Clinical Microbiology: Open Access
ISSN: 2327-5073
OPEN ACCESS Freely available online
Research Article

Exploring Crimean Congo Hemorrhagic Fever Virus Glycoprotein M to Predict Multi-Epitopes Based Peptide Vaccine Using Immunoinformatics Approach

Samra Obai Mohamed[1,2], Yassir A. Almofti[1*], Khoubieb Ali Abd Elrahman[3]

[1]Department of Molecular Biology and Bioinformatics, College of Veterinary Medicine, University of Bahri, Khartoum, Sudan;[2]Department of Pharmacy, National Ribat University, Khartoum, Sudan;[3]Department of Pharmaceutical Technology, College of Pharmacy, University of Medical Science and Technology, Khartoum, Sudan

ABSTRACT

Crimean Congo Hemorrhagic Fever (CCHF) is hemorrhagic viral disease caused by CCHF Virus (CCHFV) with case fatality rate of up to 40%. This study aimed to design multi epitopes vaccine from glycoprotein M to elicit immune response. Strains of CCHFV were used to construct a phylogenetic tree. IEDB tools were exploited to predict B and T cell epitopes and calculation of the population coverage of each predicted epitope. The vaccine protein comprises 599 amino acids and was potentially antigenic and non-allergic. Physical and chemical properties showed that the vaccine was stable, contains aliphatic side chains, hydrophilic and with thermal stability. The vaccine demonstrated no homology with human proteins. Secondary and tertiary structures of the vaccine were predicted, refined and validated by rampage plot. Structural errors were assessed by proSA web server that showed a Z-score of -2.97. The vaccine was soluble in comparison to solubility of *E. coli* proteins. Molecular docking with TLR4 provided binding energy of -1135.5 Kcal/mol and -1301.4 Kcal/mol for chain A and chain B, respectively. *In silico* cloning demonstrated the potential clonability of the vaccine protein in pET28a (+) vector resulting in efficient expression and translation. Clinical trial analysis *via in vivo* and *in vitro* studies is required.

Keywords: Glycoprotein M; Vaccine; Immune informatics

Abbreviations: CCHF: Crimean-Congo Hemorrhagic Fever; IEDB: Immune Epitope Data Base; NCBI: National Center for Biotechnology Information

INTRODUCTION

Crimean-Congo Hemorrhagic Fever (CCHF) is an important hemorrhagic viral disease that infect human. The disease is caused by CCHF virus that belongs to the genus Nairovirus (family *Bunyaviridae*) [1]. The disease is considered as a tick-borne illness occurring in many continents such as in Asia, Africa and Europe. CCHF symptoms are generally including high sudden onset of fever, severe headache, and chills. In addition to that symptoms may include gastrointestinal tract disturbances, such as nausea diarrhea, and vomiting. Hemorrhagic manifestations mainly associated with severe cases and appeared in the form of large regions of ecchymosis, instead of frank bleeding [1]. Exposure to ticks particularly the genus *Hyalomma* or upon direct contact with the people or animals infected with the virus are considered as the major factors for infection [1].

The CCHFV genome demonstrated three segments of negatively-stranded RNA. These segments are named as L, M, and S segments [2]. The L segment is shown to be responsible for coding the non-structural RNA polymerases. While the M segment is responsible for coding the surface glycoproteins and the S segment is responsible for coding nucleocapsid [3]. The virions of the CCHFV are characterized by spherical structure with nearly a diameter of 90nm to 100 nm. Moreover, the virus considered as an enveloped particle with negatively ssRNA genome. All the three segments of the genome contain one Open Reading Frame (ORF) that is flanked by noncoding regions [3]. However, the structural proteins (four structural proteins) are recognized to be encoded by the RNA-dependent RNA polymerase. For instance, the Large (L) segment expresses and encodes the L protein; the Medium (M) segment encodes the mature glycoproteins Gn and Gc and the small (S) genome segment encodes the N nucleoprotein as for other negative-sense RNA viruses [4].

The geographical distribution of the tick vector *Hyalomma*

Correspondence to: Almofti YA, Department of Molecular Biology and Bioinformatics, College of Veterinary Medicine University of Bahri, Khartoum, Sudan, Tel: 00249901589515; Fax: +249 (155) 882052; P.O Box: 1660 – Khartoum; E-mail: yamofti99@gmail.com

Received: May 7, 2021 **; Accepted:** May 21, 2021**; Published:** May 28, 2021

Citation: Mohamed SO, Almofti YA, Abd Elrahman KA (2021) Exploring Crimean Congo Hemorrhagic Fever Virus Glycoprotein M to Predict Multi-Epitopes Based Peptide Vaccine Using Immunoinformatics Approach. Clin Microbiol. 10: 122.

marginatum marginatum is commonly corresponded with the distribution of CCHFV. It is noteworthy that sporadic outbreaks cases of CCHFV previously occurred in regions of Europe, Asia, Middle East and Africa. However, by the end of the 21st century, Turkey, Iran and the former Yugoslavia demonstrated frequent outbreaks. Accordingly, the CCHFV is suggested as a migrating virus pathogen [5]. Moreover, the adaptability of the tick vectors to different environments in the geographic regions played an important role in the wide distribution and dissemination of CCHFV [6].

Although many outbreaks were reported for CCHFV infections in animals and/or humans, there is no efficient vaccine or good drug to combat the virus infection [7]. The genome of the CCHFV is shown to incorporate mutations that mainly attributed to the error-prone nature of the polymerase segment. Also, the virus demonstrated high rate of recombination in its RNA. Therefore, the process of initiating a vaccine or developing antiviral drug to combat the action of CCHFV is associated with difficulty [8]. The design or construction of multi-epitopes-based vaccines against the deadly viruses became more popular due to its increased safety and potency [9,10]. Recently the application of reverse vaccinology or the use of bioinformatics tools associated with the immunological data (termed Immuno informatics) is becoming a prominent field for vaccine design. Immuno informatics aid in construction of new vaccines *via* identification of possible B and T lymphocytes epitopes that properly interacted with Human Leukocyte Antigens (HLA) [11,12]. This novel approach demonstrated potential efficacy in some human diseases such as malaria, multiple sclerosis, AIDS and tuberculosis with satisfaction results [13-16]. The *in silico* vaccine design with potential conserved multiple epitopes *via* the immune informatics approach in order to reduce the deleterious effects of the CCHFV is crucially needed. The purpose of such vaccine is to elicit an immune response by provoking the B and T cells of the vaccinated individual upon exposures CCHFV. In some cases, the difference in individual's immune systems might result in inadequate response to the virus infection. Thus, inadequate response would be generated with no protection against the second exposure to the virus. This is considered as one of the factors that explaining the immunization failure in those individuals [17].

The preparation of vaccines that mainly relayed on the biochemical experiments are considered as time consuming, costly and are not always appropriately efficient. Moreover, such kinds of vaccine types might constitute multiple unnecessary particles from the virus proteins causing many undesirable reactions and allergic responses [18,19]. Therefore, *in silico* epitopes prediction of proper protein residues would assist in synthesis of peptide-based vaccines with good immunogenicity and least allergenic impact [18]. This study aimed to design a vaccine against Crimean-Congo hemorrhagic fever virus using glycoprotein M for peptides prediction that stimulates protective immune response.

MATERIALS AND METHODS

Viral proteome retrieval

The viral whole proteome of Crimean Congo Hemorrhagic Fever Virus (CCHFV) was retrieved from National Center for Biotechnology Information (NCBI). The virus had three proteins namely glycoprotein M (NP_950235.1), putative polyprotein (YP_325663.1) and the nucleoprotein (NP_950237.1). The length of these proteins was 1684, 3945 and 482 amino acids, respectively.

Proteins antigenicity and transmembrane topology

The VaxiJen v2.0 server was used to examine the antigenicity of each viral protein and to determine their potent antigenicity. The antigenic proteins were further analyzed for transmembrane topology using TMHMM. Proteins that demonstrated antigenicity and transmembrane topologies were allowed for further analysis. In this essence Glycoprotein M (NP_950235.1) and putative polyprotein (YP_325663.1) were shown to be antigenic in VaxiJen v2.0 server passing the threshold of (0.4) and they scored 0.5145 and 0.4306, respectively, while the nucleoprotein (NP_950237.1) was shown to be non-antigenic (scored 0.3025). For the presence of the transmembrane helices using TMHMM, only glycoprotein M demonstrated five transmembrane helices compared to polyprotein (YP_325663.1) with no transmembrane helices predicted. Accordingly, the Glycoprotein M (NP_950235.1) was targeted for prediction of multi epitopes that act as vaccine candidates eliciting both B and T lymphocytes.

Protein sequence retrieval of glycoprotein M

A set of available 43 virulent strains of CCHFV glycoprotein M were retrieved from the NCBI. These sequences were retrieved from different geographical areas (Spain, South Africa, India, UAE, Nigeria, USA, Pakistan, Uganda, Namibia, Afghanistan, Iran and Sudan). The accession numbers, year of collection and country of the retrieved strains were provided (Table 1).

Table 1: CCHFV glycoprotein M retrieved strains, their accession numbers, date of collection and their countries.

Accession number	Date of collection	Country	Accession number	Date of collection	Country
ATG31911	2016	Spain	ARB51462	1987	South Africa
AIE16137	1990	South Africa	ARB51455	1996	Nigeria
AIE16136	1990	South Africa	AAA86616	2001	USA
AEO72051	2011	India	ABA39299	2002	South Africa
AEO72050	2011	India	ABA39298	1999	South Africa
AEO72049	2011	India	CAD61345	2000	Pakistan
AEO72048	2011	India	ASV45881	2014	Spain
ASW20660	1998	UAE	ARB51465	1981	Uganda
ASW20657	1995	UAE	APB09910	2012	Nigeria
NP_950235*	2008	Nigeria	ASW22387	1984	South Africa
AHL45281	2007	Iran	ASW22385	1985	South Africa

AEI70589	2009	Sudan	AAM48106	1999	Nigeria
AEI70587	2008	Sudan	ASW22383	1981	South Africa
AEI70586	2008	Sudan	ASW22382	1987	Namibia
ADQ57289	2009	Afghanistan	APZ76784	2016	India
AAW84281	1997	South Africa	ASW22386	1984	South Africa
AAW84280	2004	South Africa	APG38034	1965	Nigeria
AF467768_1	2005	Nigeria	QBW95916	2015	Sudan
ABB30034	1987	South Africa	AZS19000	2017	Iran
ABD98125	2005	Iran	AIN41199	2015	USA
ABD98124	2007	Iran	AIN41197	2015	USA
AWX63620	2018	USA			

Note: *Ref: Reference sequence.

Phylogenetic tree construction

The closed relationship between the sequences of glycoprotein M from the retrieved strains was analyzed using MEGA7.0.26 (7170509) software [20]. The analysis aimed to construct a phylogenetic tree to assess the common ancestor of the retrieved strains.

Sequence alignment and epitopes conservancy

Glycoprotein M sequences were aligned using Clustal W, Multiple Sequence Alignment (MSA) tools, in the BioEdit program, version 7.0.9.0 [21]. MSA analysis was performed to obtain 100% conserved epitopes from the retrieved strains interacting against B and T cells lymphocytes.

B-cell epitopes prediction methods

B cell epitopes is a part of the immunogen that bound to B lymphocytes. Thus, B cells differentiated into memory cells and plasma cells that excrete immunoglobulins. The epitopes of the B cells are surface accessible and antigenic [22]. The classical propensity scale and Hidden Markov Model (HMM) from Immune Epitope Data Base (IEDB) analysis resource were used for the following prediction aspects: BepiPred linear epitopes prediction was used to provide linear B-cell epitopes from the conserved regions of the retrieved strains [23]. Surface accessible epitopes were analyzed by Emini surface accessibility prediction tool [24]. For antigenic epitopes, kolaskar and tongaonker method was performed to obtain the antigenic epitopes [25].

Prediction of cytotoxic T-cell epitopes interacting with MHC class I

The peptide-MHCI binding prediction was analyzed by the IEDB MHC-I tool. Peptides bound to MHC molecules were analyzed by Artificial Neural Network method (ANN 4.0). The analysis was only performed for the most frequent alleles (HLA-A, HLA-B and HLA-C). Epitope's length was set to 9 mers and the half-maximal inhibitory concentration (IC50) was equal or less than 100 nm (≤ 100 nm) [26-33].

Prediction of helper T-cell epitopes interacting with MHC class II

The antigenicity of the predicted B cell epitopes was based on the score of the kolaskar and tongaonker antigenicity method since this prediction method was impeded within the IEDB server [25]. While the antigenicity of T cell epitopes was assessed by VaxiJen v2.0 server with the default threshold (0.4). Allertop server was used to investigate the allergic and nonallergic epitopes [34] and the toxicity of the epitopes was assessed by Toxinpred server [35].

Determination of the antigenicity, allergenicity and toxicity of the predicted epitopes

The peptides-MHC-II molecules binding prediction was assessed by the IEDB MHC-II prediction tool [31]. The human reference allele's set were investigated. MHC-II groove has the potential to bind to numerous epitopes lengths making the epitopes prediction process inaccurate [32]. The NN-align 2.3 (NetMHCII 2.3) was used to obtain the binding core epitopes. The peptides length was set to the default value (15 mer). Half-maximal inhibitory concentrations (IC50) score was equal or less than 1000 nm (≤ 1000 nm) [33].

Determination of the population coverage

All probable binders of MHC class I and MHC class II epitopes from CCHFV Glycoprotein M strains were evaluated for population coverage compared to the whole world population using the IEDB population coverage tool.

Assemblage of the multi-epitope vaccine sequence

The sequence of the vaccine was assembled by joining epitopes of B cells, T cytotoxic and T helper lymphocytes that passed the criteria in each prediction tool and were shown to be antigenic, nonallergic and nontoxic. To promote the immunogenicity of the putative vaccine the 50S ribosomal protein L7/L12 of Mycobacterium tuberculosis (strain ATCC 25618/H37Rv, uniprot, P9WHE3) was used as an adjuvant. The adjuvant was attached to putative vaccine *via* Glu-Ala-Ala-Ala-Lys (EAAAK) linker at the amino terminal. The cytotoxic T lymphocytes epitopes were linked *via* Tyr-Ala-Ala (YAA) linkers while the helper T lymphocytes and B cell epitopes were linked *via* Gly-Pro-Gly-Pro-Gly (GPGPG) linkers. A six His-tag was added to the C terminal to purify and identify the putative vaccine construct.

Physical and chemical properties of the putative vaccine

ProtParam server is a tool used to calculate numerous physical and chemical characteristic of a given protein sequence such as the Molecular Weight MW, isoelectric point (pI), atomic composition, amino acid composition, extinction coefficient, instability index, estimated half-life, aliphatic index and grand average of hydropathicity (GRAVY) [36]. Thus, this tool was exploited to compute the physical and chemical properties of the putative vaccine.

Homology and transmembrane topology assessment

Protein BLAST for the vaccine protein was performed to analyze the homology of the vaccine protein sequence with the human proteome *via* NCBI BLASTp. The purpose of the homology of vaccine protein with human proteome was to avoid autoimmunity [36]. The result of

the homology score must be no or least homology (<40%) [37]. Also, the presence of transmembrane helixes of the vaccine protein was analyzed *via* TMHMM server. Transmembrane helixes were shown to impact the process of the expression of the proteins [38-40].

Secondary structure prediction and solvent accessibility prediction

Raptor X server was to predict the Secondary Structure (SS), solvent accessibility (ACC), and disorder regions (DISO) in the vaccine sequence [41,42]. The SS provided results in two modes, among them, is the 3-state secondary structure that comprised of α-helix (H), β-sheet (E) and coiled regions (C). The ACC model was predicted by the 3-state solvent accessibility method of RaptorX as solvent exposed (E), medium (M) and buried residues (B). The order/disorder prediction was dependent on a cutoff value at 0.25 [43].

Prediction of vaccine tertiary structure, re nement and validation

I-TASSER server is a hierarchical approach to protein structural and functional prediction. The server integrates the platform for automated protein structure and function prediction based on the sequence-to-structure-to-function paradigm [44]. The server was used to predict the three-dimensional structure (3D) of the putative vaccine. For the refinement of the tertiary structure (3D) obtained by I-TASSER server, the GalaxyRefine server was used. GalaxyRefine improves the quality of both global and local structures of the refined protein [45-47]. Model validation was initially performed using ProSA-web which calculates the overall quality score for a specific input protein PDB structure. Secondly the model validation was performed using Ramachandran plot in RAMPAGE server [48,49].

Prediction of the discontinuous B-cell epitopes

ElliPro in the IEDB epitopes analysis resources was used to predict the discontinuous B-cell epitopes. Ellipro server used different algorithms for the output epitopes. These includes protein shape approximation as an ellipsoid, Protrusion Index (PI) estimation, and based on PI scores, the cluster neighbouring residues would be determined [50]. Thus, an ellipsoid with PI value of 0.9 demonstrated 90% of the protein residues are located inside the Ellipsoid. Residues that showed greater scores provided greater solvent accessibility. Discontinuous epitopes are defined based on PI values and are clustered based on the distance R (in Å between residue's centers of mass). The greatest R distance the greater predicted discontinues epitopes. The more prominent R value the more will be the number of discontinuous epitopes predicted [50].

Solubility of the putative vaccine

Solubility of the putative vaccine is a paramount feature through vaccine design, from the stage of recombinant protein production to the stage of bio-therapeutics. The online server Protein–Sol is available at processes amino acid sequence and calculates predicted solubility and other properties. The result returned in a graphical format and as a text file and used for prediction of the putative vaccine solubility [51]. Accordingly, the protein solubility scores larger than 0.45 is expected to be soluble compared to the average solubility of *E. coli* proteins from the experimental solubility dataset and vice versa [51].

Molecular docking of putative vaccine with Toll like Receptor 4 (TLR4)

ClusPro 2.0 is a web-based server for the direct docking of two interacting proteins. The computational steps of the server were as follows: firstly, rigid body docking by sampling billions of conformations, secondly, root-mean-square deviation (RMSD) based clustering of the 1000 lowest energy structures produced to obtain the largest fine clusters model of the complex. Thirdly refinement of selected structures assessed using energy minimization [52,53]. Thus, the putative vaccine PDB file ameliorated by Galaxyrefiner was considered as the ligand and submitted to the server with TLR4 (PDB: 4G8A) as receptors. The docking process was performed with TLR4 chain A and chain B separately. The advanced method in the server was used as a docking method. The complex of the receptor and the vaccine was visualized by PyMOL visualization tool.

In silico cloning

To guarantee the translation of the putative vaccine in the selected host, an *in silico* cloning was performed *via* Java Codon Adaptation Tool (JCAT). This server performs codon usage adaptation to the most sequences of the prokaryotic and some eukaryotic organisms [54,55]. The JCAT server was used to convert the protein sequence of the chimeric vaccine into DNA sequence. In the JCAT, Codon Adaptation Index (CAI) best score is 1.0 but more than 0.8 is considered a good score [56]. The favorable GC content of a sequence ranged between 30%–70%. The BamH1 and Xho1 restriction enzymes sequences were added to the ends of the DNA sequence. The SnapGene restriction cloning module was used to insert the DNA sequence into pET28a (+) vector between BamH1 and Xho1 [57,58].

RESULTS

Phylogeny analysis

As shown in Figure 1 the African strains were clustered closed to each other, however, strains from USA and Spain were clustered with closed relationship with the African strains. This might indicate the immigration between these countries participated in the transfer of the virus between them. Moreover, strains from the Asian countries were clustered together indicating closed molecular evolution between them. In general, all the strains demonstrated molecular divergence concerning their common ancestors.

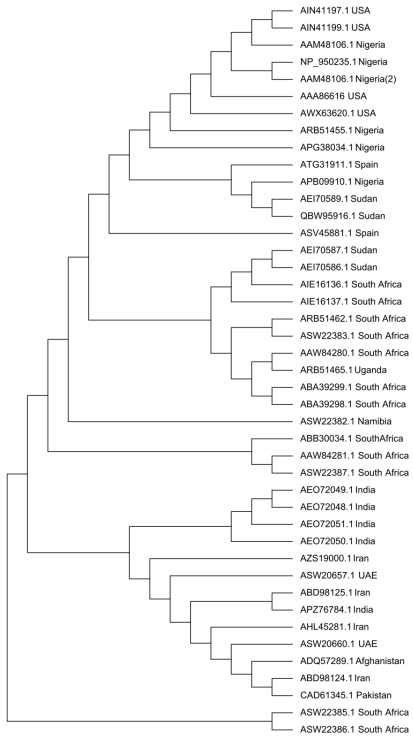

Figure 1: Phylogenetic tree of glycoprotein M retrieved strains. The retrieved strains demonstrated divergence in their common ancestors.

Epitopes conservancy

The retrieved sequences showed areas of conservancy and non-conservancy upon sequence alignment. Conserved regions (conserved epitopes) were realized by identical amino acids residues among all retrieved sequences. Epitopes that showed 100% conservancy in the tools of B and T lymphocytes were incorporated in the vaccine structure while the non-conserved epitopes were excluded.

B-cell epitopes prediction

The thresholds for B cells epitope predicition tools. Multiple epitopes showed interaction against B cells. These epitopes were found to be linear, surface accessible and demonstrate antigenicity upon using the prediction tools of B cells. Thus, five epitopes were elected as B cell epitopes candidates depending on their high Emini and kolaskar scores and were shown in Table 2. Moreover, the five epitopes were shown to be nonallergic in Allertop server and nontoxic in toxinpred server (Figure 2).

Prediction of cytotoxic T-cell epitopes interacting with MHC class I

Based on ANN 4.0 with (IC50) ≤100, a total of 32 epitopes were shown interacting with MHC-1 alleles (HLA-A, HLA-B and HLA-C). The 32 epitopes were further analyzed for antigenicity, allergenicity and toxicity and only 9 out of 32 epitopes passed these tools. The population coverage of each of the 9 epitopes was computed against the whole world (Table 3).

Prediction of T helper cell epitopes interacting with MHC class Π

Based on NN-align 2.3 (NetMHCII 2.3) and IC50 ≤ 1000, total of 113 core epitopes interacted with MHC-Π alleles. Upon antigenicity, allergenicity and toxicity analysis only 21 out of 113 epitopes passed these criteria. The population coverage of each of the 21 epitopes was computed against the whole world (Table 4).

Assembly of multi-epitopes putative vaccine

The vaccine construct was designed from epitopes interacting with B and T lymphocytes. The total number of the amino acids in the vaccine construct was 599 amino acids including the adjuvants, linkers used and the six His-tag sequence (Table 5). The putative vaccine demonstrated antigenicity in Vaxigen server and was nonallergen in the Allertop server.

The physical and chemical properties of the putative vaccine

The physical and chemical properties were shown in Table 5. The pI of the vaccine classified the vaccine as acidic in nature. The total number of negative -R and positive +R charged residues were 52 and 44 respectively. The instability index (II) classified the protein as stable and the GRAVY (-0.110) classified the vaccine construct as hydrophilic.

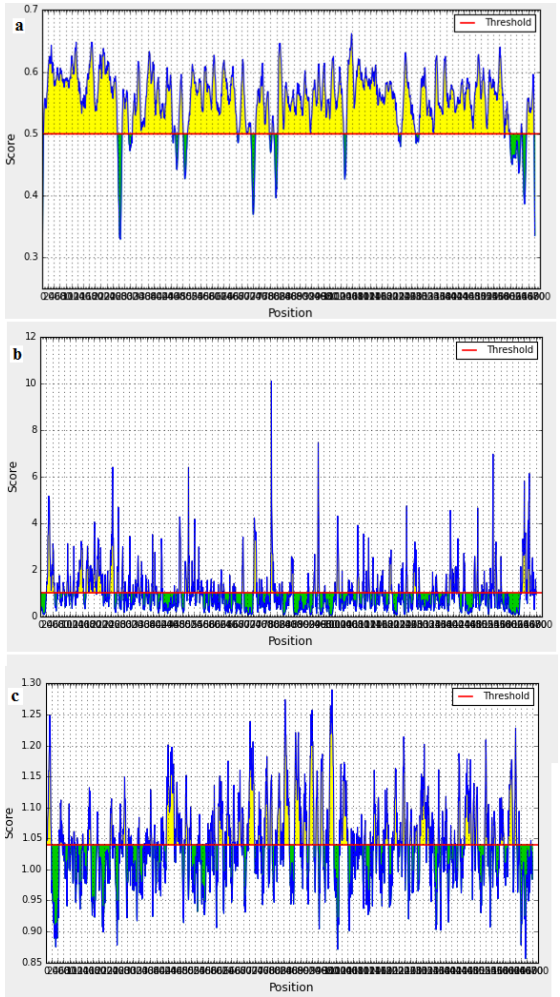

Figure 2: B-cell epitopes prediction tools were (a) Bepipred linear epitope, threshold value 0.35 (b) Emini surface accessibility, threshold value 1.000 and (c) Kolaskar and Tongaonkar antigenicity methods, threshold value 1.040. The red line showed the threshold for each tool.

Table 2: List of conserved B-cell epitopes predicted by different scales from Glycoprotein M protein of Crimean Congo Hemorrhagic Fever Virus.

Peptide	Start	End	Length	Emini surface accessibility *	kolskar antigenicity *	Allergenicity	Toxicity
PVQSAP	836	841	6	1.003	1.1	Nonallergen	Nontoxin
EWPKAT	1159	1164	6	1.98	1.045	Nonallergen	Nontoxin
IQQKLP	1270	1275	6	1.293	1.071	Nonallergen	Nontoxin
YTGVTQH	1370	1376	7	1.071	1.051	Nonallergen	Nontoxin
PQLDLKARPTY	1409	1419	11	2.993	1.041	Nonallergen	Nontoxin

Note: * Emini surface accessibility and Kolskar antigenicity thresholds were 1.000 and 1.040 respectively.

Table 3: The 9 predicted epitopes that interacted against MHC1 and were shown to be antigenic, non-allergic and nontoxic. The Population coverage (PC) of each epitope was calculated against the whole world.

Peptide	Start	End	Vaxigen antigenicity*	Allergenicity	Toxicity	Population coverage
RLGSELGCY	565	573	0.8291 Antigenic	nonallergen	Non-Toxin	10.67%
TECLCPYEA	1025	1033	0.7933 Antigenic	nonallergen	Non-Toxin	3.45%
CLCPYEALV	1027	1035	0.5713 Antigenic	nonallergen	Non-Toxin	39.08%
TSLETSLSI	1056	1064	0.8463 Antigenic	nonallergen	Non-Toxin	4.41%
TSLSIEAPW	1060	1068	0.9887 Antigenic	nonallergen	Non-Toxin	7.26%
ESKLLTVSV	1129	1137	0.5304 Antigenic	nonallergen	Non-Toxin	2.50%
TEAIVCVEL	1232	1240	1.1768 Antigenic	nonallergen	Non-Toxin	7.81%
IEAGTRFNL	1249	1257	1.0181 Antigenic	nonallergen	Non-Toxin	11.13%
FSAMPKTSL	1510	1518	0.6103 Antigenic	nonallergen	Non-Toxin	10.42%

Note: * Vaxigen antigenicity threshold was 0.4.

Table 4: The 21 predicted epitopes that interact against MHCII were shown to be antigenic, non-allergic, and non-toxic. The Population Coverage (PC) of each epitope was calculated against the whole world.

Core epitope	Peptide	Start	End	Vaxijen antigenicity*	Allergenicity	Toxicity	Population coverage
LLEWCKRNL	EDTEGLLEWCKRNLG	278	292	0.4192 Antigenic	Nonallergen	Non-Toxin	64.39%
LEWCKRNLG	DTEGLLEWCKRNLGL	279	293	0.4411 Antigenic	Nonallergen	Non-Toxin	10.54%
LGCYTINRV	QRLGSELGCYTINRV	564	578	1.4743 Antigenic	Nonallergen	Non-Toxin	18.41%
YLNLERIPW	VEETELYLNLERIPW	787	801	0.5035 Antigenic	Nonallergen	Non-Toxin	68.60%
NLERIPWVV	TELYLNLERIPWVVR	790	804	1.2163 Antigenic	Nonallergen	Non-Toxin	18.41%
LERIPWVVR	TELYLNLERIPWVVR	790	804	1.2965 Antigenic	Nonallergen	Non-Toxin	2.33%
TSLETSLSI	LNSTSLETSLSIEAP	1053	1067	0.8463 Antigenic	Nonallergen	Non-Toxin	28.34%
STSLETSLS	LNSTSLETSLSIEAP	1053	1067	0.5684 Antigenic	Nonallergen	Non-Toxin	11.53%
TSLSIEAPW	NSTSLETSLSIEAPW	1054	1068	0.9887 Antigenic	Nonallergen	Non-Toxin	49.64%
WPHSRNWRC	TCLHKEWPHSRNWRC	1179	1193	0.4645 Antigenic	Nonallergen	Non-Toxin	18.23%
WCWGVGTGC	WRCNPTWCWGVGTGC	1191	1205	1.7192 Antigenic	Nonallergen	Non-Toxin	56.45%
WGVGTGCTC	WGVGTGCTCCGLDVK	1199	1213	0.4198 Antigenic	Nonallergen	Non-Toxin	56.45%
TEAIVCVEL	IKTEAIVCVELTSQE	1229	1243	1.1768 Antigenic	Nonallergen	Non-Toxin	27.13%
VELTSQERQ	VCVELTSQERQCSLI	1235	1249	1.8525 Antigenic	Nonallergen	Non-Toxin	67.30%
IEAGTRFNL	RQCSLIEAGTRFNLG	1244	1258	1.0181 Antigenic	Nonallergen	Non-Toxin	64.95%
EAGTRFNLG	RQCSLIEAGTRFNLG	1244	1258	0.7560 Antigenic	Nonallergen	Non-Toxin	84.37%
KVLSASTVC	QKVLSASTVCKLQSC	1296	1310	0.4824 Antigenic	Nonallergen	Non-Toxin	22.71%
ASTVCKLQS	VLSASTVCKLQSCTH	1298	1312	0.5485 Antigenic	Nonallergen	Non-Toxin	10.54%
AFSAMPKTS	KAFSAMPKTSLCFYI	1508	1522	0.7394 Antigenic	Nonallergen	Non-Toxin	74.36%
FSAMPKTSL	FKAFSAMPKTSLCFY	1507	1521	0.6103 Antigenic	Nonallergen	Non-Toxin	68.15%
MPKTSLCFY	FSAMPKTSLCFYIVE	1510	1524	0.4436 Antigenic	Nonallergen	Non-Toxin	43.67%

Note: *Vaxigen antigenicity threshold was 0.4.

Table 5: The physical and chemical properties, antigenicity and the number of the predicted transmembrane helices of the vaccine protein.

Vaccine Sequence*	Molecular weight (D)	Instability index#	Aliphatic index	Theoretical PI	No amino acids	Extinction coefficient	GRAVY†	Vaxijen antigenicity $	TMHs
KFEETFEVTAAAPVAVAAAGAAPAGA AVEAAEEQSEFDVILEAAGDKKIGVIK VVREIVSGLGLKEAKDLVDGAPKPLLE KVAKEAADEAKAKLEAAGATVTVK EAAAKRLGSELGCYYAATECLCPYEA YAACLCPYEALVYAATSLETSLSIYAA TSLSIEAPWYAAESKLLTVSVYAATEA IVCVELYAAIEAGTRFNLYAAFSAMPK TSLYAAPVQSAPGPGPGEWPKATGPG PGIQQKLPGPGPGYTGVTQHGPGPGP QLDLKARPTYGPGPGLLEWCKRNLGP GPGLEWCKRNLGGPGPGLGCYTINRV GPGPGYLNLERIPWGPGPGNLERIPW VVGPGPGLERIPWVVRGPGPGTSLET SLSIGPGPGSTSLETSLSGPGPGTSLSI EAPWGPGPGWPHSRNWRCGPGPGW CWGVGTGCGPGPGWGVGTGCTCGP GPGTEAIVCVELGPGPGVELTSQERQ GPGPGIEAGTRFNLGPGPGEAGTRFN LGGPGPGKVLSASTVCGPGPGASTVC KLQSGPGPGAFSAMPKTSGPGPGFSA MPKTSLGPGPGMPKTSLCFYHHHHHH	60787.23	28.61	72.57	5.67	599	96830	-0.110	0.4577	0

*Note: In the vaccine sequence, the linkers were shown in bold #instability index <40 considered the protein stable †GRAVY negative sign indicated the protein is hydrophilic $the threshold for the Vaxijen antigenicity is 0.4 THMs: Transmembrane helices.

Homology and transmembrane topology assessment

The high similarity between the sequence of the vaccine and the host proteome sequence could guide to autoimmune diseases due to molecular mimicry and the chances of cross reactivity [37,59,60]. The result demonstrated that no homology between the vaccine protein and the human proteins using BLASTp tool providing the vaccine as an excellent candidate with no autoimmunity. Moreover, the vaccine demonstrated no transmembrane helixes that might hindered the expression of the vaccine.

Molecular secondary structure prediction

The detailed prediction results for the 3-state secondary structure (SS3), solvent accessibility (ACC), and disordered regions (DISO). The SS3 provided 17%, 14% and 68% of the residues as alpha helix, beta-sheets and coil residues, respectively. The ACC provided 56% exposed residues, 21% medium residues and 21% were buried residues. A total of 76 (12%) of the residues were predicted as disordered regions (DISO) (Figure 3).

Vaccine tertiary structure prediction, refinement and validation

As shown in Figure 4, the 3D structure of the putative vaccine was predicted by I-TASSIER server. The model was refined by Galaxyrefiner. The overall structure of the vaccine was showed a Z-score of -2.97 in ProSA web program. This prediction demonstrated no errors in the structure of the vaccine. Ramachandran plot after refinement demonstrated that 83.9% of the residues were in the favoured region, 11.6% in the allowed region and 4.5% in the outlier region.

Prediction of the discontinuous B-cell epitopes

B-cell with 6 discontinuous epitopes using ElliPro server with scores from 0.574 to 0.806. Total of 318 residues were shown positioned as B cell discontinuous epitopes. The conformation epitopes size was from 6 to 137 residues (Table 6) (Figure 5).

Solubility of the putative vaccine

Solubility of the putative vaccine with solubility QuerySol scaled value of 0.461. This value was greater than that of the protein experimental dataset of *E. coli* (PopAvrSol) of 0.45. This result showed that the putative chimeric vaccine is potentially soluble in nature (Figure 6).

Molecular docking

The putative vaccine PDB file was considered as a ligand for the docking process with TLR4 (PDB: 4G8A) as receptors. Figure 7 demonstrated the docking process with negatively binding energies between the ligand and the receptors. The negative values indicating the low entropy between the reactants and at the same time the strong binding. For instance, the binding energy score of the putative vaccine with TLR4 was -1135.5 Kcal/mol and -1301.4 Kcal/mol for chain A and chain B respectively. This strong binding strengthens the potentiality of the putative vaccine to provoke the immune system of the host.

In silico cloning

The improved DNA sequence of the chimeric vaccine provided codon adaptation value (CAI-Value) of 1.0. This result is acceptable for good expression of the sequence in pET28a (+) vector sequence within *E. coli* strain K12. The GC-content of the improved sequence was 57.929 where the optimum rang was 30% to 70% indicating favorable GC content. Figure 8 demonstrating the cloning process of the DNA sequence into pET28a (+) vector after addition of BamHl and Xho1 restriction enzymes cutting sites to the extremities of the sequence.

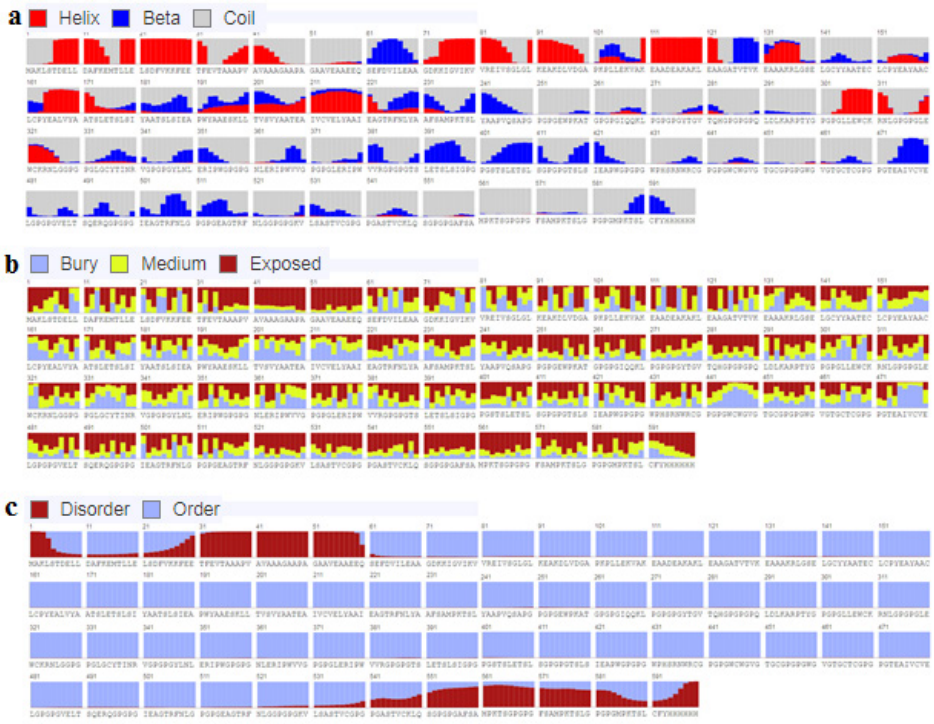

Figure 3: The secondary structure was predicted using Raptor X server. (a) Showed the 3-class secondary structure composition (b) showed solvent accessibility prediction and (c) shoed disordered residues prediction.

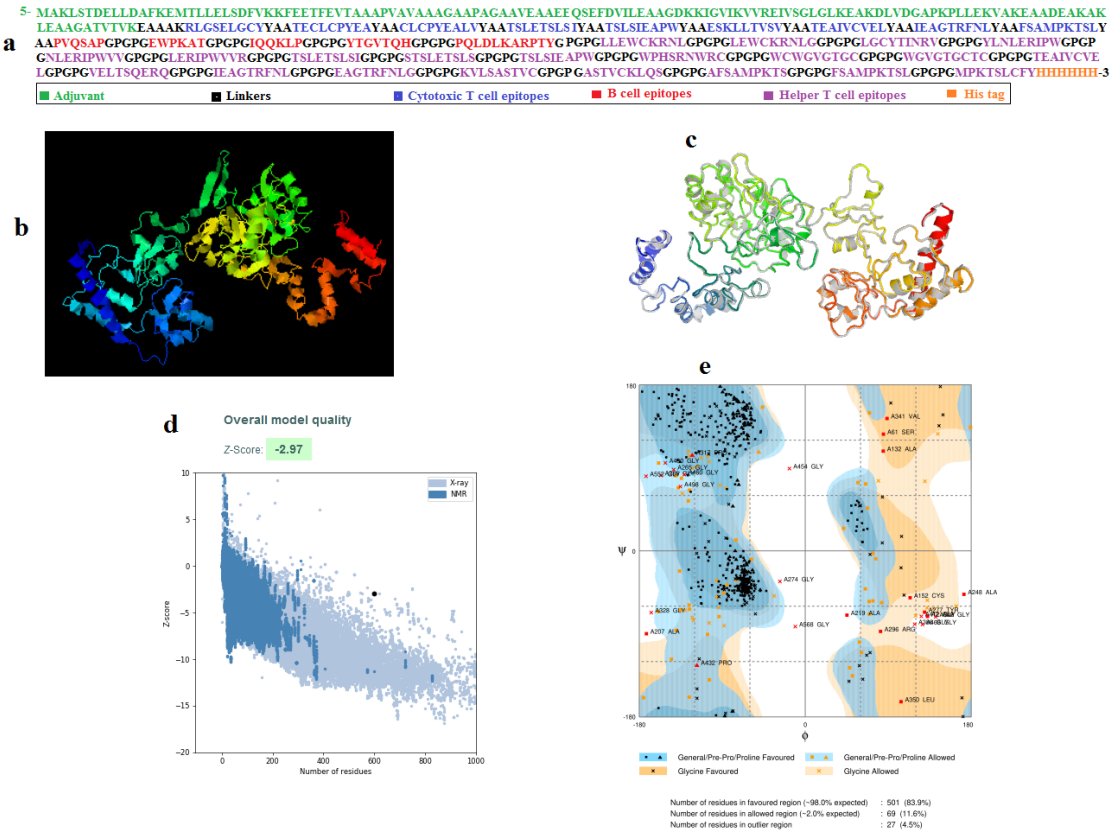

Figure 4: (a) showed the sequence of the putative vaccine. The sequence comprises the B and T cells epitopes, adjuvants, linkers and the His tag. (b) Vaccine tertiary structure as obtained by I-TASSER server. (c) The tertiary structure was refined in Galaxyrefiner server. (d) ProSA server Z score was -2.97 (e) finally the structure was validated by Ramachandran plot analysis.

Table 6: The showed the number of the predicted discontinuous B cell epitopes with the number of the residues and their scores.

No	Residue	Number of residues	Score
1	A:G510, A:P511, A:G512, A:P513, A:G514, A:E515, A:T518, A:N521, A:L522, A:G523, A:G524, A:P525, A:G526, A:P527, A:G528, A:K529, A:V530, A:L531, A:S532, A:A533, A:S534, A:T535, A:V536, A:C537, A:G538, A:P539, A:G540, A:P541, A:G542, A:A543, A:S544, A:T545, A:V546, A:C547, A:K548, A:L549, A:Q550, A:S551, A:G552, A:P553, A:G554, A:P555, A:G556, A:A557, A:F558, A:S559, A:A560, A:M561, A:P562, A:K563, A:T564, A:S565, A:G566, A:P567, A:G568, A:P569, A:G570, A:F571, A:S572, A:A573, A:M574, A:P575, A:K576, A:T577, A:S578, A:L579, A:G580, A:P581, A:G582, A:P583, A:G584, A:M585, A:P586, A:K587, A:T588, A:S589, A:L590, A:C591, A:F592, A:Y593, A:H594, A:H595, A:H596, A:H597, A:H598, A:H599	86	0.806
2	A:M1, A:A2, A:K3, A:L4, A:S5, A:T6, A:D7, A:E8, A:L9, A:L10, A:D11, A:A12, A:F13, A:K14, A:E15, A:M16, A:T17, A:L18, A:L19, A:E20, A:L21, A:S22, A:D23, A:F24, A:V25, A:K26, A:K27, A:F28, A:E29, A:E30, A:T31, A:F32, A:E33, A:V34, A:T35, A:A36, A:A37, A:A38, A:P39, A:V40, A:A41, A:V42, A:A43, A:A44, A:A45, A:G46, A:E55, A:A56, A:A57, A:E58, A:E59, A:Q60, A:S61, A:E62, A:F63, A:D64, A:V65, A:I66, A:E68, A:G76, A:V77, A:I78, A:K79, A:V80, A:V81, A:R82, A:E83, A:I84, A:V85, A:S86, A:G87, A:L88, A:G89, A:L90, A:K94, A:P101, A:P103, A:L104, A:L105, A:E106, A:K107, A:V108, A:A109, A:K110, A:E111, A:A112, A:A113, A:D114, A:E115, A:A116, A:K117, A:A118, A:K119, A:L120, A:E121, A:A122, A:A123, A:G124, A:A125, A:T126, A:V127, A:T128, A:V129, A:K130, A:E131, A:A132, A:A133, A:A134, A:K135, A:R136, A:L137, A:G138, A:S139, A:E140, A:L141, A:C143, A:Y145, A:T148, A:E149, A:C150, A:L151, A:C152, A:P153, A:Y154, A:E155, A:A156, A:Y157, A:A158, A:A159, A:C160, A:L161, A:C162, A:P163, A:Y164, A:E165, A:V168, A:Y169	137	0.708
3	A:A242, A:A243, A:P244, A:V245, A:Q246, A:S247, A:A248, A:P249, A:G250, A:P251, A:G252, A:P253, A:G254, A:P257, A:K258, A:A259, A:T260, A:G261, A:P262, A:H283, A:G284	21	0.684
4	A:G274, A:Y277, A:T278, A:T281, A:Q282, A:L293, A:K294, A:A295, A:R296, A:P297, A:T298, A:Y299, A:G300, A:P301, A:G302, A:P303, A:G304, A:L305, A:L306, A:E307, A:W308, A:C309, A:K310, A:R311, A:N312, A:L313, A:G314, A:P315, A:G328, A:P329, A:G330, A:P331, A:G332, A:L333, A:G334, A:C335, A:Y336, A:T337, A:I338, A:N339, A:R340, A:V341, A:G342, A:P343, A:G344, A:I365, A:P366, A:W367, A:V368, A:V369	50	0.63
5	A:S185, A:A206, A:A207, A:T208, A:E209, A:A210, A:I211, A:Y217, A:A218, A:A219, A:I220, A:E221, A:A222, A:G223, A:T224, A:N227, A:L228, A:Y229	18	0.592
6	A:P354, A:G356, A:P357, A:G360, A:N361, A:R364	6	0.574

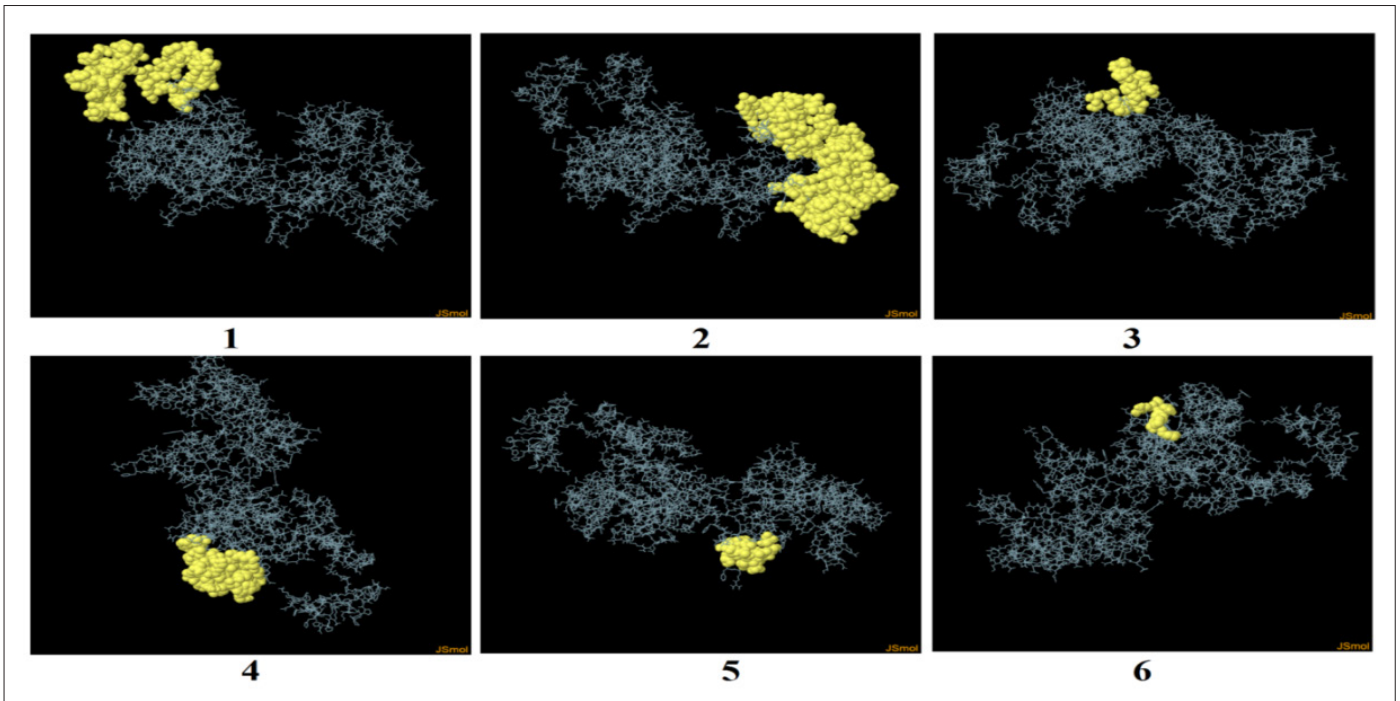

Figure 5: The 3D structures of the 6 discontinuous B-cell epitopes predicted by the ElliPro (1-6). Epitopes were shown in yellow color, while grey color showed the chimeric protein.

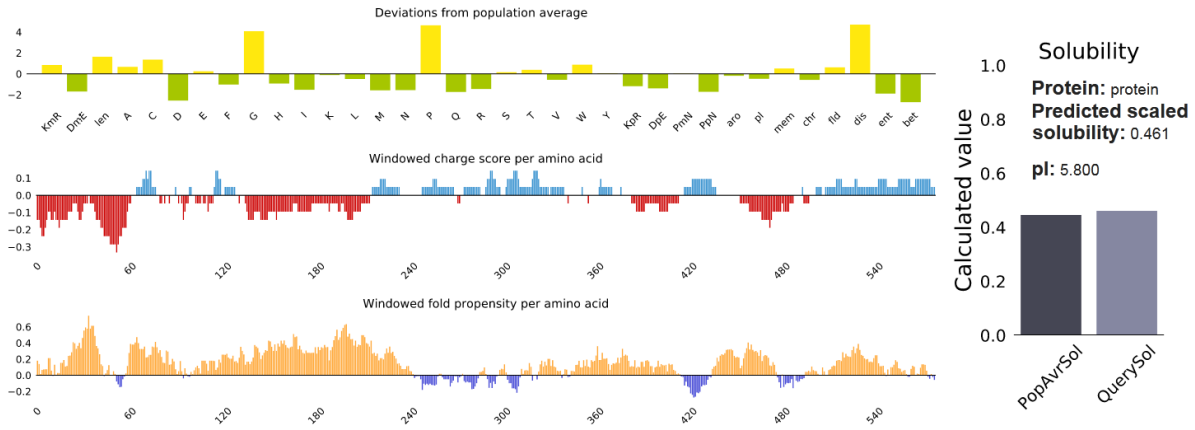

Figure 6: Vaccine construct solubility as obtained by protein sol server. The solubility of the vaccine construct was shown to be 0.461 compared to 0.45 of the population average solubility of *E. coli*.

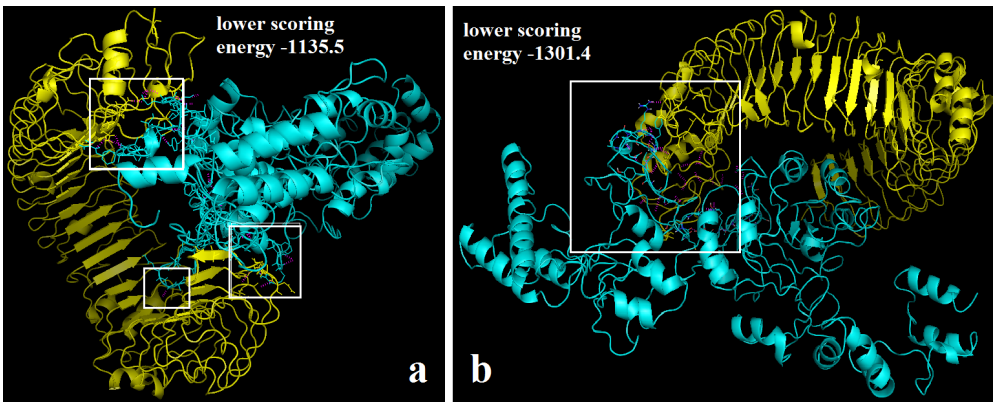

Figure 7: (a) Molecular docking of the chimeric vaccine with TLR4 chain A while (b) with TLR4 chain B. The vaccine was shown in cyan color when docked with TLR4 (yellow color). The interaction regions were shown in magenta color in the boxes.

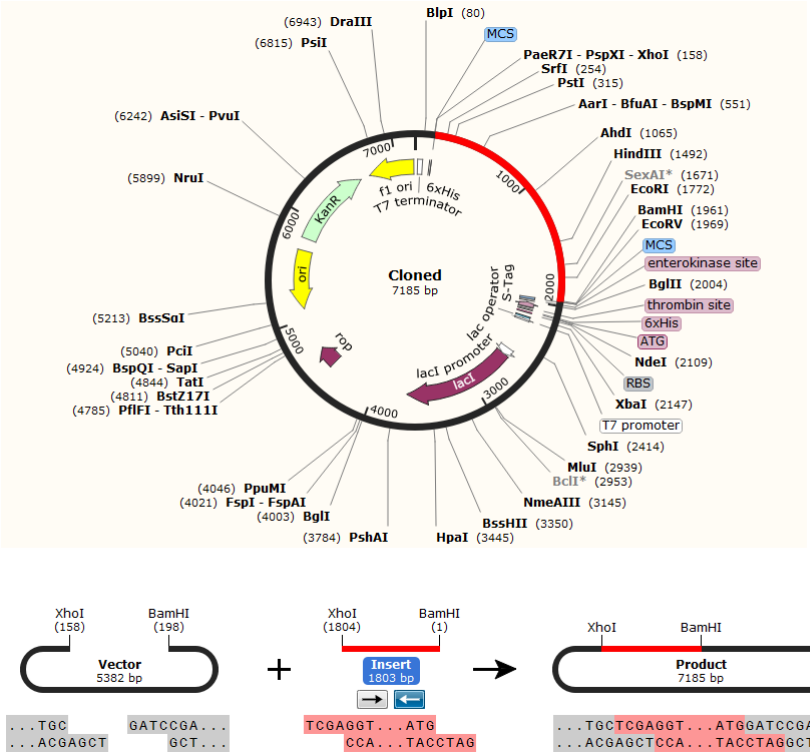

Figure 8: In silico cloning of the DNA sequence (red color) into the pET30a (+) expression vector (black color). The enzymes used in the cloning process and the length of the insert were also shown.

DISCUSSION

For the normal this study aimed to predict multi-epitopes vaccine candidate against CCHFV. The conventional vaccines development demonstrated multiple drawbacks as multiple organisms difficult to grow and culture and some demonstrates problems in the process of vaccine attenuation. Moreover, some organisms may revert to more virulent strains causing adverse events and threatens the human life [61]. Therefore, the shift to a new era of effective, safe, less cost vaccine became of great importance such as *in silico* vaccine [62,63]. Recently many efforts were paid to develop vaccine against CCHFV but these attempts were hampered by the lack of a challenge model and efficacy could not be demonstrated [64] or antibodies were not tested for neutralization titers or the vaccine not tested for protection [65]. Currently there is no available an effective vaccine to combat the action of the CCHFV. Glycoprotein M of CCHFV is a paramount protein in viral assembly and localization; multiple reports have targeted this glycoprotein as a potent immunogen for vaccine construction [66-70]. Thus, glycoprotein M was used to predict epitopes that work as vaccine candidate against CCHFV eliciting B and T lymphocytes.

For epitopes prediction, the B and T cell epitopes were predicted from glycoprotein M using various tools in the IEDB. In this study 5 B cell epitopes, 9 cytotoxic T cell and 21 helper T cell epitopes were predicted as a vaccine candidate and were shown to be antigenic, eliciting immune response, not causing allergenicity and nontoxic epitopes. The epitope 1231-TEAIVCVEL-1239 was previously shown be as a good binder to MHC-1 alleles [71]. In our study this epitope demonstrated high binding affinity to MHC-1 alleles as well as to MHC-11 alleles, got the higher antigenicity score by vaxijen server and shown to be non-allergic and nontoxic. Therefore, we proposed this epitope as vaccine candidates against T cell. Furthermore, the epitopes 1060-TSLSIEAPW-1068, 1056-TSLETSLSI-1064, 1249-IEAGTRFNL-1257 and 1510-FSAMPKTSL-1518 and bound to both MHC-1 and MHC-II alleles, antigenic, nonallergic and nontoxic epitopes. All the proposed epitopes were investigated for the population coverage against the whole world that had the potential to develop immune response against these epitopes.

To develop a global vaccine against CCHFV, sequences of glycoprotein M of 43 strains of CCHFV from different geographical areas were retrieved from the NCBI. The evolutionary relationship between the retrieved strains was studied since it assisted in demonstrating the molecular distribution of the virus and circumstances in vaccine design in the different geographical areas. Moreover, a conservation analysis of the predicted epitopes from the glycoprotein M of the retrieved strains was performed. Epitopes sequences with 100% conservancy were used in the structure of the vaccine. Moreover, the population coverage of each epitope was calculated from the allelic interaction of each epitope against the whole world.

To construct the vaccine, short peptides were used as linkers between the B and T cells epitopes [72]. Theses linkers were shown to provoke minimal junctional immunogenicity between the epitopes as previously described [57,58,72-75] and to reach a high level of expression and improve bioactivity of the vaccine [57,72]. An adjuvant was added in the N terminal of the vaccine as an immunomodulator and to enhance the activity of multiple vaccines [76,77].

Most importantly for the vaccine to be considered as better vaccine it should possess antigenic properties, which are important to elicit the immune response of the host. In addition to that it must not present homology with human proteins to escape the potentiality of causing autoimmune response [37,59,60]. Moreover, the predicted vaccine must not contain transmembrance helix regions, to ease their expression [38-40]. In this study the designed vaccine demonstrated antigenicity in Vaxijen server (scored 0.4577 passing 0.4 the threshold of the server), showed no homology to human proteins, with no transmembrane helix predicted. These results guarantee that the vaccine could elicit the immune system without provoking autoimmune diseases and ease of the vaccine expression. Also, bioinformatics and immunologic analysis of vaccine structure should contain MHCI, MHCII epitopes accompanied by linear and discontinuous B-cell epitopes as previously described [74]. In addition to that the physio-chemical properties of the vaccine were analyzed demonstrating that the vaccine protein was stable, contains aliphatic side chains, hydrophilic and with thermal stability. These features provided the suitability of vaccine construct as good vaccine against CCHFV.

The study of the quality of proteins secondary and tertiary structures are of crucial importance for efficient presentation of antigenic peptides on MHC for triggering strong immune reactions [78]. In this study the determination of the quality and potential errors in the structural model of the vaccine protein were analyzed. Secondary structure showed the vaccine with helixes, beta strands and loops. For solvent accessibility the majority of the residues were exposed to solvents. This further provided the high solubility and polarity of the vaccine model structure. Tertiary structure was predicted and refined and then the model was assessed in ProSA web server. The model Z-score was -2.97, which falls within those commonly observed in similar size-native proteins [79]. Furthermore, the model was assessed in Ramachandran plot analysis. The result showed that the majority of the residues were in most-favored region and very few residues were in disallowed regions. This result remarkably demonstrated that the excellent quality and stability of the final refined model based on Ramachandran plot predictions.

One important feature of the designed vaccine is to analyze the solubility of the vaccine since the vaccine will be administered in water milieu in the host body. Silva, et al. [80] previously showed that protein vaccines with low solubility demonstrated disadvantages in production of large amounts of virus proteins. Thus, the solubility of the vaccine was assessed compared to the solubility of *E. coli* proteins using protein sol server [51]. The solubility of the vaccine scored 0.461 more than that of the *E. coli* proteins (0.45) providing the good solubility of the vaccine protein in this study.

To strengthen the interaction between the vaccine protein and the host toll-like receptor (TLR4) molecular docking was performed to determine the favorable protein-protein interaction. The complex docked process of the vaccine protein with the receptor provided lower binding energy scores (negative values) conferring the highest binding between the molecules. Also, molecular cloning of the vaccine protein was performed for immunoreactivity [81]. *E. coli* expression systems were shown as the most preferable choice for cloning [82,83]. To ensure the complete expressing of the designed vaccine protein, codon usage optimization was performed. The CAI was 1.0, showing the high proportional level of the abundant codons, while the GC-content was 57.929 where the optimum rang was 30% to 70% indicating expression of the protein with high level in bacteria.

CONCLUSION

The synthesis or construction of a vaccine or antiviral drug against CCHFV is very crucial. The most important feature of an epitope-based vaccine is their high conservancy among all retrieved strains sequences. Another important criterion of the peptide vaccine is not to be associated with adverse effects. The proposed epitopes in this study were examined *in silico* and found to be non-allergenic and nontoxic. By identifying CCHF virus glycoproteins epitopes that successfully elicited B and T cells we concluded that our proposed epitopes as vaccine candidates were promising candidates. Further studies to propose a peptide vaccine for other strains of Crimean Congo hemorrhagic fever virus is recommended. Moreover, these proposed epitopes require applying *in vitro* and *in vivo* methods to evaluate their efficacy as efficient vaccine.

ACKNOWLEDGMENT

Authors would like to thank the staff members of University of Bahri, Sudan, College of Veterinary Medicine, Department of Molecular Biology and Bioinformatics, for their valuable support during the study.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

REFERENCES

1. Flick R, Whitehouse CA. Crimean-Congo hemorrhagic fever virus. Curr Mol Med. 2005; 5 (8): 753-760.
2. Bergeron É, Albariño CG, Khristova ML, Nichol ST. Crimean-Congo hemorrhagic fever virus-encoded ovarian tumor 6 (OTU) protease activity is dispensable for virus RNA polymerase 7 function. J Virol. 2010; 84 (1): 216-226.
3. Alam MM, Khurshid A, Sharif S, Shaukat S, Rana MS, Angezet A, et al. Genetic analysis and epidemiology of crimean congo hemorrhagic fever viruses in Baluchistan Province of Pakistan. BMC Infect Dis. 2013; 13: 201.
4. Sherifi K, Cadar D, Muji S, Robaj A, Ahmeti S, Jakupi X, et al. Crimean-Congo hemorrhagic fever virus clades v and vi (Europe 1 and 2) in ticks in Kosovo. PLoS Negl Trop Dis. 2012; 8 (9): e3168.
5. Mild M, Simon M, Albert J, Mirazimi A. Towards an understanding of the migration of Crimean-Congo hemorrhagic fever virus. J Gen Virol. 2010; 91 (1): 199-207.
6. Estrada-Peña A, Ruiz-Fons F, Acevedo P, Gortazar C, Fuente J la. 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.
7. Hoogstraal H. The epidemiology of tick-borne Crimean–Congo hemorrhagic fever in Asia, Europe, and Africa. J Med Entomol. 1979; 15 (4): 307–417.
8. Kraus AA, Mirazimi A. Molecular biology and pathogenesis of Crimean–Congo hemorrhagic fever virus. Future Virol. 2010; 5 (4): 469–479.
9. Holland J, Domingo E. Origin and evolution of viruses. Virus Genes. 1998; 16 (1): 13–21.
10. Sette A, Newman M, Livingston B, McKinney D, Sidney J, Ishioka G, et al. Optimizing vaccine design for cellular processing, MHC binding and TCR recognition. Tissue Antigens. 2002; 59 (6): 443–451.
11. Sette A, Fikes J. Epitope-based vaccines: An update on epitope identification, vaccine design and delivery. Curr Opin Immunol. 2003; 15 (4): 461–470.
12. Poland GA, Ovsyannikova IG, Jacobson RM. Application of pharmacogenomics to vaccines. Pharmacogenomics. 2009; 10 (5): 837–852.
13. Petrovsky N, Brusic V. Computational immunology: The coming of age. Immunol Cell Biol. 2002; 80 (3): 248–254.
14. Bourdette DN, Edmonds E, Smith C. A highly immunogenic trivalent T cell receptor peptide vaccine for multiple sclerosis. Mult Scler. 2005; 11 (5): 552–561.
15. López JA, Weilenman C, Audran R. 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.
16. Wilson CC, McKinney D, Anders M, Ma Whinney S. 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.
17. Emhoff IA, Fugate E, Eyal N. Is there a moral right to nonmedical vaccine exemption? Am J Law Med. 2016; 42: 598-620.
18. Lo YT, Pai TW, Wu WK, Chang HT. Prediction of conformational epitopes with the use of a knowledge-based energy function and geometrically related neighboring residue characteristics. BMC Bioinformatics. 2013; 14: S3.
19. Shayan S, Bokaean M, Shahrivar MR, Chinikar S. Crimean-Congo hemorrhagic fever. Lab Med. 2015; 46 (3): 180-9.
20. Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016; 33 (7): 1870-1874.
21. Hall TA. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT in Nucleic acids symposium series. [London]: Information Retrieval Ltd. Am J Micro Biol Res. 1999; c1979-c2000.
22. Zheng J, Lin X, Wang X, Zheng L. *In Silico* analysis of epitope-based vaccine candidates against hepatitis B virus polymerase protein. Viruses. 2017; 9 (5): 112.
23. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: Improvingsequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res. 2017; 45 (W1): W24–W29.
24. Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J Virol. 1985; 55 (3): 836-839.
25. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett. 1990; 276 (1-2): 172-174.
26. Kim Y, Ponomarenko J, Zhu Z, Tamang D, Wang P, Greenbaumet J, et al. Immune epitope database analysis resource. Nucleic Acids Res. 2012; 40: W525-30.
27. Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lamberth K. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci. 2003; 12 (5): 1007-1017.
28. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: Accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. Nucleic Acids Res. 2008; 36: W509-W12.
29. Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. BMC Bioinformatics. 2005; 6:132.

30. Sidney J, Assarsson E, Moore C, Ngo S, Pinilla C, Sette A, et al. Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. Immunome Res. 2008; 4: 2.

31. Hawman DW, Feldmann H. Recent advances in understanding Crimean–Congo hemorrhagic fever virus. 2018; 7: F1000 Faculty Rev:1715.

32. Aslam S, Latif MS, Daud M, Tabassum B, Riaz MS, Khan A, et al. Crimean-Congo hemorrhagic fever: Risk factors and control measures for the infection abatement. Biomed Rep. 2016; 4: 15-20.

33. Wang P, Sidney J, Dow C, Mothe B, Sette A. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comput Biol. 2008; 4 (4): e1000048.

34. Dimitrov I, Bangov I, Flower DR, Doytchinova IA. AllerTOP v.2- a server for *in silico* prediction of allergens. J Mol Model. 2013; 20 (6): 2278.

35. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GP. *In silico* approach for predicting toxicity of peptides and proteins. PLoS One. 2013; 8 (9): e73957.

36. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein Identification and Analysis Tools on the ExPASy Server: (In) John M. Walker (ed): The Proteomics Protocols Handbook. Humana Press. 2005; pp.571-607.

37. Rojas M, Restrepo-Jimenez P, Monsalve DM, Pacheco Y, Acosta-Ampudia Y, Ramírez-Santana C, et al. Molecular mimicry and autoimmunity. J Autoimmun. 2018; 95: 100–123.

38. Monterrubio G, Gonzalez-Y-Merchand JA, Ribas-Aparicio RM. Identification of novel potential vaccine candidates against tuberculosis based on reverse vaccinology. Biomed Res Int. 2015; (12): 1–16.

39. Naz A, Awan FM, Obaid A. Identification of putative vaccine candidates against Helicobacter pylori exploiting exoproteome and secretome: A reverse vaccinology-based approach. Infect Genet Evol. 2015; 32: 280–291.

40. Nezafat N, Eslami M, Negahdaripour M. Designing an efficient multi-epitope oral vaccine against Helicobacter pylori using immunoinformatics and structural vaccinology approaches. Mol Biosyst. 2017; 13 (4): 699–713.

41. Wang S, Peng J, Ma J, Xu J. Protein secondary structure prediction using deep convolutional neural fields. Sci Rep. 2016; 6: 18962–18973.

42. Schaarschmidt J, Monastyrskyy B, Kryshtafovych A, Bonvin AMJJ. Assessment of contact predictions in CASP12: Co-evolution and deep learning coming of age. Proteins. 2018; 86: 51–66.

43. Yang Y, Gao J, Wang J, Heffernan R, Hanson J, Paliwal K, et al. Sixty-five years of the long march in protein secondary structure prediction: the final stretch? Brief Bioinform. 2018; 19 (3): 482–494.

44. Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophys J. 2011; 101 (10): 2525–2534.

45. Heo L, Park H, Seok C. GalaxyRefine: Protein structure refinement driven by side-chain repacking. Nucleic Acids Res. 2013; 41: W384–W388.

46. Shin WH, Lee GR, Heo L, Lee H, Seok C. Prediction of protein structure and interaction by GALAXY protein modeling programs. Bio Design. 2014; 2 (1): 1-11.

47. Ko J, Park H, Heo L, Seok C. Galaxy WEB server for protein structure prediction and refinement, Nucleic Acids Res. 2012; 40 (W1): W294-W297.

48. Lovell SC, Davis IW, Arendall WB, Bakker PIW, Word JM, Prisant MG, et al. Structure validation by Calpha geometry: phi, psi and Cbeta deviation, Proteins. 2003; 50 (3): 437–450.

49. Al-Hakim M, Hasan R, Ali MF, Rabbee MF, Marufatuzzahan. *In silico* characterization and homology modeling of catechol 1,2 dioxygenase involved in processing of catechol- an intermediate of aromatic compound degradation pathway. Glob J Sci Front Res G Bio-Tech Genet. 2015; 15: 1–13.

50. Ponomarenko JV, Bui H, Li W, Fusseder N, Bourne PE, Sette A, et al. ElliPro: A new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics. 2008; 9: 514.

51. Hebditch M, Carballo-Amador MA, Charonis S, Curtis R, Warwicker J. Protein-Sol: A web tool for predicting protein solubility from sequence. Bioinformatics. 2021; 33 (19): 3098-3100.

52. Vajda S, Yueh C, Beglov D, Bohnuud T, Mottarella SE, Xia B, et al. New additions to the ClusPro server motivated by CAPRI. Proteins. 2017; 85 (3): 435-444.

53. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusPro web server for protein-protein docking. Nat Protoc. 2017; 12 (2): 255-278.

54. Grote A, Hiller K, Scheer M, Munch R, Nortemann B, Hempel DC, et al. JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res. 2005; 33: W526-31.

55. Sharp PM, Li WH. The codon adaptation index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 1987; 15 (3): 1281-95.

56. Morla S, Makhija A, Kumar S. Synonymous codon usage pattern in glycoprotein gene of rabies virus. Gene. 2016; 584 (1): 1–6.

57. Shey RA, Ghogomu SM, Esoh KK, Nebangwa ND, Shintouo CM, Nongley NF, et al. *In silico* design of a multi-epitope vaccine candidate against onchocerciasis and related fiearial diseases. Sci Rep. 2019; 9 (1): 4409.

58. Pandey RK, Ojha R, Aathmanathan VS, Krishnan M, Prajapati VK. Immunoinformatics approaches to design a novel multi-epitope subunit vaccine against HIV infection. Vaccine. 2018; 36 (17): 2262–2272.

59. Kanduc D. Peptide cross-reactivity: The original sin of vaccines. Front Biosci. 2012; 4: 1393–1401.

60. Ojha R, Pareek A, Pandey RK, Prusty D, Prajapati VK. Strategic development of a next-generation multi-epitope vaccine to prevent Nipah virus zoonotic infection. ACS Omega. 2019; 4 (8): 13069-13079.

61. Purcell AW, Mc Cluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. Nat Rev Drug Discov. 2007; 6 (5): 404–414.

62. Mora M, Veggi D, Santini L, Pizza M, Rappuoli R. Reverse vaccinology. Drug Discov. 2003; 8 (10): 459–464.

63. Rappuoli R, Bottomley MJ, D'Oro U, Finco O, De Gregorio E. Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design. J Exp Med. 2016; 213 (4): 469–481.

64. Spik K, Shurtleff A, McElroy AK, Guttieri MC, Hooper JW, SchmalJohn C. Immunogenicity of combination DNA vaccines for Rift Valley fever virus, tickborne encephalitis virus, Hantaan virus, and Crimean Congo hemorrhagic fever virus. Vaccine. 2006; 24 (21): 4657–4666.

65. Ghiasi SM, Salmanian AH, Chinikar S, Zakeri S. Mice orally immunized with a transgenic plant expressing the glycoprotein of Crimean-Congo hemorrhagic fever virus. Clin Vaccine Immunol. 2011; 18 (12): 2031–2037.

66. Saijo M, Morikawa S, Kurane I. Recent progress in the treatment of Crimean–Congo hemorrhagic fever and future perspectives. Fut Virol. 2010; 5 (6): 801–809.

67. Strandin T, Hepojoki J, Vaheri A. Cytoplasmic tails of bunyavirus Gn glycoproteins—could they act as matrix protein surrogates? Virology. 2013; 437 (2): 73–80.

68. Buttigieg KR, Dowall SD, Findlay-Wilson S, Miloszewska A, Rayner E, Hewson R, 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.

69. Dowall SD, Carroll MW, Hewson R. Development of vaccines against Crimean-Congo haemorrhagic fever virus. Vaccine. 2017; 35 (44): 6015–6023.

70. Wu Y, Zhu Y, Gao F, Jiao Y, Oladejo BO, Chai Y, et al. Structures of phlebovirus glycoprotein Gn and identifcation of a neutralizing antibody epitope. Proc Natl Acad Sci USA. 2017; 114 (36): E7564–E7573.

71. Tipu HN. Immunoinformatic analysis of Crimean Congo hemorrhagic fever virus glycoproteins and epitope prediction for synthetic peptide vaccine. J Coll Phys Surg Pak. 2016; 26 (2): 108-112.

72. Meza B, Ascencio F, Sierra-Beltrán AP, Torres J, Angulo C. A novel design of a multi-antigenic, multistage and multi-epitope vaccine against Helicobacter pylori: An *in silico* approach. Infect Genet Evol. 2017; 49: 309–317.

73. Hasan M, Ghosh PP, Azim KF, Mukta S, Abir RA, Nahar J, et al. Reverse vaccinology approach to design a novel multi-epitope subunit vaccine against avian influenza A (H7N9) virus. Microb Pathog. 2019; 130: 19-37.

74. Ali M, Pandey RK, Khatoon N. Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. Sci Rep. 2017; 9232.

75. Khatoon N, Pandey R K, Prajapati VK. Exploring Leishmania secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach. Sci Rep. 2017; 7: 82-85.

76. Mohan T, Verma P, Rao DN. Novel adjuvants and delivery vehicles for vaccines development: A road ahead, Indian J Med Res. 2013; 138 (5): 779-95.

77. Solanki V, Tiwari V. Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against Acinetobacter baumannii. Sci Rep. 2018; 8 (1): 9044.

78. Scheiblhofer S, Laimer J, Machado Y, Weiss R, Thalhamer J. Influence of protein fold stability on immunogenicity and its implications for vaccine design. Expert Rev Vaccines. 2017; 16 (5): 479–89.

79. Urrutia-Baca VH, Gomez-Flores R, De La Garza-Ramos MA, Tamez-Guerra P, Lucio-Sauceda DG, Rodríguez-Padilla MC. Immunoinformatics approach to design a novel epitope-based oral vaccine against helicobacter pylori. J Comput Biol. 2019; 26 (10): 1177-1190.

80. Silva HC Junior, Pestana CP, Galler R, Medeiros MA. Solubility as a limiting factor for expression of hepatitis A virus proteins in insect cell-baculovirus system. Mem Inst Oswaldo Cruz. 2016; 111 (8): 535-538.

81. Gori A, Longhi R, Peri C, Colombo G. Peptides for immunological purposes: Design, strategies and applications. Amino Acids. 2013; 45 (2): 257–268.

82. Chen R. Bacterial expression systems for recombinant protein production: *E. coli* and beyond. Biotechnol Adv. 2012; 30 (5): 1102–7.

83. Rosano GL, Ceccarelli EA. Recombinant protein expression in Escherichia coli: Advances and challenges. Front Microbiol. 2014; 5: 172.