-cell epitopes were selected as multi-epitope regions.

Evaluation of Conservation of Selected Multi-Epitope Regions Across CCHFV Clades

Sequence conservation was assessed at the residue level for both percent identity and percent similarity. Each of the 11 selected Turkey2004 multi-epitope epitope sequences were compared individually to the same region of 50 representative CCHFV sequences, as mentioned previously, using William Pearson's *lalign* program. The average percent identity and percent similarity of each multi-epitope region across the 50 sequences was compared to the average percent identity of the GPC protein where the multi-epitope region originated.

Construction of the Multi-Epitope Antigen and Antigenicity Prediction

The 11 multi-epitope regions were joined together *in silico* using a flexible linker of -glycine-glycine-glycine-serine- (-GGGS-). A start codon was placed at the N-terminus, and a six residue polyhistidine tag (6XHis) was added before a stop codon at the C-terminus. A signal peptide was not included at the N-terminus of the antigen to prevent localization for processing through the secretory pathway, as is the case for the GPC. The final model multi-epitope antigen, termed EPItope Construct (*EPIC*), with linkers and tag was 853 residues in length and will be referred to as *EPIC* throughout this manuscript. Subcellular localization and the residues important for localization of *EPIC* were predicted using the DeepLoc-1.0 server run through the Department of Bio and Health Informatics at the Technical University of Denmark (https://services.healthtech.dtu.dk/service.php?DeepLoc-1.0). Antigenicity of *EPIC* and individual GPC proteins was predicted with two independent servers, ANTIGENpro (http://scratch.proteomics.ics.uci.edu/) and VaxiJen (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html).

RESULTS

Evaluation of CCHFV GPC Protein Diversity Across Clades

Fifty GPC sequences from ticks, animals, and clinical cases were selected for evaluation with representative sequences from each clade of CCHFV and spanning the widespread geographical distribution of the virus (Figure 1). The percent identity and similarity of each CCHFV GPC protein, MLD, GP38, G_N, NS_M, and G_C, from each sequence to the Turkey2004 sequence was calculated using William Pearson's *lalign* program. The Turkey2004 sequence aligns with clade V-Europe 1, and unsurprisingly, the representative sequences from this clade have the greatest homology to the Turkey2004 sequence. Sequences from clades I-IV and VI displayed greater differences in percent identity and similarity to the Turkey2004 sequence. There is a gradient in the homology of the GPC, with the N-terminus displaying a low level of homology, with increasing homology at the C-terminus of the GPC. The structural glycoproteins G_N and G_C display the greatest average percent identity and similarity across the six clades. The N-terminal nonstructural proteins MLD and GP38 display the lowest average percent identity and similarity across the six clades.

CTL and HTL Epitope Prediction and Epitope Alignment to the GPC

The NetCTL 1.2 Server predicted 256 CTL peptides that met the initial inclusion criteria, and the NetMHCII 2.3 Server predicted 837 HTL peptides that met the initial inclusion criteria (Table 1). These 1093 predicted epitopes were aligned to the full-length GPC using Clustal Omega. After alignment, 4 CTL ligands and 17 HTL ligands were excluded for their location across known GPC cleavage sites, leaving 252 CTL ligands and 820 HTL ligands for consideration (Table 1). It was expected that the number of peptides that aligned to the individual proteins would be relative to the proportion of the GPC that the individual proteins comprise. The individual GPC proteins, MLD (249 residues), GP38 (272 residues), G_N (288 residues), NS_M (233 residues), and G_C (647 residues), comprise 14.7%, 16.1%, 17.1%, 13.8%, and 38.3% of the polyprotein, respectively. The number of peptides that aligned to the individual proteins resulted in similar proportions of 19.1%, 16.2%, 13.1%, 14.9%, and 34.8% for MLD, GP38, G_N, NS_M, and G_C, respectively (Table 1). Interestingly, the peptides predicted for the nonstructural proteins MLD, GP38, and NSM resulted in peptide prediction proportions higher than the proportion of the polyprotein the proteins comprise. At the same time, there were a lower proportion of peptides predicted for the structural glycoproteins G_N and G_C. The pattern of the predicted epitopes aligned to the GPC follows closely with graphs of antibody reactivity data and T-cell recall responses across the CCHFV GPC[15,17], giving confidence in this prediction method for identifying potentially immunogenic regions of the GPC.

Identification of Multi-Epitope Regions

The 1072 CTL and HTL ligands that met the inclusion criteria were quantified and graphed by assigning numerical values for the number of predicted peptides that overlapped a given residue in the GPC sequence. This graphing method allowed for comparison to experimentally demonstrated CD8+ T-cell epitopes (Figure 2A)[13], and for identification of regions with the most predicted epitopes (Figure 2B). There were 11 regions of the GPC, referred to as GPC-01 through GPC-11, chosen for further evaluation, which had the greatest number of overlapping predicted epitopes and included the two experimentally shown human T-cell epitopes (Figure 2C). The location of the 11 multi-epitope regions are as follows, using the residue numbers of the Turkey2004 GPC sequence: residues 96–172 (GPC-01), residues 205–244 (GPC-02), residues 253–283 (GPC-03), residues 303–399 (GPC-04), residues 643–742 (GPC-05), residues 923–1007 (GPC-06), residues 1043–1105 (GPC-07), residues 1270–1325 (GPC-08), residues 1335–1435 (GPC-09), residues 1472–1529 (GPC-10), and residues 1572–1663 (GPC-11). These 11 regions include 812 (75.7%) of all predicted CTL and HTL ligands and include 805 residues (47.7%) of the GPC sequence with regions selected from all five proteins of the GPC (Table 2).

Evaluation of Conservation of Selected Multi-Epitope Regions Across CCHFV Clades

Multiple studies have demonstrated variability of immune responses to homologous or heterologous CCHFV strains[15,17,25–30]. Thus, it was essential to assess the residue conservation of the selected GPC regions between various CCHFV sequences. Each selected multi-epitope region was evaluated for residue identity across 50 different CCHFV sequences (Figure 3). As expected from the evaluation of identity across the whole CCHFV

proteins, the multi-epitope regions originating from the MLD had the lowest average percent identity, while multi-epitope regions originating from the G_C had the highest average percent identity. The average percent identity of each region was compared to its originating GPC protein, and 5/11 regions (GPC-04, GPC-05, GPC-06, GPC-10, and GPC-11) were less conserved than their respective whole proteins, and 6/11 regions (GPC-01, GPC-02, GPC-03, GPC-07, GPC-08, and GPC-09) had equal to, or greater, conservation than their respective whole proteins.

Constuction of EPIC and Antigenicity Prediction

For the exercise of constructing a model multi-epitope antigen for CCHFV, *EPIC* was generated *in silico* using all 11 selected GPC regions. These regions included three transmembrane domains in GPC-05, GPC-06, and GPC-11, originating from G_N, NS_M, and G_C, respectively. It was hypothesized that the lack of signal peptide in the protein, and the inclusion of the transmembrane domain from the structural glycoprotein G_C, would localize *EPIC* to the cell membrane, rather than signaling *EPIC* to be processed through the secretory pathway in a manner similar to the full-length GPC. Localization prediction by the DeepLoc 1.0 server predicted that *EPIC* was likely to localize to the cell membrane, with a 78.7% likelihood (Figure 4A). The residues most important for the predicted subcellular localization of *EPIC* (Figure 4B) correspond to the three transmembrane domains included in the construct (Figure 4C). These results suggest that the inclusion of the transmembrane domains, but not signal peptide, from the GPC can result in a different predicted subcellular localization than the GPC, and the localization of future multi-epitope antigens can be modified by inclusion of different GPC regions.

The antigenicity of *EPIC* and the GPC proteins was predicted using two independent bioinformatic servers. ANTIGENpro predicted that *EPIC* was more likely to be antigenic than any of the GPC proteins alone with a predicted value of 0.94 out of 1 (MLD: 0.91; GP38: 0.56; G_N: 0.73; NS_M: 0.25; G_C: 0.87). The VaxiJen server uses a threshold of 0.4 for indicating protein antigenicity. The VaxiJen server predicted that *EPIC* (0.49) would be less antigenic than the MLD (0.50), NS_M (0.58) and G_C (0.57) proteins, but more antigenic than GP38 (0.42) and G_N (0.45). These results suggest that *EPIC* has antigenic potential, and further suggest that multi-epitope antigens such as *EPIC* should be evaluated *in vitro* and *in vivo* in future studies.

DISCUSSION

CCHF is a widespread and medically important tick-borne viral disease, yet, no vaccine is widely available for human use[3]. Vaccine development strategies for CCHFV have evaluated numerous platform and antigen combinations, and current CCHFV DNA vaccines being explored primarily encode the full-length GPC, or other whole protein-encoding sections from the GPC, including GP38, G_N, and G_C, due to the complex processing of the glycoproteins[14–16,31,32]. It has been suggested that the size and complex processing of the CCHFV GPC may hinder the GPC as a vaccine antigen and that the use of a multi-epitope antigen would be an improvement from current CCHFV DNA vaccine candidates[18–21]. Bioinformatic analyses towards generation of a CCHFV multi-epitope

vaccine have previously evaluated the NP, GPC, or RdRp, for prediction of T- and B-cell epitopes, but these studies have been limited to evaluation of a single strain of CCHFV. The results presented herein are the first to provide bioinformatic epitope prediction using the Turkey2004 strain of the CCHFV GPC, evaluation of multi-epitope region homology between sequences from all CCHFV clades, and provide information for potential targets within the GPC for vaccine development.

Previous studies have discussed the diversity of the CCHFV M segment. However, this is the first study to compare the Turkey2004 GPC sequence to other CCHFV isolates spanning the widespread geographical distribution of CCHFV (Figure 1). Unsurprisingly, the Turkey2004 GPC displays high levels of similarity and identity to other isolates within Clade V-Europe 1, but greater diversity from other clades. The Turkey2004 GPC sequence displays greater homology to isolates from Asia and East Africa, than strains from West Africa, Central Africa, and the Mediterranean. This may be explained by migratory bird patterns, which play a role in the spread of *Hyalomma* ticks, the primary vector and reservoir of CCHFV[1,33]. These results highlight the importance of choosing specific regions of the GPC for vaccine development, with certain regions of the GPC demonstrating greater cross-clade homology than others.

The bioinformatic analyses in this manuscript predicted CTL and HTL epitopes using the amino acid sequence of the GPC from strain Turkey2004, followed by the evaluation of residue conservation of selected GPC regions across CCHFV clades. A total of 1093 T-cell epitopes were predicted across the Turkey2004 GPC sequence (Table 1). Alignment of these predicted epitopes to the GPC revealed 21 predicted epitopes that were located across experimentally demonstrated GPC cleavage sites, and were excluded from consideration (Table 1). All included epitopes were quantified and graphed to identify multi-epitope regions (Figure 2). Previous studies have shown that immunogenicity is not evenly distributed across the GPC, and certain regions of the GPC generate stronger antibody binding and T-cell recall responses than others[12,15–17]. Results of antibody epitope mapping of pooled Turkish or South African CCHFV convalescent sera to linear peptide pools generated from the Turkey-Kelkit06 strain of the CCHFV GPC demonstrated that the greatest reactivity was seen within the nonstructural glycoproteins, specifically, the C-terminus of MLD, N-terminus of G_N, and middle of NS_M[17]. Reactivity was also seen at the C-terminus of the structural glycoprotein G_C[17]. Similarly, peptides generated from the MLD and NS_M glycoproteins generate robust T-cell recall responses in GPC DNA vaccined mice and *Cynomologous macaques*[15,16]. Conversely, the N-terminus of G_C generates stronger T-cell recall responses than the C-terminus in animals, however, the two specific human T-cell epitopes identified using PBMCs from convalescent patients are found in the middle of G_C[13]. Herein, the *in silico* analyses of CTL and HTL epitopes within the Turkey2004 GPC demonstrate an uneven distribution across the GPC, with certain regions displaying more predicted epitopes than others. These results confirm the published results of splenocyte stimulation with peptide pools spanning the GPC, where splenocytes are not uniformly stimulated across the GPC. Additionally, the graphing pattern of the predicted T-cell epitopes was similar to the results of linear antibody epitope mapping of pooled Turkish sera to peptides generated against the CCHFV strain Turkey-Kelkit06[17]. Together, results from the reactivity of T-cells[15,16], antibody binding[17], and prediction of epitopes across

the CCHFV GPC (this manuscript)[18,20,21], provides information on the regions of the GPC which are predicted to be the most immunogenic, and can act as potential targets for vaccine development: the nonstructural glycoproteins and G_C.

To generate a model multi-epitope antigen *in silico*, 11 regions of the GPC that contained the greatest number of overlapping epitopes, or included previously published T-cell epitopes[13], were selected for further evaluation (Table 2). Multi-epitope regions were selected from each of the GPC proteins and contained 812 (75.7%) of all predicted CTL and HTL ligands and included 805 residues (47.7%) of the GPC sequence (Table 2). Interestingly, more epitopes were predicted in the nonstructural proteins than the structural glycoproteins, relative to protein size. This finding is similar to data showing that splenocyte stimulation with peptides homologous to the vaccine strain generates greater recall responses to nonstructural proteins than the structural glycoproteins, but not with heterologous peptides[15]. It was hypothesized that multi-epitope regions would demonstrate greater cross-clade homology than their whole originating GPC proteins, theoretically eliminating issues with reduced T-cell recall responses from heterologous strains. However, only 6/11 regions had equal to, or greater, conservation than their respective whole proteins (Figure 3). The MLD is the most diverse protein of the GPC (Figure 1), and this was the only GPC protein where both multi-epitope regions demonstrated greater than or equal levels of residue identity and similarity across 50 CCHFV sequences than the whole protein (Figure 3). These results suggest that a multi-epitope antigen derived strictly from regions of the most predicted epitopes may not be able to overcome the significant diversity of CCHFV between strains. In combination with published data showing variability in immune reactivity between strains[15,17], these data suggest that given the widespread distribution of CCHFV, future work should focus on GPC regions with the most significant residue conservation to generate a vaccine candidate that produces heterologous protection across CCHFV clades. Alternatively, the development of regionally strain-specific vaccines is worth consideration.

The model multi-epitope antigen, *EPIC*, was constructed *in silico* by connecting each selected multi-epitope region with flexible linkers. The final multi-epitope antigen was predicted to localize to the cell membrane in mammalian cells, with the included transmembrane domains being important for localization prediction (Figure 4), and the protein was predicted to be antigenic by two different prediction servers. These results suggest that a multi-epitope antigen for CCHFV can be designed *in silico*, with favorable antigenic properties. This information can be used in downstream studies to modify the subcellular localization, and therefore processing and immunogenicity of multi-epitope antigens by inclusion of various domains of the GPC. Since a multi-epitope vaccine for HTNV, SEOV, and PUUV, was able to induce both cellular and humoral immunity to each virus in mice[22], future *in vitro* and *in vivo* experiments utilizing *EPIC* or other CCHFV multi-epitope antigens should be undertaken to evaluate the immunogenicity and efficacy of this type of vaccine development strategy

The bioinformatic servers used within this manuscript predict epitopes based on linear amino acid sequences for protein cleavage and binding of peptides to MHC complexes. The *in silico* analysis is designed to provide information about epitopes that may bind to MHC complexes; however, the complex nature of this biological process is difficult

to predict, so, likely, at least some or many of the predicted epitopes included within *EPIC* are not immunogenic. Sequence variability can affect the prediction of epitopes, leading to vastly different prediction results. These results from epitope prediction using the Turkey2004 strain of the CCHFV GPC can be combined with epitope prediction using other CCHFV strains to provide a broader understanding of how CCHFV sequence diversity may impact immunogenicity. Given the wide range of sequence variability within some of the GPC proteins, epitope predictions may be improved by completing the analyses with multiple CCHFV GPC sequences from diverse strains, then choosing regions based on homologously predicted peptides, rather than assessing conservation after regions were chosen from epitope predictions from a single sequence of GPC.

In conclusion, the results presented in this paper demonstrate that a model multi-epitope antigen using the Turkey2004 strain of the CCHFV GPC can be designed *in silico* by combining bioinformatic epitope predictions and T-cell immunogenicity data from the literature. The predicted antigenicity of *EPIC* suggests that a multi-epitope antigen based vaccine for CCHFV should be evaluated as a future vaccine development strategy for CCHFV.

ACKNOWLEDGEMENTS

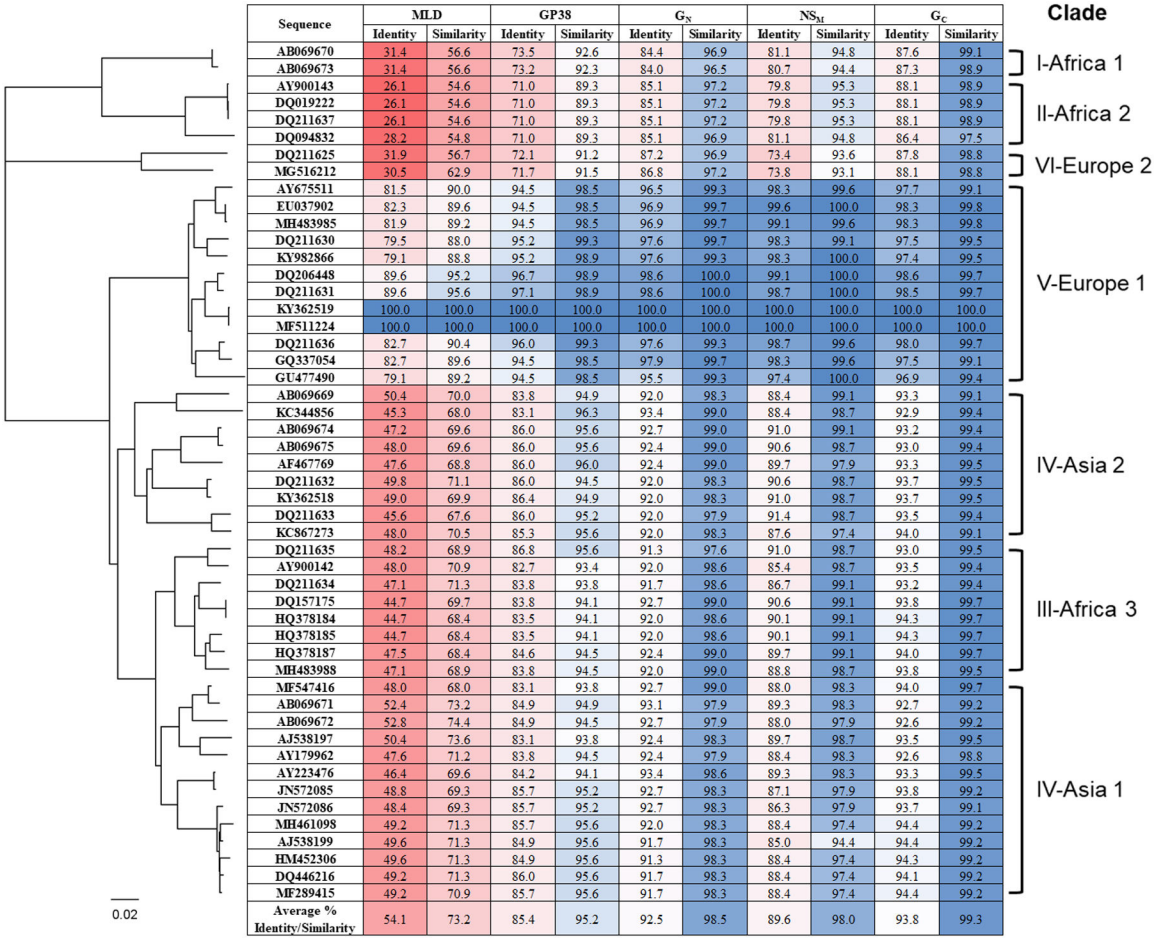
MCM was supported by pre-doctoral fellowships from the UTMB Sealy Institute for Vaccine Sciences and NIAID Emerging and Tropical Infectious Diseases T32 (AI007526-20, PI: Dr. Lynn Soong).

REFERENCES

1. Bente DA et al. Crimean-Congo hemorrhagic fever: History, epidemiology, pathogenesis, clinical syndrome and genetic diversity. Antiviral Res. 100, 159–189 (2013). [PubMed: 23906741]
2. Control, E. C. for D. P. and. Factsheet about Crimean-Congo haemorrhagic fever. European Centre for Disease Prevention and Control Available at: https://www.ecdc.europa.eu/en/crimean-congo-haemorrhagic-fever/facts/factsheet.
3. Ergönül Ö Crimean-Congo haemorrhagic fever. Lancet Infect. Dis 6, 203–214 (2006). [PubMed: 16554245]
4. Whitehouse CA Crimean–Congo hemorrhagic fever. Antiviral Res. 64, 145–160 (2004). [PubMed: 15550268]
5. Keshtkar-Jahromi M et al. Crimean-Congo hemorrhagic fever: Current and future prospects of vaccines and therapies. Antiviral Res. 90, 85–92 (2011). [PubMed: 21362441]
6. Adams MJ et al. Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2017). Arch. Virol 162, (2017).
7. Erickson BR, Deyde V, Sanchez AJ, Vincent MJ & Nichol ST N-linked glycosylation of Gn (but not Gc) is important for Crimean Congo hemorrhagic fever virus glycoprotein localization and transport. Virology 361, 348–355 (2007). [PubMed: 17197010]
8. Sanchez AJ, Vincent MJ & Nichol ST Characterization of the Glycoproteins of Crimean-Congo Hemorrhagic Fever Virus. J. Virol 76, 7263 (2002). [PubMed: 12072526]
9. Rodriguez SE et al. Immunobiology of Crimean-Congo hemorrhagic fever. Antiviral Res. 199, 105244 (2022). [PubMed: 35026307]
10. Hawman DW et al. T-Cells and Interferon Gamma Are Necessary for Survival Following Crimean-Congo Hemorrhagic Fever Virus Infection in Mice. Microorganisms 9, 1–17 (2021).
11. Dowall SD et al. Protective effects of a Modified Vaccinia Ankara-based vaccine candidate against Crimean-Congo Haemorrhagic Fever virus require both cellular and humoral responses. PLoS One 11, e0156637 (2016). [PubMed: 27272940]

12. Buttigieg KR et al. A Novel Vaccine against Crimean-Congo Haemorrhagic Fever Protects 100% of Animals against Lethal Challenge in a Mouse Model. PLoS One 9, e91516 (2014). [PubMed: 24621656]
13. Goedhals D, Paweska JT & Burt FJ Long-lived CD8+ T cell responses following Crimean-Congo haemorrhagic fever virus infection. PLoS Negl. Trop. Dis 11, e0006149 (2017). [PubMed: 29261651]
14. Garrison AR et al. A DNA vaccine for Crimean-Congo hemorrhagic fever protects against disease and death in two lethal mouse models. PLoS Negl. Trop. Dis 11, (2017).
15. Suschak JJ et al. A CCHFV DNA vaccine protects against heterologous challenge and establishes GP38 as immunorelevant in mice. npj Vaccines 2021 61 6, 1–11 (2021). [PubMed: 33398010]
16. Hawman DW et al. A DNA-based vaccine protects against Crimean-Congo haemorrhagic fever virus disease in a Cynomolgus macaque model. Nat. Microbiol (2020). doi:10.1038/s41564-020-00815-6
17. Fritzen A et al. Epitope-mapping of the glycoprotein from Crimean-Congo hemorrhagic fever virus using a microarray approach. PLoS Negl. Trop. Dis 12, (2018).
18. Nosrati M, Behbahani M & Mohabatkar H Towards the first multi-epitope recombinant vaccine against Crimean-Congo hemorrhagic fever virus: A computer-aided vaccine design approach. J. Biomed. Inform 93, 103160 (2019). [PubMed: 30928513]
19. Tipih T & Burt FJ Crimean-congo hemorrhagic fever virus: Advances in vaccine development. BioResearch Open Access 9, 137–150 (2020). [PubMed: 32461819]
20. Khan MSA et al. Computational formulation and immune dynamics of a multi-peptide vaccine candidate against Crimean-Congo hemorrhagic fever virus. Mol. Cell. Probes 55, 101693 (2021). [PubMed: 33388416]
21. Tahir Ul Qamar M et al. Development of a Novel Multi-Epitope Vaccine Against Crimean-Congo Hemorrhagic Fever Virus: An Integrated Reverse Vaccinology, Vaccine Informatics and Biophysics Approach. Front. Immunol 12, 2313 (2021).
22. Zhao C et al. Immunogenicity of a multi-epitope DNA vaccine against hantavirus. Hum. Vaccin. Immunother 8, 208–215 (2012). [PubMed: 22426376]
23. Dowall SD et al. A Crimean-Congo hemorrhagic fever (CCHF) viral vaccine expressing nucleoprotein is immunogenic but fails to confer protection against lethal disease. Hum. Vaccines Immunother 12, 519–527 (2016).
24. Ramírez de Arellano E et al. Phylogenetic Characterization of Crimean-Congo Hemorrhagic Fever Virus, Spain. Emerg. Infect. Dis 23, 2078 (2017). [PubMed: 29148370]
25. Spengler JR et al. Heterologous protection against Crimean-Congo hemorrhagic fever in mice after a single dose of replicon particle vaccine. Antiviral Res. 170, 104573 (2019). [PubMed: 31377243]
26. Hawman DW et al. Crimean-Congo Hemorrhagic Fever Mouse Model Recapitulating Human Convalescence. J. Virol 93, (2019).
27. Golden JW et al. GP38-targeting monoclonal antibodies protect adult mice against lethal Crimean-Congo hemorrhagic fever virus infection. Sci. Adv (2019). doi:10.1126/sciadv.aaw9535
28. Rodriguez SE et al. Vesicular Stomatitis Virus-Based Vaccine Protects Mice against Crimean-Congo Hemorrhagic Fever. Sci. Rep (2019). doi:10.1038/s41598-019-44210-6
29. Spengler JR et al. Viral replicon particles protect IFNAR−/− mice against lethal Crimean-Congo hemorrhagic fever virus challenge three days after vaccination. Antiviral Res. 191, 105090 (2021). [PubMed: 34044061]
30. Zivcec M, Spiropoulou CF & Spengler JR The use of mice lacking type I or both type I and type II interferon responses in research on hemorrhagic fever viruses. Part 2: Vaccine efficacy studies. Antiviral Res. 174, 104702 (2020). [PubMed: 31982149]
31. Hinkula J et al. Immunization with DNA Plasmids Coding for Crimean-Congo Hemorrhagic Fever Virus Capsid and Envelope Proteins and/or Virus-Like Particles Induces Protection and Survival in Challenged Mice. J. Virol 91, (2017).
32. Spik K et al. Immunogenicity of combination DNA vaccines for Rift Valley fever virus, tick-borne encephalitis virus, Hantaan virus, and Crimean Congo hemorrhagic fever virus. Vaccine 24, 4657–4666 (2006). [PubMed: 16174542]

33. Jameson LJ, Morgan PJ, Medlock JM, Watola G & Vaux AGC Importation of Hyalomma marginatum, vector of Crimean-Congo haemorrhagic fever virus, into the United Kingdom by migratory birds. Ticks Tick. Borne. Dis 3, 95–99 (2012). [PubMed: 22300969]



Sequence	MLD		GP38		G_N		NS_M		G_C		Clade
	Identity	Similarity	Identity	Similarity	Identity	Similarity	Identity	Similarity	Identity	Similarity	
AB069670	31.4	56.6	73.5	92.6	84.4	96.9	81.1	94.8	87.6	99.1	I-Africa 1
AB069673	31.4	56.6	73.2	92.3	84.0	96.5	80.7	94.4	87.3	98.9	
AY900143	26.1	54.6	71.0	89.3	85.1	97.2	79.8	95.3	88.1	98.9	II-Africa 2
DQ019222	26.1	54.6	71.0	89.3	85.1	97.2	79.8	95.3	88.1	98.9	
DQ211637	26.1	54.6	71.0	89.3	85.1	97.2	79.8	95.3	88.1	98.9	
DQ094832	28.2	54.8	71.0	89.3	85.1	96.9	81.1	94.8	86.4	97.5	
DQ211625	31.9	56.7	72.1	91.2	87.2	96.9	73.4	93.6	87.8	98.8	VI-Europe 2
MG516212	30.5	62.9	71.7	91.5	86.8	97.2	73.8	93.1	88.1	98.8	
AY675511	81.5	90.0	94.5	98.5	96.5	99.3	98.3	99.6	97.7	99.1	V-Europe 1
EU037902	82.3	89.6	94.5	98.5	96.9	99.7	99.6	100.0	98.3	99.8	
MH483985	81.9	89.2	94.5	98.5	96.9	99.7	99.1	99.6	98.3	99.8	
DQ211630	79.5	88.0	95.2	99.3	97.6	99.7	98.3	99.1	97.5	99.5	
KY982866	79.1	88.8	95.2	98.9	97.6	99.3	98.3	100.0	97.4	99.5	
DQ206448	89.6	95.2	96.7	98.9	98.6	100.0	99.1	100.0	98.6	99.7	
DQ211631	89.6	95.6	97.1	98.9	98.6	100.0	98.7	100.0	98.5	99.7	
KY362519	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	
MF511224	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	
DQ211636	82.7	90.4	96.0	99.3	97.6	99.3	98.7	99.6	98.0	99.7	
GQ337054	82.7	89.6	94.5	98.5	97.9	99.7	98.3	99.6	97.5	99.1	
GU477490	79.1	89.2	94.5	98.5	95.5	99.3	97.4	100.0	96.9	99.4	
AB069669	50.4	70.0	83.8	94.9	92.0	98.3	88.4	99.1	93.3	99.1	IV-Asia 2
KC344856	45.3	68.0	83.1	96.3	93.4	99.0	88.4	98.7	92.9	99.4	
AB069674	47.2	69.6	86.0	95.6	92.7	99.0	91.0	99.1	93.2	99.4	
AB069675	48.0	69.6	86.0	95.6	92.4	99.0	90.6	98.7	93.0	99.4	
AF467769	47.6	68.8	86.0	96.0	92.4	99.0	89.7	97.9	93.3	99.5	
DQ211632	49.8	71.1	86.0	94.5	92.0	98.3	90.6	98.7	93.7	99.5	
KY362518	49.0	69.9	86.4	94.9	92.0	98.3	91.0	98.7	93.7	99.5	
DQ211633	45.6	67.6	86.0	95.2	92.0	97.9	91.4	98.7	93.5	99.4	
KC867273	48.0	70.5	85.3	95.6	92.0	98.3	87.6	97.4	94.0	99.1	
DQ211635	48.2	68.9	86.8	95.6	91.3	97.6	91.0	98.7	93.0	99.5	III-Africa 3
AY900142	48.0	70.9	82.7	93.4	92.0	98.6	85.4	98.7	93.5	99.4	
DQ211634	47.1	71.3	83.8	93.8	91.7	98.6	86.7	99.1	93.2	99.4	
DQ157175	44.7	69.7	83.8	94.1	92.7	99.0	90.6	99.1	93.8	99.7	
HQ378184	44.7	68.4	83.5	94.1	92.0	98.6	90.1	99.1	94.3	99.7	
HQ378185	44.7	68.4	83.5	94.1	92.0	98.6	90.1	99.1	94.3	99.7	
HQ378187	47.5	68.4	84.6	94.5	92.4	99.0	89.7	99.1	94.0	99.7	
MH483988	47.1	68.9	83.8	94.5	92.0	99.0	88.8	98.7	93.8	99.5	
MF547416	48.0	68.0	83.1	93.8	92.7	99.0	88.0	98.3	94.0	99.7	IV-Asia 1
AB069671	52.4	73.2	84.9	94.9	93.1	97.9	89.3	98.3	92.7	99.2	
AB069672	52.8	74.4	84.9	94.5	92.7	97.9	88.0	97.9	92.6	99.2	
AJ538197	50.4	73.6	83.1	93.8	92.4	98.3	89.7	98.7	93.5	99.5	
AY179962	47.6	71.2	83.8	94.5	92.4	97.9	88.4	98.3	92.6	98.8	
AY223476	46.4	69.6	84.2	94.1	93.4	98.6	89.3	98.3	93.3	99.5	
JN572085	48.8	69.3	85.7	95.2	92.7	98.3	87.1	97.9	93.8	99.2	
JN572086	48.4	69.3	85.7	95.2	92.7	98.3	86.3	97.9	93.7	99.1	
MH461098	49.2	71.3	85.7	95.6	92.0	98.3	88.4	97.4	94.4	99.2	
AJ538199	49.6	71.3	84.9	95.6	91.7	98.3	85.0	94.4	94.4	99.2	
HM452306	49.6	71.3	84.9	95.6	91.3	98.3	88.4	97.4	94.3	99.2	
DQ446216	49.2	71.3	86.0	95.6	91.7	98.3	88.4	97.4	94.1	99.2	
MF289415	49.2	70.9	85.7	95.6	91.7	98.3	88.4	97.4	94.4	99.2	
Average % Identity/Similarity	54.1	73.2	85.4	95.2	92.5	98.5	89.6	98.0	93.8	99.3	

Figure 1: The percent identity and similarity of each GPC protein from 50 CCHFV GPC sequences compared to the Turkey2004 sequence.

Fifty GPC sequences from ticks, animals, and clinical cases were selected with representative sequences from each clade of CCHFV and spanning the widespread geographical distribution of the virus. A phylogenic tree of the 50 sequences was generated using Geneious Tree Builder, and clades were assigned based on previous publications[1,24]. Percent identity and similarity was calculated using William Pearson's *lalign* program run through the Swiss Institute of Bioinformatics ExPASy Bioinformatics Resource Portal (now available through the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) https://www.ebi.ac.uk/Tools/psa/lalign/). A color gradient was applied to the percent identity and similarity, with red highlighting areas with the lowest homology, and blue highlighting areas with the greatest homology to the Turkey2004 sequence.

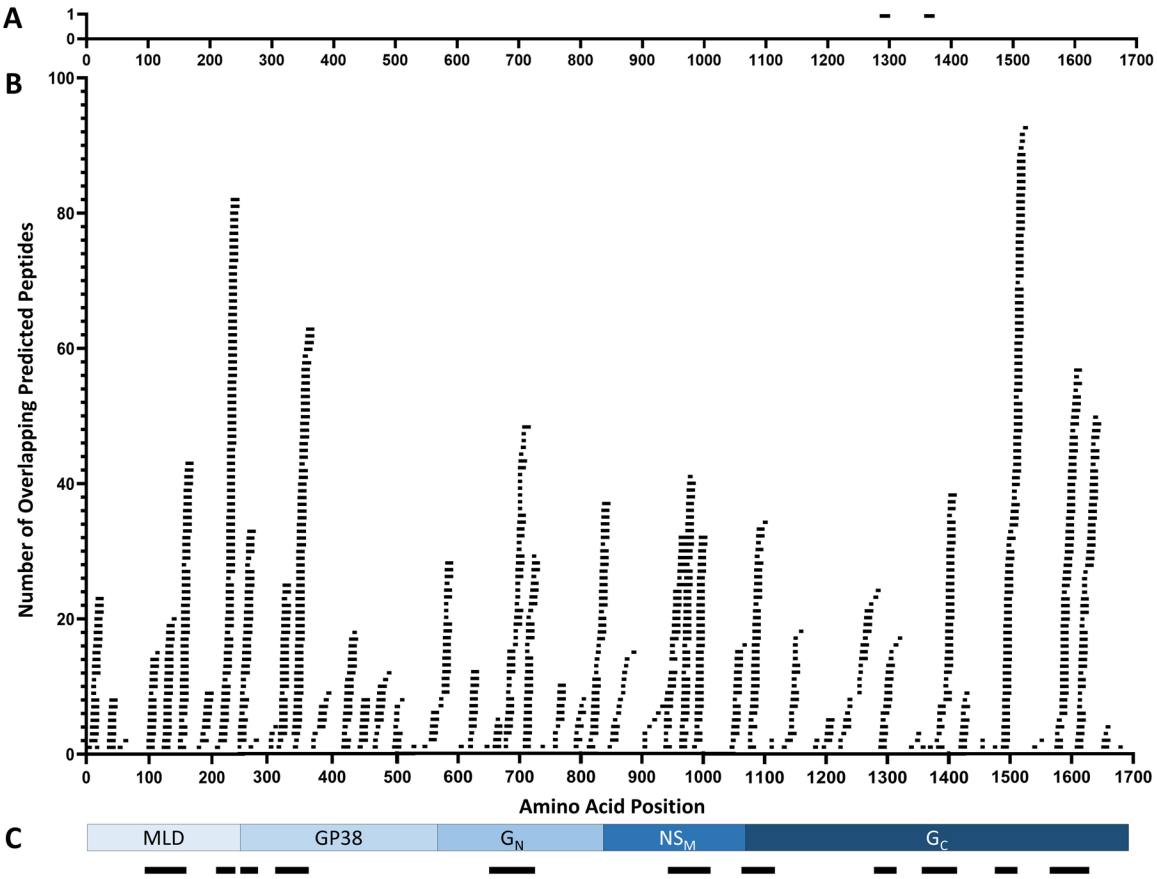
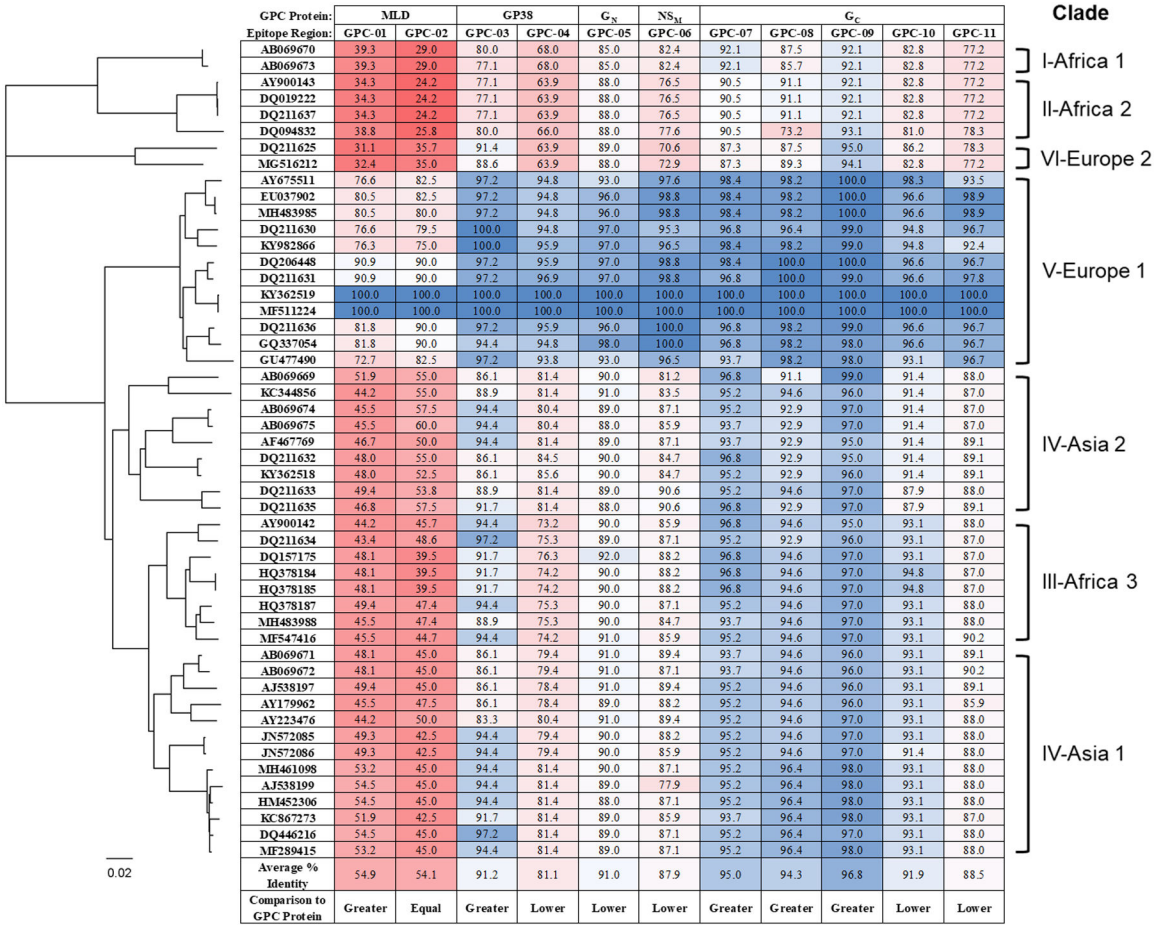


Figure 2: Alignment of predicted epitopes to the GPC.

(A) Location of two experimentally demonstrated T cell epitopes aligned to the GPC[13]. (B) Alignment of all predicted epitopes to the GPC, excluding the 21 epitopes predicted across known GPC cleavage sites, for a total of 1072 aligned epitopes. (C) Schematic diagram of the five GPC proteins and 11 regions of the GPC chosen for generation of the multi-epitope antigen (black bars).



GPC Protein:	MLD		GP38		G_N	NS_M	G_C				
Epitope Region:	GPC-01	GPC-02	GPC-03	GPC-04	GPC-05	GPC-06	GPC-07	GPC-08	GPC-09	GPC-10	GPC-11
AB069670	39.3	29.0	80.0	68.0	85.0	82.4	92.1	87.5	92.1	82.8	77.2
AB069673	39.3	29.0	77.1	68.0	85.0	82.4	92.1	85.7	92.1	82.8	77.2
AY900143	34.3	24.2	77.1	63.9	88.0	76.5	90.5	91.1	92.1	82.8	77.2
DQ019222	34.3	24.2	77.1	63.9	88.0	76.5	90.5	91.1	92.1	82.8	77.2
DQ211637	34.3	24.2	77.1	63.9	88.0	76.5	90.5	91.1	92.1	82.8	77.2
DQ094832	38.8	25.8	80.0	66.0	88.0	77.6	90.5	73.2	93.1	81.0	78.3
DQ211625	31.1	35.7	91.4	63.9	89.0	70.6	87.3	87.5	95.0	86.2	78.3
MG516212	32.4	35.0	88.6	63.9	88.0	72.9	87.3	89.3	94.1	82.8	77.2
AY675511	76.6	82.5	97.2	94.8	93.0	97.6	98.4	98.2	100.0	98.3	93.5
EU037902	80.5	82.5	97.2	94.8	96.0	98.8	98.4	98.2	100.0	96.6	98.9
MH483985	80.5	80.0	97.2	94.8	96.0	98.8	98.4	98.2	100.0	96.6	98.9
DQ211630	76.6	79.5	100.0	94.8	97.0	95.3	96.8	96.4	99.0	94.8	96.7
KY982866	76.3	75.0	100.0	95.9	97.0	96.5	98.4	98.2	99.0	94.8	92.4
DQ206448	90.9	90.0	97.2	95.9	97.0	98.8	98.4	100.0	100.0	96.6	96.7
DQ211631	90.9	90.0	97.2	96.9	97.0	98.8	96.8	100.0	99.0	96.6	97.8
KY362519	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
MF511224	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
DQ211636	81.8	90.0	97.2	95.9	96.0	100.0	96.8	98.2	99.0	96.6	96.7
GQ337054	81.8	90.0	94.4	94.8	98.0	100.0	96.8	98.2	98.0	96.6	96.7
GU477490	72.7	82.5	97.2	93.8	93.0	96.5	93.7	98.2	98.0	93.1	96.7
AB069669	51.9	55.0	86.1	81.4	90.0	81.2	96.8	91.1	99.0	91.4	88.0
KC344856	44.2	55.0	88.9	81.4	91.0	83.5	95.2	94.6	96.0	91.4	87.0
AB069674	45.5	57.5	94.4	80.4	89.0	87.1	95.2	92.9	97.0	91.4	87.0
AB069675	45.5	60.0	94.4	80.4	88.0	85.9	93.7	92.9	97.0	91.4	87.0
AF467769	46.7	50.0	94.4	81.4	89.0	87.1	93.7	92.9	95.0	91.4	89.1
DQ211632	48.0	55.0	86.1	84.5	90.0	84.7	96.8	92.9	95.0	91.4	89.1
KY362518	48.0	52.5	86.1	85.6	90.0	84.7	95.2	92.9	96.0	91.4	89.1
DQ211633	49.4	53.8	88.9	81.4	89.0	90.6	95.2	94.6	97.0	87.9	88.0
DQ211635	46.8	57.5	91.7	81.4	88.0	90.6	96.8	92.9	97.0	87.9	89.1
AY900142	44.2	45.7	94.4	73.2	90.0	85.9	96.8	94.6	95.0	93.1	88.0
DQ211634	43.4	48.6	97.2	75.3	89.0	87.1	95.2	92.9	96.0	93.1	87.0
DQ157175	48.1	39.5	91.7	76.3	92.0	88.2	96.8	94.6	97.0	93.1	87.0
HQ378184	48.1	39.5	91.7	74.2	90.0	88.2	96.8	94.6	97.0	94.8	87.0
HQ378185	48.1	39.5	91.7	74.2	90.0	88.2	96.8	94.6	97.0	94.8	87.0
HQ378187	49.4	47.4	94.4	75.3	90.0	87.1	95.2	94.6	97.0	93.1	88.0
MH483988	45.5	47.4	88.9	75.3	90.0	84.7	93.7	94.6	97.0	93.1	88.0
MF547416	45.5	44.7	94.4	74.2	91.0	85.9	95.2	94.6	97.0	93.1	90.2
AB069671	48.1	45.0	86.1	79.4	91.0	89.4	93.7	94.6	96.0	93.1	89.1
AB069672	48.1	45.0	86.1	79.4	91.0	87.1	93.7	94.6	96.0	93.1	90.2
AJ538197	49.4	45.0	86.1	78.4	91.0	89.4	95.2	94.6	96.0	93.1	89.1
AY179962	45.5	47.5	86.1	78.4	89.0	88.2	95.2	94.6	96.0	93.1	85.9
AY223476	44.2	50.0	83.3	80.4	91.0	89.4	95.2	94.6	97.0	93.1	88.0
JN572085	49.3	42.5	94.4	79.4	90.0	88.2	95.2	94.6	97.0	93.1	88.0
JN572086	49.3	42.5	94.4	79.4	90.0	85.9	95.2	94.6	97.0	91.4	88.0
MH461098	53.2	45.0	94.4	81.4	90.0	87.1	95.2	96.4	98.0	93.1	88.0
AJ538199	54.5	45.0	94.4	81.4	89.0	77.9	95.2	96.4	98.0	93.1	88.0
HM452306	54.5	45.0	94.4	81.4	88.0	87.1	95.2	96.4	98.0	93.1	88.0
KC867273	51.9	42.5	91.7	81.4	89.0	85.9	93.7	96.4	98.0	93.1	87.0
DQ446216	54.5	45.0	97.2	81.4	89.0	87.1	95.2	96.4	97.0	93.1	88.0
MF289415	53.2	45.0	94.4	81.4	89.0	87.1	95.2	96.4	98.0	93.1	88.0
Average % Identity	54.9	54.1	91.2	81.1	91.0	87.9	95.0	94.3	96.8	91.9	88.5
Comparison to GPC Protein	Greater	Equal	Greater	Lower	Lower	Lower	Greater	Greater	Greater	Lower	Lower

Figure 3: The percent identity of each multi-epitope region from 50 CCHFV GPC sequences compared to the Turkey2004 sequence.

Each of the 11 CCHFV Turkey2004 multi-epitope region sequences were compared individually to the same region in 50 selected sequences across all CCHFV clades using William Pearson's *align* program run through the Swiss Institute of Bioinformatics ExPASy Bioinformatics Resource Portal (https://embnet.vital-it.ch/software/LALIGN_form.html). The average percent identity of each region was compared to the average percent identity of the GPC protein the region originated from. A color gradient was applied to the percent identity and similarity, with red highlighting areas with the lowest homology, and blue highlighting areas with the greatest homology to the Turkey2004 sequence.

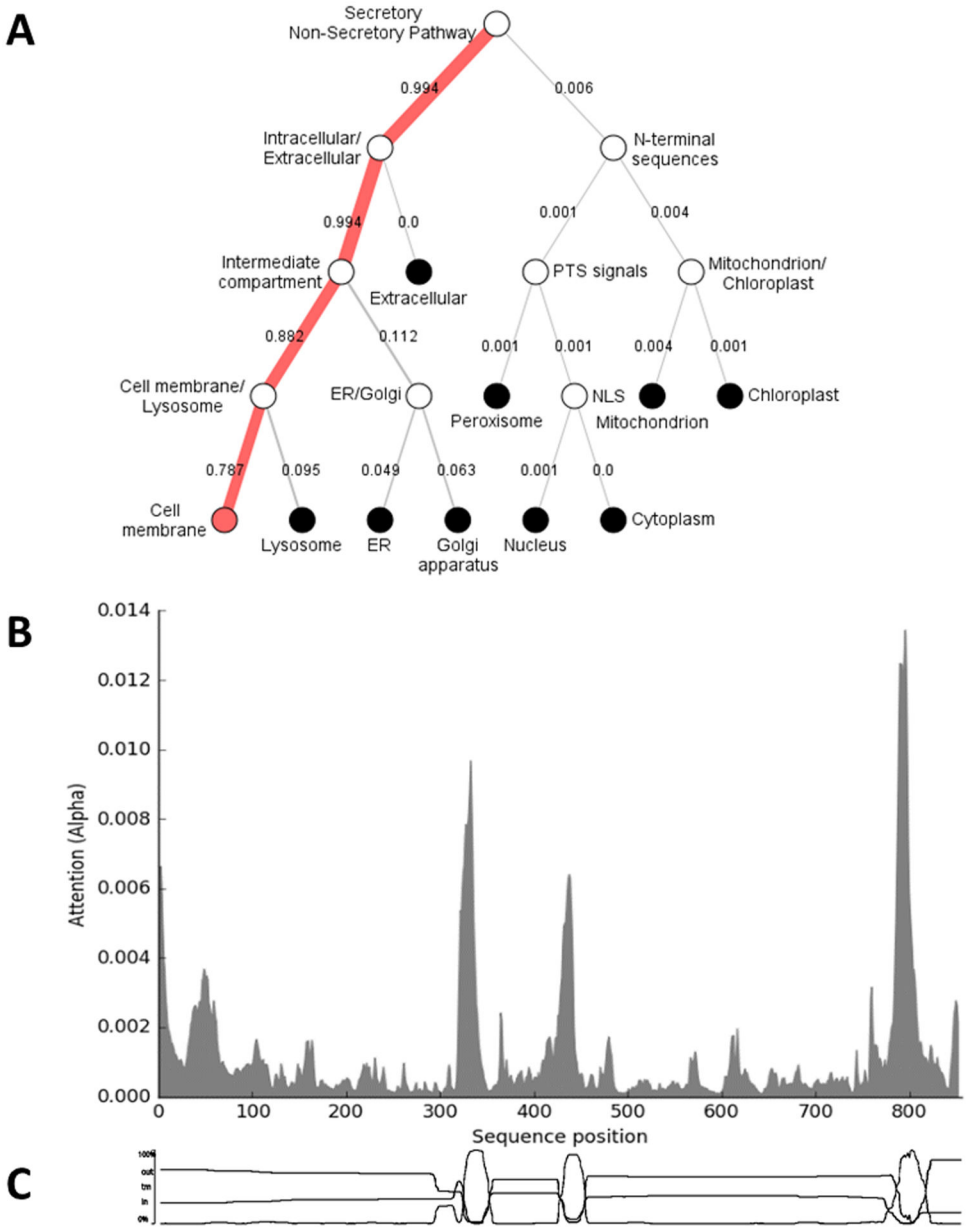


Figure 4: Subcellular localization prediction and residues important for localization of *EPIC*.
Results of the submission of the multi-epitope antigen sequence (*EPIC*) to the DeepLoc 1.0 Server (https://services.healthtech.dtu.dk/service.php?DeepLoc-1.0) for prediction of subcellular localization. (A) Hierarchical tree output showing the most likely localization of the antigen with numbers representing percent likelihood. (B) Graph of the importance of each residue for localization. (C) Location of the three transmembrane domains with EPIC, as shown by the transmembrane prediction function of Geneious Prime software.

Table 1:
Predicted epitopes by type per GPC region.

Total number of predicted CTL and HTL epitopes from using the NetCTL 1.2 and NetMHCII 2.3 servers. Epitopes were aligned to the Turkey2004 GPC sequence to determine their corresponding GPC region.

Number of Predicted Epitopes					
GPC Region	Type of Epitope				Total
	CTL	HLA-DR	HLA-DQ	HLA-DP	
MLD	19	39	140	11	209 (19.1%)
MLD-GP38 *	0	0	0	0	0 (0%)
GP38	41	94	17	25	177 (16.2%)
GP38-G_N *	1	0	0	0	1 (0.1%)
G_N	54	42	10	37	143 (13.1%)
G_N-NS_M *	2	5	5	0	12 (1.1%)
NS_M	40	43	66	14	163 (14.9%)
NS_M-G_C *	1	3	4	0	8 (0.7%)
G_C	98	130	100	52	380 (34.8%)
Total	256	356	342	139	1093 (100%)

Epitopes located across known cleavage sites (indicated with *) were excluded from further consideration.

Table 2:

Predicted epitopes by type per selected multi-epitope region.

Eleven multi-epitope regions were selected from alignment of the predicted epitopes to the GPC to generate the multi-epitope antigen. The length in residues of each region, GPC protein the region originated from, and the number and type of predicted epitopes included in each selected region are shown. A total of 812 epitopes (75.7% of total predicted epitopes) were included within the selected regions.

Number of Predicted Epitopes							
Epitope Region	**Length (AA)**	**GPC Protein**	**Type of Epitope**				**Total**
			CTL	**HLA-DR**	**HLA-DQ**	**HLA-DP**	
GPC-01	77	MLD	6	11	62	0	**79**
GPC-02	40	MLD	4	15	63	0	**82**
GPC-03	36	GP38	9	19	0	7	**35**
GPC-04	97	GP38	17	52	14	18	**101**
GPC-05	100	G_N	32	9	7	35	**83**
GPC-06	85	NS_M	15	36	45	10	**106**
GPC-07	63	G_C	11	16	23	0	**50**
GPC-08	56	G_C	8	2	10	0	**20**
GPC-09	101	G_C	14	22	16	0	**52**
GPC-10	58	G_C	11	39	44	0	**94**
GPC-11	92	G_C	21	43	0	46	**110**
Total			**148**	**264**	**284**	**116**	**812**