International Immunopharmacology 78 (2020) 106020

Contents lists available at ScienceDirect

International Immunopharmacology

journal homepage: www.elsevier.com/locate/intimp

Introducing of an integrated artificial neural network and Chou's pseudo amino acid composition approach for computational epitope-mapping of Crimean-Congo haemorrhagic fever virus antigens

Mokhtar Nosrati, Hassan Mohabatkar, Mandana Behbahani*

Department of Biotechnology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan, Iran

ARTICLE INFO

Keywords:
Crimean-Congo haemorrhagic fever virus
Chou's pseudo amino acid composition
Epitope
Machine learning

ABSTRACT

This study was aimed to introduce a novel algorithm for determining linear B- and T-cell epitopes from Crimean-Congo haemorrhagic fever virus (*CCHFV*) antigens. To this end, 387 approved B- and T-cell epitopes, as well as 331 non-epitope peptides from different serotypes of the virus were collected from IEDB database for generating of the train datasets. After that, the physicochemical properties of the epitopes were expressed as the numeric vectors using Chou's pseudo amino acid composition method. The vectors then were used for training of four machine learning algorithms including artificial neural network (ANN), k-nearest neighbors (kNN), support vector machine (SVM) and Random forest (RF). The results confirmed that ANN was the most accurate algorithm for discriminating between the epitopes and non-epitopes with the accuracy of 0.90. Furthermore, for evaluating the performance of the ANN algorithm, an epitope prediction challenge was performed to a random peptide library from envelopment polyprotein of *CCHFV*. Moreover, the efficiency of the predicted epitopes in term of antigenicity and affinity to MHC-II were compared to the predicted epitope by standard epitope prediction tools based on their VaxiJen 2.0 score and molecular docking outputs. Finally, the ability of the screened epitopes to stimulation of humoral and cellular responses was evaluated by an in silico immune simulation process thought C-Immsim 10.1 server. The results confirmed that this method has more accuracy for epitope-mapping than the standard tools and could considered as an effective algorithm to develop a serotype independent one-click automated epitope based vaccine design tool.

1. Introduction

Crimean-Congo hemorrhagic Fever (CCHF) is a virulent lethal human disease caused by CCHF virus (*CCHFV*), a member of the *Orthonairovirus* genus of the *Nairoviridae* family. The viral genome, include a three-segmented single strand negative sense RNA which encode RNA dependent polymerase (L segment), the envelope spike proteins (M segments) and nucleoprotein (N segment) respectively. The virus is transmitted to humans via tick bites or by exposure to the blood, body fluids or tissue of infected animals or humans [2,1].

Due to severity, high mortality rate, wide geographical distribution and therapeutic difficulties, CCHF is considered as one of the significant threat to public health. Furthermore, unfortunately, there is currently no internationally-accepted vaccine for CCHF. Therefore, the development of a new effective vaccine for CCHF is necessary [3,4].

First developed vaccine against CCHF was chloroform-inactivated CCHFV, which licensed in Bulgaria and has been used in endemic areas for military personnel since 1974. Despite the claims of the Bulgarian Ministry of Health about the efficacy of the vaccine, obtained results in 2012 showed that immune responses in individuals vaccinated with the Bulgarian vaccine were low and required three booster vaccinations to improve immunity [2,5]. After that, some researches have been aimed at studying the molecular mechanism of *CCHFV* infection as well as vaccine development against the disease [6–9].

Regarding vaccine development for *CCHFV*, Hinkula et al., investigated the efficacy of a DNA vaccine expressing the virus glycoproteins (Gc and Gn) as well as nucleoprotein. Their study demonstrated that the vaccine could provoke humoral and cellular immune responses against *CCHFV* in interferon-alpha receptor (IFNAR) knockout mouse model [10].

In another study, Spik and colleagues had demonstrated that a DNA vaccine encodes *CCHFV* glycoprotein precursor induced humoral immune responses in half of the vaccinated mice [11].

Similarly, Ghiasi et al. confirmed that transgenic tobacco leaves

* Corresponding author.
E-mail address: ma_behbahani@yahoo.com (M. Behbahani).

https://doi.org/10.1016/j.intimp.2019.106020
Received 19 September 2019; Received in revised form 9 October 2019; Accepted 31 October 2019
Available online 24 November 2019

expressing Gn and Gc could inducing specific IgG and IgA in tested mice [12].

Buttigieg and colleagues demonstrated that the modified vaccinia virus Ankara expressing *CCHFV* glycoproteins raised both humoral and cellular immune responses in IFNAR mice. This vaccine protected all vaccinated the animal model from *CCHFV* in a viral challenge [13].

Despite promising results in the previous studies, however, due to some common weak points including serotype-limited protection, the possibility of allergenicity and necessity of both cellular and humoral immune responses against vaccine candidate, there is no approved commercially available vaccine for CCHFV [1,14]. Therefore, development of effective platforms of none serotype-limited and safe *CCHFV* vaccines is necessary.

Epitope-based vaccines (EBVs) are considered as an alternative to conventional vaccine platforms and can address the possible common vaccine platform side effects such as serotype-limited protection, low immunogenicity and allergenicity [15,16].

Generally, an epitope also called antigenic determinant refers to the part of an antigen that interacts with the antibody or the antigen receptor. Experimental mapping of epitopes is a complicate, costly and tedious process. Therefore, in the recent years computational methods are highly regarded as alternative or complementary approaches for empirical determination of epitopes and EBVs design [17,18].

To date, many algorithms, software and online web server are introduced for B- and T-cell epitope prediction as well as in silico vaccine design [19,20]. Furthermore, recently many EBVs are proposed against infectious diseases including Hepatitis C virus (HCV) [21], human immunodeficiency viruses-1 (HIV-1) [22], Classical swine fever virus (CSFV) [23], influenza virus [24], Dengue virus [25], Helicobacter pylori [26], Enterohemorrhagic Escherichia coli [27], Staphylococcus aureus [28], Plasmodium falciparum [29], Toxoplasma gondii [30] and many more using in silico approaches.

Despite dramatic progression in increasing accuracy as well as the number of immunoinformatics tools for epitope prediction, the method is faced with severe challenges. As the first limitation, presence of many epitope-prediction tools with different algorithms causes researchers' confusion in choosing the best tool. Moreover, due to separate database and algorithms used in the tools, obtained results may not be compatible or repeatable. Furthermore, in the most of the available immunoinformatics tools, few physicochemical and structural properties of epitopes are considered for data classification and learning of used algorithms; while, more decisive characteristics as well as the simultaneous effect of the features should be regarded to increase precision and reliability of in silico epitope mapping and vaccine development [31–34].

To address the mentioned issues we used a unique pipeline for increase accuracy and repeatability of epitope determination from CCHFV proteins. For this purpose, experimental and insilico determined B- and T-cell epitopes with appropriate physicochemical properties were retrieved from particular databases and previously reported studies as positive datasets. Furthermore, a set of non-epitope peptides were consider as negative data. Afterwards, negative and positive peptides were expressed in to binary forms through Chou's pseudo amino acid composition method. Finally, different statistical classification models were trained using the datasets and best algorithm was selected for epitope prediction based on its accuracy and repeatability. Then the screened epitope(s) were compared with predicted epitopes from some known epitope prediction servers and top-scored peptides were selected as high immunogenic epitopes for different therapeutics goals especially EBV design against *CCHFV*.

2. Material and methods

2.1. Data collection and datasets preparation

As the first step, a set of experimentally determined B- and T-cell epitopes from nucleoprotein, envelope polyproteins and RNA dependent RNA polymerase of *CCHFV* were separately extracted from the IEDB server at https://www.iedb.org/ as well as previously reported researches as the primary positive dataset. IEDB is an online free resource for searching and exporting more than 100,000 unique immune epitopes. Furthermore, a set of experimentally validated non-epitope peptides from *CCHFV* proteins were also retrieved from the IEDB server as negative data set.

2.2. Sequence retrieving, epitope prediction and completing datasets

The amino acid sequences of CCHFV proteins including nucleoprotein, envelope polyprotein and RNA dependent polymerase were obtained from Uniprot database at https://www.uniprot.org/ in Fasta format as inputs for epitope prediction process. Furthermore, to complete the primary datasets as well as increase the accuracy, B- and T-cell epitope prediction was also done. To this, the linear B-cell epitope(s) from CCHFV proteins were predicted using three different online servers including ABCpred at https://webs.iiitd.edu.in/raghava/abcpred/, BCPerd at http://ailab.ist.psu.edu/bcpred/predict.html and IEDB analysis resource at http://tools.iedb.org/bcell/. Subsequently, the top five highest scored B-cell epitopes, especially shared epitopes between the servers were added in to experimentally confirmed B-cell epitopes dataset. Similarly, MHC-II restricted T-cell epitopes from the proteins were determined using a set of online web servers including IEDB analysis resource/MHC-II binding prediction at http://tools.iedb.org/mhcii/, NetMHCIIpan 3.2 at http://www.cbs.dtu.dk/services/NetMHCIIpan/ and EpiTOP 3.0 at http://www.ddg-pharmfac.net/EpiTOP3/. Like linear B-cell epitopes, the predicted T-cell epitopes with highest scores were added in to T-cell epitopes dataset. Finally, two datasets including B- and T-cell epitopes (empirically approved + predicted), and non-epitopes, were prepared for further analysis.

2.3. Amino acid composition and physicochemical properties analysis

To investigate the proportion of each amino acid in the sequences of B- and T-cell epitopes as well as non-epitope peptides, the amino acid composition analysis was used through COPid web server at https://webs.iiitd.edu.in/raghava/COPid/index.html. Furthermore, to evaluate the physicochemical properties differences between the epitope and non-epitope peptides five physicochemical properties of the peptides including molecular weight, theoretical pI, extinction coefficient, aliphatic index and grand average of hydropathicity were computed by Prot Param server at http://web.expasy.org/protparam. Finally, probable statistically differences between the epitopes and non-epitopes datasets were evaluated by unpaired *t*-test using JASP 0.8.6 software.

2.4. Pseudo-amino acid composition (PseAAC) and data classification

Performing biochemical experiments to acquire the desired information about biomacromolecules especially proteins is tedious and costly. Therefore, recently the computational methods especially machine-learning based tools are highly regarded for fast and accurately identifying various characteristics of proteins based on their sequences. PseAAC is a powerful interdisciplinary method with the advantage of avoiding loss of the sequence-order information, which has been proven very useful in studying proteome and genome analysis [35]. Psein- One is a flexible web-server available at http://bioinformatics.hitsz.edu.cn/Pse-in-One2.0/ which can be used to generate any desired feature vectors for protein/peptide and DNA/RNA sequences according to the need of users' studies. In this study, the B- and T-cell epitopes from *CCHFV* proteins as well as non-epitope peptides were subjected to the classic Pseudo-amino acid composition method by Pse-in-One to analyze and compare based on their physicochemical properties. In Classic PseAAC the sequence-order pattern is reflected by $20 + \lambda$ separate

Table 1
Top five scored predicted linear B-cell epitopes from CCHFV proteins by three different online tools (ABCpred, BepiPred-2.0 and IEDB).

Antigen	ABCpred		BepiPred-2.0		IEDB	
	Sequence	Score	Sequence	Score	Sequence	Score
Envelopment polyprotein	LRKPLFLDSTAKGMKNLL	0.95	PHPVSNRPPTPPATAQGPTE	1	PGDNPSS	1.497
	SIFKEHREVEINVLLPQV	0.93	KTDTMTTPGDNPSSEPPVST	1	NYGGPGD	1.480
	DNLIDLGCPKIPLLGKMA	0.90	PGPDETSTPSGTGKESSATS	1	DNYGGPG	1.480
	PKTTMAFLFWFSFGYVIT	0.89	LDPSTVTPTTPASGLEGSGE	0.999	GDNYGGP	1.480
	EVLQFRTPGTLSTTESTP	0.88	IHGDNYGGPGDKITICNGST	0.999	TPGDNPS	1.43
Nucleoprotein	QDMDIVASEHLLHQSLVG	0.88	LILNRGGDENPRGPVSHEHV	0.998	NKSGRSG	1.357
	LLKHIAKAQELYKNSSAL	0.88	GTIPVANPDDAAQGSGHTKS	0.993	NRGGDEN	1.341
	RGGDENPRGPVSHEHVDW	0.87	SQFLFELGKQPRGTKKMKKA	0.980	RGGDENP	1.336
	EGKGIFDEAKKTVEALNG	0.84	YIMAFNPPWGDINKSGRSGI	0.978	GGDENPR	1.336
	TKFCAPIYECAWVSSTGI	0.83	QAALKWRKDIGFRVNANTAA	0.978	GDENPRG	1.336
RNA dependent polymerase	TTYHENLLKVKHLVDCST	0.93	SDYFEIVRQPGDGNCFYHSI	0.997	DSGSSSS	1.453
	LDESAISEVKPTKVDFSN	0.93	LSTFLYGSNNKNKKKFITNC	0.943	SGSSSSS	1.449
	REITFALEGRFEESYKIR	0.93	DGDPAEQGNQQSITEHESSS	0.990	GDSNRSG	1.421
	GRLFVPTYSGLVSSAVAL	0.91	TNHCSNCHPNGVNISNTSNV	0.953	DSWDGND	1.413
	APKAQLGGARDLLVQETG	0.91	KAGTATKTPVSTKDVLETWE	0.952	PDSSNPR	1.410

Table 2
Top five scored predicted MHC-II restricted T-cell epitopes from CCHFV proteins by three different online tools (EpiTOP3, NetMHCIIpan and IEDB).

Antigen	EpiTOP3		NetMHCIIpan		IEDB	
	Sequence	Bonded alleles	Sequence	%Rank	Sequence	IC50
Envelope polyprotein	LCLQLCGLG	11	FYGLKNMLS	1.5	INRVRSFKL	2.90
	WCKRNLGLD	11	LKNMLSGIF	5	FYGLKNMLS	3.80
	YFAKGFLSI	12	FDLMHVQKV	4.5	LFFMFGWRI	4.60
	INLKASIFK	11	VQKVLSAST	5	INRVRSFKL	4.80
	VLLPQVAVN	11	FTVSLSPVQ	6.5	LFLDSTAKG	5.30
Nucleoprotein	MSVKEMLSD	10	FVFQMASAT	2.50	QELYKNSSALRAQSA	9.70
	IRRRNLILN	11	FRVNANTAA	1.80	AQELYKNSSALRAQS	10.50
	IQDMDIVAS	10	YIMAFNPPW	1.50	KAQELYKNSSALRAQ	13.50
	LKWRKDIGF	12	YKNSSALRA	1.80	ELYKNSSALRAQSAQ	13.80
	VLAEYKVPG	11	IYMHPAVLT	1.60	KGKYIMAFNPPWGDI	11.60
RNA dependent polymerase	FNHREIADL	10	FISVVSGLN	1.60	VNSDRQLIF	2.70
	YVSNPRFNI	12	FYMLKGNLM	3.5	LNHALSLMF	2.90
	YIKRLTESA	12	YKVLGNLGN	1.70	VNSDRQLIF	3.00
	YQEEPEARL	11	FISTSGRAM	2	LNHALSLMF	3.00
	FNCKLCVEI	12	VRVLKNSVS	1.40	MQLLHSEMI	3.10

Table 3
The mean of amino acid proportion in B- and T-cell epitopes as well as non-epitope peptides from Envelope polyprotein, nucleoprotein and RNA dependent polymerase of CCHFV.

Amino acid	Studied dataset		
	Non-epitope	B-cell epitope	T-cell epitope
Glycine	7.732	8.67	4.048
Alanine	6.298	6.667	6.285
Leucine	8.843	6.521	13.697
Methionine	1.365	1.533	2.748
Phenylalanine	3.876	2.233	7.599
Tryptophan	2.73	2.865	2.742
Lysine	7.741	5.485	7.196
Glutamine	6.348	6.018	5.44
Glutamic Acid	7.732	8.67	4.048
Serine	7.787	8.945	5.735
Proline	3.92	5.17	3.09
Valine	6.247	5.129	6.215
Isoleucine	5.719	5.862	8
Cysteine	3.792	3.205	3.256
Tyrosine	2.73	2.865	2.742
Histidine	2.465	2.854	1.532
Arginine	3.708	4.8	5.464
Asparagine	4.806	6.82	3.031
Aspartic Acid	5.081	5.497	5.567
Threonine	6.524	7.342	3.9

components. Here, six different physicochemical properties including hydrophilicity, hydrophobicity, molecular weight, pK-C, pK-N and Isoelectric point with weight factor = 0.05 and $\lambda = 1$ incorporated to compare the datasets.

Discrimination between the datasets based on PseAAC outputs was characterized using different machine learning algorithms including support vector machine (SVM), Random forests, K-Nearest Neighbor (KNN) and artificial neural network (ANN) in Orange 3.20 software. The performance of used classifiers was defined by accuracy (ACC), precision or positive predictive value (PPV), Recall (or sensitivity) and F1-score which were calculated by the following equations:

$$\mathrm{ACC} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\mathrm{PPV} = \frac{TP}{TP + FP} \quad (2)$$

$$\mathrm{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\mathrm{F1} = 2 \times \frac{PPV \times Recall}{PPV + Recall} \quad (4)$$

where, TP, TN, FP and FN in the listed above equations are the total numbers of positive samples, the number of the positive samples wrongly predicted to be the negative, the total number of negative

Table 4
The means of six physicochemical properties of studied B- and T-cell epitopes as well as non-epitope peptides from CCHFV proteins (GRAVY is abbreviation for Grand average of hydropathicity).

Dataset	Molecular weight (Da)	Theoretical pI	Half-life (hour)	Instability index	Aliphatic index	GRAVY
B-cell epitopes	1667.2	29.62	10.27	40.58	68.9	0.9
T-cell epitopes	1204.62	6.94	18.02	36.78	108.66	1.62
Non-epitopes	1823.05	6.808	12.73	33.15	77.97	0.58

Table 5
The performances of the studied classifier algorithms for classifying between epitope and non-epitope peptides from CCHFV proteome based on PseAAC outputs. The Neural network and SVM had most and least accuracy respectively.

Classifier algorithm	AUC	Accuracy	F1-score	Precision	Recall
kNN	0.838	0.764	0.766	0.700	0.846
SVM	0.529	0.558	0.569	0.512	0.640
Random forest	0.898	0.820	0.809	0.786	0.834
Neural network	0.922	0.900	0.887	0.890	0.914

samples and the number of the negative samples falsely predicted to be the positive respectively. Finally, the algorithm with the highest ACC as well as appropriate PPV, sensitivity and F1-score was selected for further analysis.

2.5. Peptide library generation and epitope prediction challenge

The accuracy of our proposed method to the prediction of high antigenic B- and T-cell epitopes was compared to the used online webservers (are mentioned in 1.2 section). To this end, high antigenic protein from the virus was selected as a test antigen based on its antigenicity score, presented by Vaxijen at http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html. Consequently, the amino acid sequence of the top-scored protein was subjected to peptide library design tool at https://www.genscript.com/peptide_screening_tools.html to generate an overlapping peptide library. The peptide library generation process is defined by two crucial factors including peptide length and offset number. The peptide length reflects the epitope length and the offset number determines the degree of overlapping. To generate the peptide library from the protein, mentioned parameters were set at 18 and three respectively. Subsequently, the overlap peptides were subjected to PseAAC to generate a new test datasets for epitope prediction using the most accurate algorithm. Finally, the predicted B- and T-cell epitope(s) by the proposed method as well as five top-scored epitopes determined by the common epitope prediction tools were selected for more analysis.

2.6. Antigenicity prediction and molecular docking

The predicted B- and T-cell epitopes by our proposed method as well as the common tools were compared in term of antigenicity and affinity to MHC-II respectively. The antigenicity and affinity of the predicted epitopes were evaluated by Vaxijen score and molecular docking method respectively. To molecular docking evaluation, three-dimension structure of MHC-II was extracted from Protein data bank at https://www.rcsb.org/ with PDB entry of 1AQD in pdb format. Furthermore, the raw structure of MHC-II was optimized in term of energy and geometry using AMBER99 force field in MOE 2010 software. Finally, the optimized structures were considered to molecular docking study by MDockPeP at http://zougrouptoolkit.missouri.edu/mdockpep/index.html. Finally, the studied epitopes were compared based on their averages of antigenicity and the results of molecular docking evaluations by one-way ANOVA analysis.

2.7. Epitope screening

The B- and T-cell epitopes predicted by our proposed method which showed appropriate antigenicity and affinity to MHC-II were further studied in term of water solubility, allergenicity and cytotoxicity thought Pepcalc at https://pepcalc.com/, AllergenFP v.1.0 at http://ddg-pharmfac.net/AllergenFP/feedback.py and Toxin pred at https://webs.iiitd.edu.in/raghava/toxinpred/index.html respectively. Subsequently, the epitope(s) with good water solubility as well as without cytotoxicity and allergenicity potential were considered for investigating possible immune-responses.

2.8. Statistics analysis

To evaluate the statistically significant difference between B- and T-cell epitopes from CCHFV proteins and non-epitope peptides in term of physicochemical properties and amino acid proportions an unpaired *t*-test analysis with a significance level of 0.05 was done using JASP 0.8.6 software. The *t*-test is an inferential statistical test, which compares the means of two independent groups to determine whether there is a statistically significant difference between the groups. Additionally, to the comparison of antigenicity and docking scores of the predicted B- and T-cell epitopes by the proposed method as well as the online tools one-way ANOVA analysis with Scheffe post hoc test in JASP 0.8.6 software was performed. One-way ANOVA also is known as one-factor ANOVA is a standard statistical test to determine whether there are any statistically significant differences between the means of two or more unrelated groups. Furthermore, to comparison the accuracy of used classifiers Receiver-Operating Characteristic analysis (ROC) was performed using Orange 3.20 software. ROC analysis also written as AUROC (Area Under the Receiver Operating Characteristics) is a beneficial tool for calculating the performance of diagnostic tests and evaluating the accuracy of a statistical model. The results of the ROC analysis reveal the classification accuracy (ACC) and the area under the curve (AUC). The AUC is an overall summary of accuracy. Generally, the ROC curve corresponds to random chance when AUC be equals to 0.5 and represents perfect accuracy when AUC be 1.

2.9. Immunogenicity validation and immune simulation

As the final step, immunogenicity of the screened B- and T-cell epitopes as well as non-epitopes which predicted by our proposed method was investigated. For this purpose affinity of the peptides to the B-cell receptor (with PDB entry of 2H32) and MHC-II molecules were calculated by molecular docking via MDockPeP server after optimization of primary structures trough AMBER99 force field in MOE2010 software. Moreover, antigenicity score and antigenicity propensity of all tested peptides were also measured by Vaxijen and online server for the Kolaskar and Tongaonkar method (available at http://imed.med.ucm.es/Tools/antigenic.pl) respectively. Additionally, to predict probable immune-responses against the screened epitope(s), computational immune simulations were performed via the C-ImmSim server at http://150.146.2.1/C-IMMSIM/index.php. C-ImmSim is a flexible online tool which uses a position-specific scoring matrix (PSSM) and machine learning techniques for prediction of immune interactions and epitope determination. The selected parameters in C-ImmSim for predicting humoral and cellular responses to the epitopes were set at default.

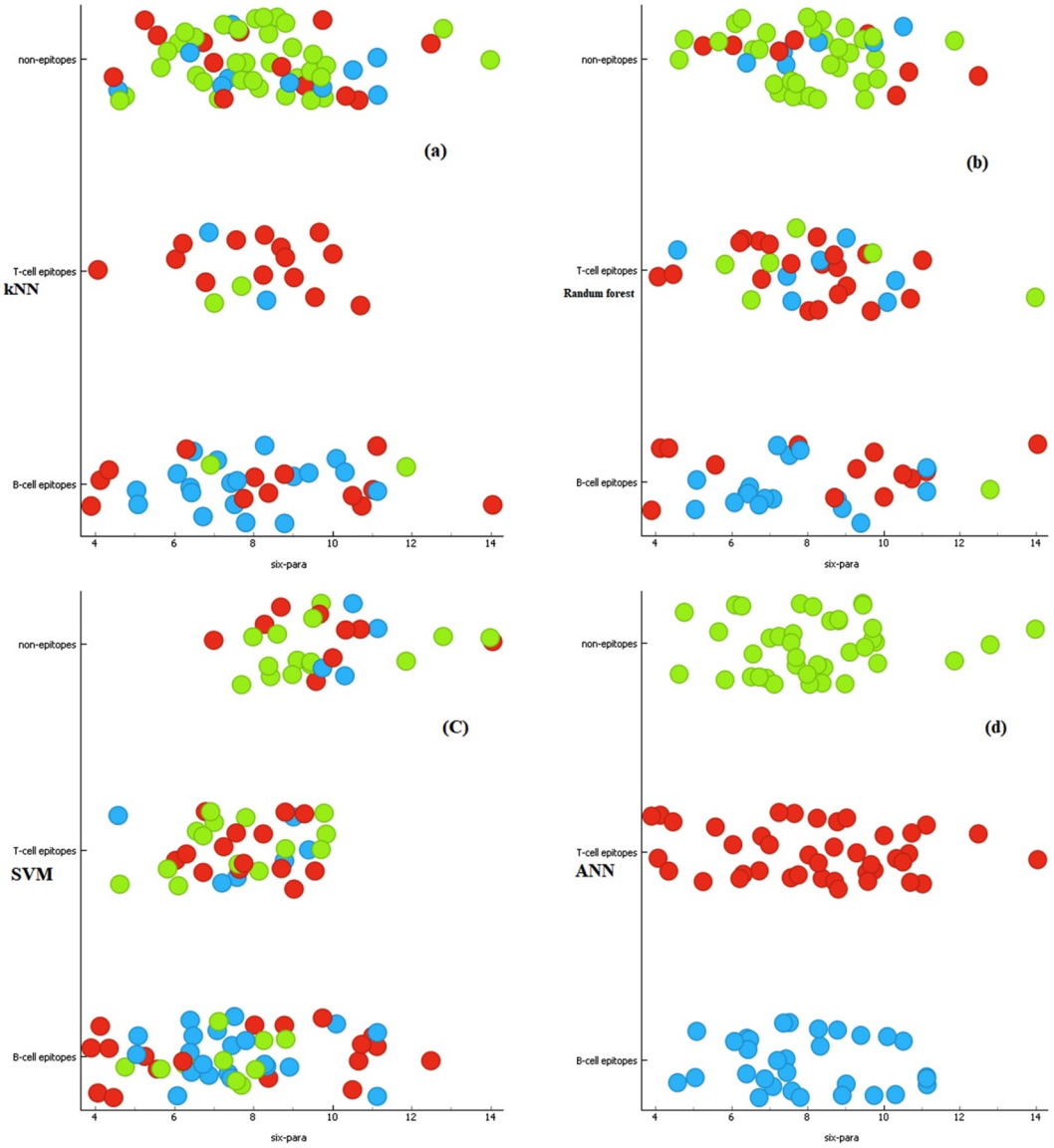

Fig. 1. The performance of four tested machine learning algorithms (a: kNN, b: random forest, c: SVM and d: ANN) for classifying B- and T-cell epitopes as well as non-epitopes peptides from envelope polyprotein of CCHFV. The ANN algorithm showed highest accuracy with score of 0.90 followed by random forest, kNN and SVM respectively. The blue, red and green circles represent the B-cell epitopes, T-cell epitopes and non-epitopes peptides. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Results

3.1. Data collection and datasets preparation

As primary datasets a total number of 150 and 147 B- and T-cell epitopes from *CCHFV* proteins were retrieved from previous studies and IEDB database respectively (supplementary Table 1). Moreover, a set of 331 experimentally confirmed non-epitope peptides from different *CCHFV* proteins were also extracted from the IEDB database as negative dataset (supplementary Table 2).

3.2. Sequence retrieving, epitope prediction and completing datasets

The amino acid sequences of *CCHFV* proteins including nucleoprotein, envelope polyprotein and RNA dependent polymerase were extracted from Uniprot with the accession number of P89522, Q8JSZ3 and Q6TQR6 respectively. To increase the accuracy of the aimed method for predicting potent B-and T-cell epitopes from *CCHFV* proteome, the primary datasets were completed by adding predicted B- and T-cell epitopes. A total number of 45 top-scored linear B-cell epitopes (Table 1) were predicted using ABCpred, BepiPred-2.0 and IEDB. Similarly, the total number of 45 top-ranked MHC-II restricted T-cell epitopes (Table 2) were determined from the subjected proteins for completing the epitopes dataset. The completed B and T-cell epitope dataset was consists of 387 epitopes.

3.3. Amino acid composition and physicochemical properties analysis

Amino acid composition and physicochemical properties are vital factors in the appearance of immunological responses against antigens as well as host recognition of pathogen patterns and epitope regions. Therefore, determining significance differences between epitope and non-epitope peptides in term of the mentioned factors in the tested antigens can improve the accuracy of the epitope predicting pipeline. Consequently, probable differences between studied datasets were investigated by unpaired *t*-test. The mean of amino acids proportions in the selected B- and T-cell epitopes as well as non-epitope peptides are presented in Table 3. Furthermore, the mean of six determining

Table 6
The results of epitope prediction challenge by the trained neural network algorithm; among 112 tested peptides 39 peptides were defined as linear B-cell epitope, while 33 and 40 peptides were predicted as T-cell epitope and non-epitope respectively.

Peptide sequence	Result of prediction	Peptide sequence	Result of prediction
MHISLMYAILCLQLCGLG	T-cell epitope	EKLNNKKGKNKLLDGERL	T-cell epitope
VTPTTPASGLEGSGEVYT	T-cell epitope	GLGETHGSHNETRHNKTD	B-cell epitope
SAPTVRTSLPNSPSTPST	B-cell epitope	VYTSPPITTGSLPLSETT	T-cell epitope
NRPPTPPATAQGPTENDS	B-cell epitope	PSTPQDTHHPVRNLLSVT	B-cell epitope
PTNRSKRNLKMEIILTLS	T-cell epitope	NDSHNATEHPESLTQSAT	B-cell epitope
QKRIEEFFITGEGHFNEV	B-cell epitope	TLSQGLKKYYGKILRLLQ	T-cell epitope
AKCYSGTSNSGLQLINIT	B-cell epitope	NEVLQFRTPGTLSTTEST	Non-epitope
SHVCDYSLDIDGAVRLPH	B-cell epitope	NITRHSTRIVDTPGPKIT	Non-epitope
QSVLRQYKTEIRIGKAST	T-cell epitope	LPHIYHEGVFIPGTYKIV	Non-epitope
CNGSTIVDQRLGSELGCY	Non-epitope	ASTGSRRLLSEEPSDDCI	Non-epitope
GHVKLSRGSEVVLDACDT	Non-epitope	GCYTINRVRSFKLCENSA	Non-epitope
MAIYICRMSNHPKTTMAF	Non-epitope	CDTSCEIMIPKGTGDILV	B-cell epitope
MHDLNCSYNICPYCASRL	B-cell epitope	MAFLFWFSFGYVITCILC	Non-epitope
ESTGVALKRSSWLIVLLV	T-cell epitope	SRLTSDGLARHVIQCPKR	Non-epitope
IVSCLMKGLVDSVGNSFF	Non-epitope	LLVLFTVSLSPVQSAPIG	Non-epitope
IHKLHLSICKKRKKGSNV	T-cell epitope	SFFPGLSICKTCSISSIN	B-cell epitope
VSTSAVEMENLPAGTWER	T-cell epitope	SNVMLAVCKLMCFRATME	B-cell epitope
NLLNSTSLETSLSIEAPW	T-cell epitope	APWGAINVQSTYKPTVST	T-cell epitope
KLEERTGISWDLGVEDAS	B-cell epitope	DASESKLLTVSVMDLSQM	T-cell epitope
ERCGCTSSTCLHKEWPHS	B-cell epitope	PHSRNWRCNPTWCWGVGT	Non-epitope
TEAIVCVELTSQERQCSL	B-cell epitope	CSLIEAGTRFNLGPVTIT	Non-epitope
DLMHVQKVLSASTVCKLQ	B-cell epitope	KLQSCTHGVPGDLQVYHI	Non-epitope
WDGCDLDYYCNMGDWPS	B-cell epitope	PSCTYTGVTQHNHASFVN	T-cell epitope
LDLKARPTYGAGEITVLV	Non-epitope	VLVEVADMELHTKKIEIS	B-cell epitope
DEPDELTVHVKSDDPDVV	B-cell epitope	DVVAASSSLMARKLEFGT	T-cell epitope
CSEEDTKKCVNTKLEQPQ	B-cell epitope	QPQSILIEHKGTIIGKQN	T-cell epitope
FILLILFFMFGWRILFCF	T-cell epitope	RGNKILVSGRSESIMKLE	T-cell epitope
ERLADRRIAELFSTKTHI	T-cell epitope	QVGEWPKATCTGDCPERC	Non-epitope
KTDTMTTPGDNPSSEPPV	B-cell epitope	FTDYMFVKWKVEYIKTEA	T-cell epitope
ETTPELPVTTGTDTLSAG	B-cell epitope	PEIITLHPRIEEGFFDLM	B-cell epitope
SVTSPGPDETSTPSGTGK	T-cell epitope	LIHKIEPHFNTSWMSWDG	Non-epitope
SATPGLMTSPTQIVHPQS	B-cell epitope	HFHSKRVTAHGDTPQLDL	Non-epitope
LLQLTLEEDTEGLLEWCK	B-cell epitope	YACSSGISCKVRIHVDEP	Non-epitope
ESTPAGLPTAEPFKSYFA	Non-epitope	TSLCFYIVEREHCKSCSE	T-cell epitope
KITNLKTINCINLKASIF	Non-epitope	SVKSFFYGLKNMLSGIFG	B-cell epitope
KIVIDKKNKLNDRCTLFT	T-cell epitope	RHLKDDEETGYRRIIEKL	B-cell epitope
DCISRTQLLRTETAEIHG	T-cell epitope	PQVAVNLSNCHVVIKSHV	T-cell epitope
NSATGKNCEIDSVPVKCR	Non-epitope	CTICETTPVNAIDAEMHD	Non-epitope
ILVDCSGGQQHFLKDNLI	T-cell epitope	CTICETTPVNCHVDAHD	B-cell epitope
ILCKAIFYLLIIVGTLGK	B-cell epitope	KQNSTCTAKASCWLESVK	Non-epitope
PKRKEKVEETELYLNLER	Non-epitope	FCFKCCRRTRGLFKYRHL	B-cell epitope
PIGQGKTIEAYRAREGYT	T-cell epitope	PPVSTALSITLDPSTVTP	T-cell epitope
SINGFEIESHKCYCSLFC	B-cell epitope	SAGDVDPSTQTAGGTSAP	B-cell epitope
TMEVSNRALFIRSIINTT	B-cell epitope	TGKESSATSSPHPVSNRP	Non-epitope
QVTETECLCPYEALVLRK	T-cell epitope	PQSATPITVQDTHPSPTN	B-cell epitope
VSTANIALSWSSVEHRGN	Non-epitope	WCKRNLGLDCDDTFFQKR	B-cell epitope
SQMYSPVFEYLSGDRQVG	B-cell epitope	YFAKGFLSIDSGYYSAKC	Non-epitope
VGTGCTCCGLDVKDLFTD	Non-epitope	SIFKEHREVEINVLLPQV	Non-epitope
TITLSEPRNIQQKLPPEI	T-cell epitope	LFTDCVIKGREVRKGQSV	Non-epitope
YHIGNLLKGDKVNGHLIH	T-cell epitope	IHGDNYGGPGDKITICNG	B-cell epitope
FVNLLNIETDYTKNFHFH	Non-epitope	KCRQGYCLRITQEGRGHV	B-cell epitope
EISGLKFASLACTGCYAC	T-cell epitope	NLIDLGCPKIPLLGKMAI	Non-epitope
FGTDSTFKAFSAMPKTSL	Non-epitope	LGKRLKQYRELKPQTCTI	Non-epitope
LFCCPYCRHCSTDKEIHK	Non-epitope	LERIPWVVRKLLQVSEST	T-cell epitope
LRKPLFLDSTAKGMKNLL	B-cell epitope	GYTSICLFVLGSILFIVS	T-cell epitope

physicochemical properties were computed to evaluate probable differences between the datasets (Table 4). The results of *t*-test demonstrated that molecular weight, theoretical pI, instability index and aliphatic index are statistically different parameters between studied B-cell epitopes and non-epitope peptides (supplementary Table 3). The results also confirmed that aliphatic index, half-life, molecular weight and instability index were suggesting statistically differences between evaluated T-cell epitopes and non-epitope peptide (supplementary Table 4). Moreover, it was cleared that there also are statistically significant differences between tested B- and T-cell epitopes in term of the aliphatic index, theoretical pI, half-life and molecular weight (supplementary Table 5).

3.4. PseAAC and data classification

Classification between epitope and non-epitope peptides from *CCHFV* proteome based on the PseAAC outputs was performed by four different classifiers. The results of the data classification are summarized in Table 5. Comparison between the performances of tested classifiers demonstrated that neural network shows most performance in categorizing of the studied datasets with an accuracy of 0.900 followed by random forest, kNN and SVM respectively. Furthermore, the results of ROC analysis also confirmed that the neural network is the accurate algorithm for labeling the tested datasets with the AUC of 0.922. Therefore, neural network was selected for further analysis. As depicted

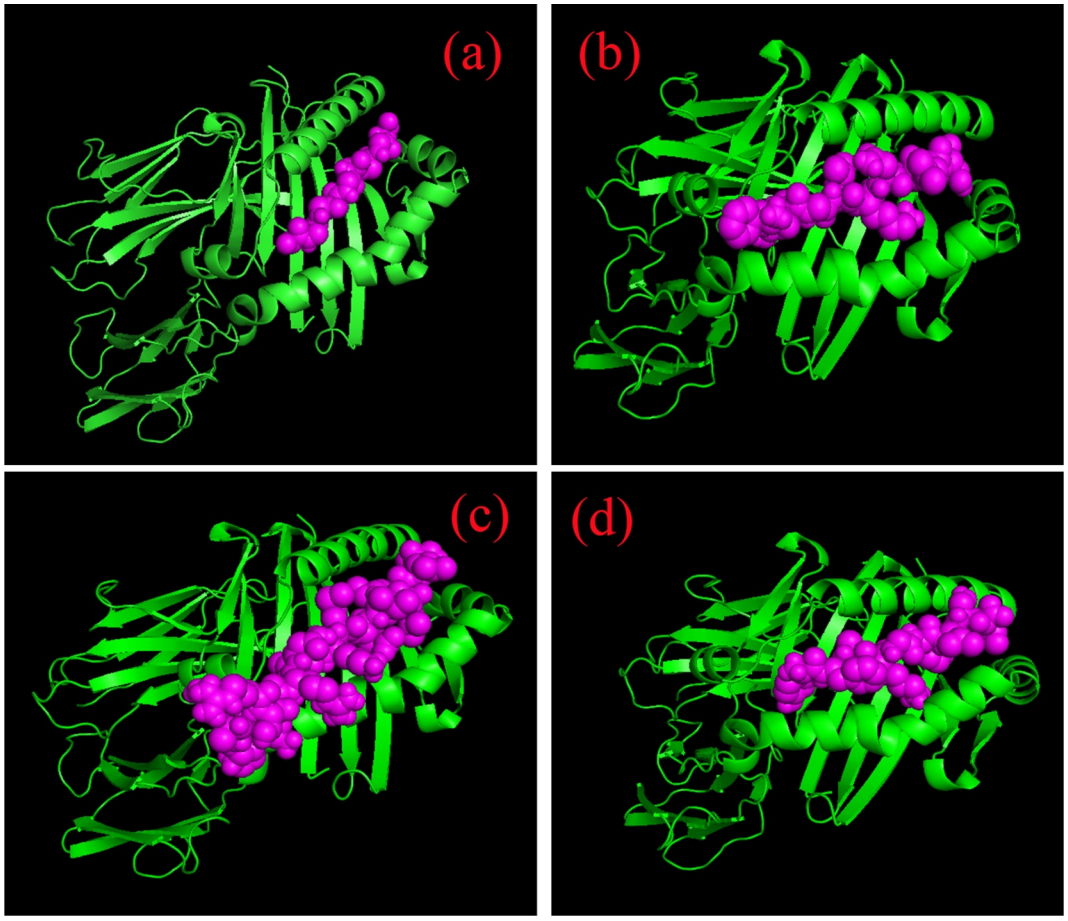

Fig. 2. Schematic representation of molecular docking results performed between the top score T- cell epitope predicted by our purposed method (c), IEDB (a), NetMHCIIpan (b) and EpiTOP3 (d) against MHC-II molecules.

Table 7
The screened B-and T-cell epitope among the predicted epitopes by our proposed method.

Screened Linear B-cell epitope	Screened T-cell epitope
NRPPTPPATAQGPTENDS	PTNRSKRNLKMEIILTLS
TEAIVCVELTSQERQCSL	VSTSAVEMENLPAGTWER
CSEEDTKKCVNTKLEQPQ	NLLNSTSLETSLSIEAPW
LLQLTLEEDTEGLLEWCK	ERLADRRIAELFSTKTHI
LRKPLFLDSTAKGMKNLL	EKLNNKKGKNKLLDGERL
VLVEVADMELHTKKIEIS	DASESKLLTVSVMDLSQM
FCFKCCRRTRGLFKYRHL	TSLCFYIVEREHCKSCSE
IHGDNYGGPGDKITICNG	–

in Fig. 1 the trained neural network algorithm can classify the selected datasets with the highest performance based on the selected amino acid character in PseAAC in comparison to the other tested classifiers.

3.5. Peptide library generation and epitope prediction challenge

To evaluation of the efficacy of our proposed to accurate epitope determination a prediction challenge was performed. For this purpose, firstly, the *CCHFV* protein with the highest antigenicity was selected based on Vaxijen score. The results confirmed that the envelopment polyprotein was the most antigenic protein with antigenicity score of 0.5145 followed by RNA dependent RNA polymerase and nucleoprotein respectively. Therefore the envelopment polyprotein was selected for peptide library generation and epitope prediction challenge. A total number of 112 overlapping peptides were generated based on the selected parameters. Furthermore, an epitope prediction challenge was carried out by the trained neural network algorithm. The results showed that among 112 tested peptides 39 peptides were defined as linear B-cell epitope, while 33 and 40 peptides were predicted as T-cell epitope and non-epitope respectively (Table 6).

3.6. Antigenicity prediction and molecular docking

To comparison between the efficacy of our proposed method and the standard epitope prediction tools in term of antigenicity and affinity to MHC-II; one-way ANOVA analysis with Scheffe post hoc test was performed. The results of one-way ANOVA analysis revealed that there is a statistical difference between the means of antigenicity scores of the linear B-cell epitopes predicted by our proposed method and other tested online tools (p-value < 0.001; $F = 8.917$). Furthermore, the results of ANOVA analysis for the means of docking scores from the T-cell epitopes also confirmed that the molecular docking scores of the T-cell epitopes predicted by our method are statistically more than the other tested methods (p-value = 0.001; $F = 7.836$). Moreover, as depicted in Fig. 2 and supplementary Figs. 1–4 the top scored T-cell epitope predicted by our method has more affinity to the receptor than the three top score T-cell epitopes predicted by other tested standard servers. The results of post hoc test (supplementary Tables 6 and 7) demonstrated that both antigenicity and docking scores of the predicted B- and T-cell epitopes by our method are significantly more than the predicted epitopes using the studied tools.

3.7. Epitope screening

To determine high antigenic and safe B- and T-cell epitopes among

Table 8
The results of evaluation immunogenicity of the screened B- and T-cell epitopes as well as non-epitopes predicted by our proposed method (BCR, AAP and N.E are abbreviation for B-cell receptor, Average antigenic propensity and non-epitope respectively).

Epitope sequence	Type	Affinity to BCR (score)	Affinity to MHC-II (score)	AAP	Vaxijen score
NRPPTPPATAQGPTENDS	B	−320.1	−196.3	0.995	0.34
TEAIVCVELTSQERQCSL	B	−321.5	−216.6	1.07	1.300
CSEEDTKKCVNTKLEQPQ	B	−311.8	−162.8	1.02	0.44
LLQLTLEEDTEGLLEWCK	B	−299.4	−180.6	1.04	0.28
LRKPLFLDSTAKGMKNLL	B	−290.1	−178.3	1.07	0.04
VLVEVADMELHTKKIEIS	B	−300.3	−120.5	1.06	1.34
FCFKCCRRTRGLFKYRHL	B	−298.7	−155.8	1.06	1.02
IHGDNYGGPGDKITICNG	B	−301.2	−169.2	0.98	0.93
PTNRSKRNLKMEIILTLS	T	−278.6	−198.3	0.92	0.98
VSTSAVEMENLPAGTWER	T	−290.1	−200.2	0.90	0.68
NLLNSTSLETSLSIEAPW	T	−261.5	−263.2	0.89	0.63
ERLADRRIAELFSTKTHI	T	–222.3	−278.4	0.91	0.27
EKLNNKKGKNKLLDGERL	T	−270.1	−290.7	0.93	0.06
DASESKLLTVSVMDLSQM	T	−218.2	−290.4	1.05	0.51
TSLCFYIVEREHCKSCSE	T	−213.4	−203.7	0.89	1.01
CNGSTIVDQRLGSELGCY	N.E	−190.1	−167.2	0.88	0.66
GHVKLSRGSEVVLDACDT	N.E	−207.1	−134.6	0.90	0.02
MAIYICRMSNHPKTTMAF	N.E	−192	−144.8	1.00	−0.24
IVSCLMKGLVDSVGNSFF	N.E	−231.4	−150.2	0.80	−0.24
LDLKARPTYGAGEITVLV	N.E	−200.3	−167.3	1.04	1.17
NEVLQFRTPGTLSTTEST	N.E	−210.4	−122.7	0.99	0.04
NITRHSTRIVDTPGPKIT	N.E	−183.5	−110.8	1.00	0.08
LPHIYHEGVFIPGTYKIV	N.E	−170.4	−178.3	0.72	0.10
ASTGSRRLLSEEPSDDCI	N.E	−182.4	−180.4	1.00	0.06
GCYTINRVRSFKLCENSA	N.E	−194.5	−191.4	1.04	0.35
MAFLFWFSFGYVITCILC	N.E	−200.1	−145.2	0.76	0.78
SRLTSDGLARHVIQCPKR	N.E	−203.8	−155.2	0.88	−0.04
LLVLFTVSLSPVQSAPIG	N.E	−199.3	−137.2	0.99	0.09
PHSRNWRCNPTWCWGVGT	N.E	−152.3	−194.2	0.83	0.82
CSLIEAGTRFNLGPVTIT	N.E	−145.8	−183.6	0.84	1.42
KLQSCTHGVPGDLQVYHI	N.E	−132.7	−200.1	0.89	0.11
QVGEWPKATCTGDCPERC	N.E	−276.6	−176.3	0.90	0.39
ESTPAGLPTAEPFKSYFA	N.E	−234.1	−181.4	0.92	0.18
KITNLKTINCINLKASIF	N.E	−167.2	−188.3	0.93	0.07
NSATGKNCEIDSVPVKCR	N.E	−180.4	−124.5	0.92	1.02
PKRKEKVEETELYLNLER	N.E	−193.2	−148.3	0.99	0.06
VSTANIALSWSSVEHRGN	N.E	−130.3	−160.2	0.91	0.16
VGTGCTCCGLDVKDLFTD	N.E	−192.7	−180.4	0.93	0.98
FGTDSTFKAFSAMPKTSL	N.E	−180.2	−179.3	0.77	0.33
LFCCPYCRHCSTDKEIHK	N.E	−200.7	−188.5	0.78	0.15
LIHKIEPHFNTSWMSWDG	N.E	−203.7	−157.3	0.87	0.05
HFHSKRVTAHGDTPQLDL	N.E	−143.9	−162.5	0.90	0.10
YACSSGISCKVRIHVDEP	N.E	−156.2	−177.4	0.92	−0.10
CTICETTPVNAIDAEMHD	N.E	−176.2	−180.5	1.00	0.53
KQNSTCTAKASCWLESVK	N.E	−180.2	−134.6	1.01	0.94
TGKESSATSSPHPVSNRP	N.E	−190.2	−133.9	1.10	0.05
YFAKGFLSIDSGYYSAKC	N.E	–222.3	−167.2	0.67	0.70
SIFKEHREVEINVLLPQV	N.E	−186.2	−202.8	0.87	0.16
LFTDCVIKGREVRKGQSV	N.E	−194.2	−172.8	0.97	0.91
NLIDLGCPKIPLLGKMAI	N.E	−200.1	−146.2	0.99	0.08
LGKRLKQYRELKPQTCTI	N.E	−194.2	−144.3	0.88	0.05

the predicted epitope by our proposed method a multi-factors screening have been done. For this purpose, the predicted epitopes from *CCHFV* envelopment polyprotein (which selected based on its antigenicity score) were evaluated in term of water solubility, allergenicity and cytotoxicity. The results of epitope screening are summarized in supplementary Table 8. Based on the epitope screening results eight and seven B- and T-cell epitopes were selected for immune response simulations respectively (Table 7).

3.8. Immunogenicity validation and immune simulation

Recognition process of different antigens by mature B lymphocytes (B cells) is based on molecular interactions between their surface receptors and the antigens. Moreover, and integral to the progress of an effective humoral response B cells utilize MHC-II antigen presentation pathway to promoting a strong cellular response trough CD4+ T cells [36]. Therefore, as a validation step affinity of all predicted B- and T-cell epitopes as well as non-epitopes to BCR and MHC-II were measured by molecular docking study. Furthermore, antigenicity score and antigenicity propensity of all tested epitopes as two determinant factors for eliciting appropriate immune responses were also monitored. The results (Table 8) confirmed that most B- and T-cell epitopes have more affinity to the receptors than the non-epitope peptides. Results also demonstrated that the epitopes have appropriate antigenicity score as well as antigenicity propensity in comparison to the studied non-epitopes.

In virtual immune response simulations the screened B- and T- cell epitopes were further evaluated for possible humoral and cellular immune responses respectively. The results confirmed that among the screened linear B-cell epitopes five of them could elicit strong humoral responses as well as proliferation of B lymphocytes. Similarly, obtained results demonstrated that four of the screened T-cell epitopes could increase of T-cell lymphocyte population. As depicted in Fig. 3, comparison the effectiveness of the epitopes showed that two peptides including “TEAIVCVELTSQERQCSL” and “PTNRSKRNLKMEIILTLS” have most efficiency among the screened B- and T-cell epitopes respectively.

4. Discussion

This study was aimed to introduce a new computational pipeline for predicting high antigenic and safe linear B- and T-cell epitopes from *CCHFV* proteome. The predicted epitopes then can be used for different therapeutics goals especially developing effective epitope based vaccine against different serotypes of *CCHFV*.

In general, developing an effective global vaccine against pathogens with high genetic variation such as *CCHFV*, *dengue fever virus* and HIV is a complex and challenging process. Therefore, recently, determining top antigenic and conserved epitopes using computational approaches is considered as an powerful alternative solution for the issue [20,37,38].

Consequently, in recent years many computational tools have been introduced for prediction of conformational and linear B-cell epitopes [31], T-cell epitopes [39], IFN-γ inducing epitopes [40], transporter associated with antigen processing (TAP) binding proteins [41] and antigenicity quantification [42].

Most available tools for immunoinformatics analysis and epitope prediction are developed based on an optimized machine learning algorithm [43,44]. Although, most of the used algorithms in available immunoinformatics tools can predict epitopes with acceptable accuracy, but the absence of specific datasets for the training of the algorithms is a serious challenge for determining epitope regions especially in variable antigens. Therefore, in this study we designed a particular epitope predictor to identify high antigenic epitopes from *CCHFV* proteins based on particular training datasets.

Based on our best knowledge, in the most available epitope prediction tools primarily linear B-cell epitope determiner servers, a set of random peptide and experimentally approved epitopes from different resources have been considered as training datasets [33,45,46]. Using random peptide as negative dataset can directly affect the accuracy of epitope prediction algorithm through increased false-positive rate [47]. Furthermore, positive data from different resources can decrease the performance of the algorithm through loss of individual motifs in origin proteins as well as amino acid propensities. Moreover, in the most available epitope prediction tools a set of fixed length peptides are considered for training the algorithms, when B- and T-cell epitopes have not a fixed amino acid length and this obstacle can decrease the performance of prediction process through reducing true-positive rate [33].

To address the mentioned limitation about the training and testing datasets in available epitope prediction tools we collected a set of

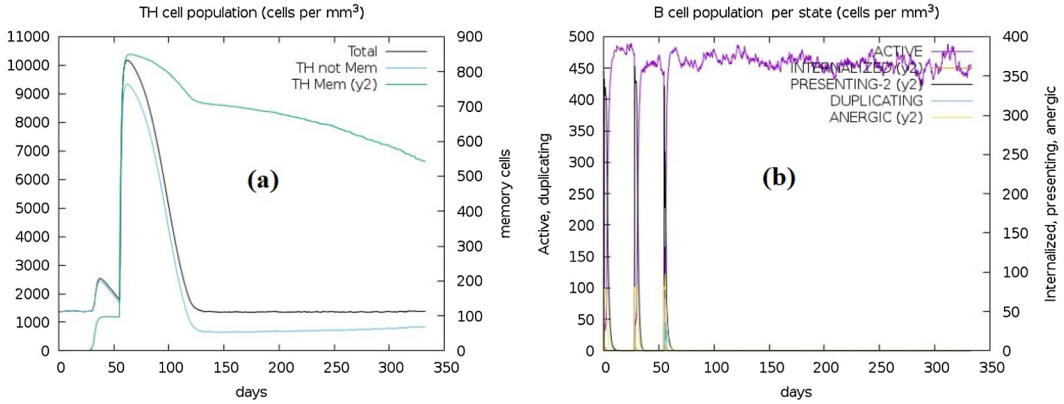

Fig. 3. Graphical representation of the results of immune response simulations after three injections of most effective B-cell epitope (TEAIVCVELTSQERQCSL) and T-cell epitope (PTNRSKRNLKMEIILTLS) predicted by our proposed method. The results showed that both B- and T-cell epitopes (a and b respectively) can increases the B- and T-cell lymphocytes population.

experimentally approved epitope and non-epitope peptides with variable length from different serotypes of *CCHFV* as positive and negative datasets respectively. Furthermore, for increasing the performance of the proposed method a set of share predicted B- and T-cell epitopes by different servers were also added in to the datasets.

Another severe restriction of standard epitope prediction tools is the presence of an optimized machine learning algorithms for only B- or T-cell epitopes when our proposed method can predict both B- and T- cell epitopes and this can decrease the time of the epitope mapping process.

To date, different machine learning algorithms such as support vector machine (SVM), random forest, artificial neural network (ANN), extremely randomized tree (ERT), gradient boosting (GB), *k*-nearest neighbors (*k*-NN) and AdaBoost (AB) have been used for developing practical epitope prediction tools. However, among the aforesaid algorithms ANN and SVM are highly utilized for immunoinformatics analysis especially for B- and T-cell epitope mapping [33,48–51]. To determine the effective machine learning algorithm for our purpose we evaluated four different algorithms including SVM, Random forests, KNN and ANN. The results confirmed that ANN has the most accuracy for classifying the epitope and non-epitope peptides.

The ANN previously has been used for developing B- and T-cell epitope prediction tools with variable efficacy [52–54]. However, using specific datasets as well as a multi-step screening process for collecting the datasets improved the algorithm accuracy in this study.

A high immunogenic B- or T-cell epitope maybe have also cytotoxic activity or inappropriate physicochemical properties. Therefore, screening of defined epitopes in term of cytotoxicity and physicochemical characteristics is a crucial step in creating training datasets. However, we screened the primary retrieved B- and T-cell epitopes for removing probable cytotoxic, allergen and poor water soluble epitopes.

To classification of desired data using the machine learning algorithms it is vital that the inputs data expressed as mathematical vectors [55,56]. Herein, we used the PseAAC method to convert amino acid composition and physicochemical properties of the epitope and non-epitope peptide from *CCHFV* proteome using Pse-in-One 2.0 online web server [57]. Ever since first appearance of the PseAAC formulation, the method has been widely applied for evaluating various problems in protein science, such as discriminating between same enzyme from different resources, protein structural classes, DNA-binding proteins, protein subcellular locations, prediction of anti-bacterial peptides, prediction of anti-cancer peptides, protein solubility, protein-protein interactions and many more [35,58–63].

However, our study was the first report about the application of PseAAC approach for epitope related data classification. As a conventional method, in available epitope prediction tools amino acid composition and physicochemical properties were considered for performing input datasets and the data classification. Using standard methods some sequence related information such as conserved motif and simultaneous effect of physicochemical properties may be lost and the accuracy of data classification decreased. Therefore, in the present study we used the parallel-correlation type of PseAAC method which generates $20 + \lambda$ discrete numbers to represent a protein [64]. The results showed that using the outputs of PseAAC method the trained ANN algorithm can discriminate between the epitope and non-epitope peptides with accuracy up to 0.90.

After training the ANN algorithm trough PseAAC outputs an epitope prediction challenge was performed on a set of random overlapping peptides from *CCHFV* envelopment polyprotein. To confirm the efficacy of the predicted B- and T- cell epitope their antigenicity and affinity to MHC-II molecule were compared to the predicted epitopes by common B- and T-cell epitope prediction tools using Vaxijen score prediction and molecular docking approach respectively. The antigenicity score prediction and molecular docking evaluation previously have been used for assessment of in silico designed epitope-based vaccine such as *hepatitis C virus* (*HCV*) [21], *Dengue fever virus* [65], *CCHFV* [1], *Mayaro virus* [66], *Human papillomavirus* (*HPV*) [67] and much more. The results confirmed that both B- and T-cell epitopes predicted by our proposed method have high efficacy in term of antigenicity and affinity to the receptor in comparison with the predicted epitopes by the standard epitope prediction tools. As a final evaluation, the predicted epitopes by our proposed method were investigated for probable humoral and cellular immune responses via an immune system simulation analysis. The obtained results demonstrated that the screened epitopes can provoke the appropriate immune responses and could be considered as safe and high immunogenic epitopes for more in vitro and in vivo evaluations.

Based on the results of the present study, the introduced method has high accuracy for predicting high immunogenic and safe B- and T-cell epitopes from different serotypes of *CCHFV* proteome. This pipeline can have been used for developing an automated one-click epitope prediction tools such as an online webserver or offline software. Furthermore, due to the serious threat to public health by pathogens with high levels of genetic variation such as *HIV*, *HCV*, Influenza virus and *dengue virus* this method can also be used for developing epitope based vaccine.

5. Conclusion

This study was planned for introducing a novel pipeline for predicting linear B- and T-cell epitopes from different serotypes of *CCHFV* proteome. To this end, we collected two sets of experimentally approved and predicted B- and T-cell epitopes as well as non-epitope

peptides from IEDB database and previously reported studies as positive and negative datasets respectively. After that, different physicochemical and compositional properties of the peptides were expressed as numerical vectors for classifying by machine learning algorithms. The results showed that the trained ANN algorithm could classify the datasets with highest accuracy. Furthermore, for evaluating the performance of the proposed method an epitope prediction challenge was done to a random peptide library from *CCHFV* envelopment polyprotein. Finally, the efficacy of this method in prediction of high immunogenic and safe B- and T-cell epitope was evaluated by comparison between the antigenicity and molecular docking scores of predicted epitopes by our proposed method and some standard epitope prediction tools. The results confirmed that this method could predict more effective epitopes than conventional methods and also can be used for developing an automated epitope based vaccine design tool.

Acknowledgment

The authors would like to acknowledge the University of Isfahan for the financial support of the study.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.intimp.2019.106020.

References

[1] M. Nosrati, M. Behbahani, H. Mohabatkar, Towards the first multi-epitope recombinant vaccine against Crimean-Congo hemorrhagic fever virus: a computer-aided vaccine design approach, J. Biomed. Inform. 93 (2019) 103160.
[2] S.D. Dowall, M.W. Carroll, R. Hewson, Development of vaccines against Crimean-Congo haemorrhagic fever virus, Vaccine 35 (44) (2017) 6015–6023.
[3] M.U. Mirza, et al., In silico structural elucidation of RNA-dependent RNA polymerase towards the identification of potential Crimean-Congo Hemorrhagic Fever Virus inhibitors, Sci. Rep. 9 (1) (2019) 6809.
[4] S. Shayan, et al., Crimean-Congo hemorrhagic fever, Lab. Med. 46 (3) (2015) 180–189.
[5] M. Mousavi-Jazi, et al., Healthy individuals' immune response to the Bulgarian Crimean-Congo hemorrhagic fever virus vaccine, Vaccine 30 (44) (2012) 6225–6229.
[6] M. Zivcec, et al., Molecular insights into Crimean-Congo hemorrhagic fever virus, Viruses 8 (4) (2016) 106.
[7] A.A. Kraus, A. Mirazimi, Molecular biology and pathogenesis of Crimean-Congo hemorrhagic fever virus, Future Virol. 5 (4) (2010) 469–479.
[8] M.E. Lindquist, et al., Exploring Crimean-Congo hemorrhagic fever virus-induced hepatic injury using antibody-mediated type I interferon blockade in mice, J. Virol. 92 (21) (2018) e01083–18.
[9] D.A. Bente, et al., Pathogenesis and immune response of Crimean-Congo hemorrhagic fever virus in a STAT-1 knockout mouse model, J. Virol. 84 (21) (2010) 11089–11100.
[10] J. Hinkula, 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 (10) (2017) e02076–16.
[11] K. Spik, 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 (21) (2006) 4657–4666.
[12] S.M. Ghiasi, et al., Mice orally immunized with a transgenic plant expressing the glycoprotein of Crimean-Congo hemorrhagic fever virus, Clin. Vaccine Immunol. 18 (12) (2011) 2031–2037.
[13] K.R. Buttigieg, et al., A novel vaccine against Crimean-Congo Haemorrhagic Fever protects 100% of animals against lethal challenge in a mouse model, PloS One 9 (3) (2014) e91516.
[14] A.R. Garrison, et al., A DNA vaccine for Crimean-Congo hemorrhagic fever protects against disease and death in two lethal mouse models, PLoS Neglected Trop. Dis. 11 (9) (2017) e0005908.
[15] C.B. Palatnik-de-Sousa, I.d.S. Soares, D.S. Rosa, Epitope discovery and synthetic vaccine design, Front. Immunol. 9 (2018) 826.
[16] A. Sette, J. Fikes, Epitope-based vaccines: an update on epitope identification, vaccine design and delivery, Curr. Opin. Immunol. 15 (4) (2003) 461–470.
[17] S. Smith-Gill, Protein epitopes: functional vs. structural definitions, Res. Immunol. 145 (1) (1994) 67–70.
[18] M. Luštrek, et al., Epitope predictions indicate the presence of two distinct types of epitope-antibody-reactivities determined by epitope profiling of intravenous immunoglobulins, PloS One 8 (11) (2013) e78605.
[19] B.N. Sobolev, et al., Computer design of vaccines: approaches, software tools and informational resources, Curr. Comput. Aided Drug Des. 1 (2) (2005) 207–222.
[20] L. He, J. Zhu, Computational tools for epitope vaccine design and evaluation, Curr. Opin. Virol. 11 (2015) 103–112.
[21] M. Nosrati, H. Mohabatkar, M. Behbahani, A novel multi-epitope vaccine for cross protection against Hepatitis C Virus (HCV): an immunoinformatics approach, Res. Mol. Med. 5 (1) (2017) 17–26.
[22] K. Xu, et al., Epitope-based vaccine design yields fusion peptide-directed antibodies that neutralize diverse strains of HIV-1, Nat. Med. 24 (6) (2018) 857.
[23] B. Zhou, et al., Multiple linear B-cell epitopes of classical swine fever virus glycoprotein E2 expressed in E. coli as multiple epitope vaccine induces a protective immune response, Virol. J. 8 (1) (2011) 378.
[24] T. Ben-Yedidia, R. Arnon, Epitope-based vaccine against influenza, Expert Rev. Vaccine 6 (6) (2007) 939–948.
[25] S. Sabetian, et al., Exploring dengue proteome to design an effective epitope-based vaccine against dengue virus, J. Biomol. Struct. Dyn. 37 (10) (2019) 2546–2563.
[26] V.H. Urrutia-Baca, et al., Immunoinformatics approach to design a novel epitope-based oral vaccine against Helicobacter pylori, J. Comput. Biol. (2019).
[27] K.A. Rahjerdi, et al., Designing and structure evaluation of multi-epitope vaccine against ETEC and EHEC, an in silico approach, Protein Peptide Lett. 23 (1) (2016) 33–42.
[28] N. Hajighahramani, et al., Computational design of a chimeric epitope-based vaccine to protect against Staphylococcus aureus infections, Mol. Cellular Probes (2019).
[29] B. Mahajan, et al., Multiple antigen peptide vaccines against Plasmodium falciparum malaria, Infection Immunity 78 (11) (2010) 4613–4624.
[30] M. Shaddel, M. Ebrahimi, M.R. Tabandeh, Bioinformatics analysis of single and multi-hybrid epitopes of GRA-1, GRA-4, GRA-6 and GRA-7 proteins to improve DNA vaccine design against Toxoplasma gondii, J. Parasitic Dis. 42 (2) (2018) 269–276.
[31] R.E. Soria-Guerra, et al., An overview of bioinformatics tools for epitope prediction: implications on vaccine development, J. Biomed. Info. 53 (2015) 405–414.
[32] K. Marciniuk, B. Trost, S. Napper, EpIC: a rational pipeline for epitope immunogenicity characterization, Bioinformatics 31 (14) (2015) 2388–2390.
[33] H. Singh, H.R. Ansari, G.P. Raghava, Improved method for linear B-cell epitope prediction using antigen's primary sequence, PloS One 8 (5) (2013) e62216.
[34] P. Sun, et al., Advances in in-silico b-cell epitope prediction, Curr. Top. Med. Chem. 19 (2) (2019) 105–115.
[35] P. Du, S. Gu, Y. Jiao, PseAAC-General: fast building various modes of general form of Chou's pseudo-amino acid composition for large-scale protein datasets, Int. J. Mol. Sci. 15 (3) (2014) 3495–3506.
[36] L.N. Adler, et al., The other function: class II-restricted antigen presentation by B cells, Front. Immunol. 8 (2017) 319.
[37] P. Oyston, K. Robinson, The current challenges for vaccine development, J. Med. Microbiol. 61 (7) (2012) 889–894.
[38] X. Qiu, V.R. Duvvuri, J. Bahl, Computational approaches and challenges to developing universal influenza vaccines, Vaccines 7 (2) (2019) 45.
[39] D.V. Desai, U. Kulkarni-Kale, T-cell epitope prediction methods: an overview, Immunoinformatics, Springer, 2014, pp. 333–364.
[40] S.K. Dhanda, P. Vir, G.P. Raghava, Designing of interferon-gamma inducing MHC class-II binders, Biol. Direct 8 (1) (2013) 30.
[41] M. Bhasin, S. Lata, G. Raghava, TAPPred prediction of TAP-binding peptides in antigens, Immunoinformatics, Springer, 2007, pp. 381–386.
[42] I.A. Doytchinova, D.R. Flower, VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines, BMC Bioinform. 8 (1) (2007) 4.
[43] S. Lata, M. Bhasin, G.P. Raghava, Application of machine learning techniques in predicting MHC binders, Immunoinformatics, Springer, 2007, pp. 201–215.
[44] N. Tomar, R.K. De, Immunoinformatics: an integrated scenario, Immunology 131 (2) (2010) 153–168.
[45] H. Singh, G. Raghava, ProPred1: prediction of promiscuous MHC Class-I binding sites, Bioinformatics 19 (8) (2003) 1009–1014.
[46] B. Manavalan, et al., iBCE-EL: a new ensemble learning framework for improved linear B-cell epitope prediction, Front. Immunol. 9 (2018) 1695.
[47] P. Sun, et al., Epitope prediction based on random peptide library screening: benchmark dataset and prediction tools evaluation, Molecules 16 (6) (2011) 4971–4993.
[48] H.-W. Wang, T.-W. Pai, Machine learning-based methods for prediction of linear B-cell epitopes, Immunoinformatics, Springer, 2014, pp. 217–236.
[49] N.D. Rubinstein, I. Mayrose, T. Pupko, A machine-learning approach for predicting B-cell epitopes, Mol. Immunol. 46 (5) (2009) 840–847.
[50] Y. Zhao, et al., Application of support vector machines for T-cell epitopes prediction, Bioinformatics 19 (15) (2003) 1978–1984.
[51] M. Bhasin, G. Raghava, Prediction of CTL epitopes using QM, SVM and ANN techniques, Vaccine 22 (23–24) (2004) 3195–3204.
[52] C. Lundegaard, O. Lund, M. Nielsen, Prediction of epitopes using neural network based methods, J. Immunol. Meth. 374 (1–2) (2011) 26–34.
[53] M.C. Honeyman, et al., Neural network-based prediction of candidate T-cell epitopes, Nat. Biotechnol. 16 (10) (1998) 966.
[54] J. Lara, et al., Artificial neural network for prediction of antigenic activity for a major conformational epitope in the hepatitis C virus NS3 protein, Bioinformatics 24 (17) (2008) 1858–1864.
[55] A.L. Tarca, et al., Machine learning and its applications to biology, PLoS Comput. Biol. 3 (6) (2007) e116.
[56] K. Lai, et al., Artificial intelligence and machine learning in bioinformatics, Encycl. Bioinfo. Comput. Biol.: ABC of Bioinfo. 55 (2018) 272.
[57] B. Liu, et al., Pse-in-One: a web server for generating various modes of pseudo components of DNA, RNA, and protein sequences, Nucleic Acids Res. 43 (W1) (2015) W65–W71.

[58] K.-C. Chou, Pseudo amino acid composition and its applications in bioinformatics, proteomics and system biology, Curr. Proteomics 6 (4) (2009) 262–274.
[59] M. Behbahani, H. Mohabatkar, M. Nosrati, Analysis and comparison of lignin peroxidases between fungi and bacteria using three different modes of Chou's general pseudo amino acid composition, J. Theor. Biol. 411 (2016) 1–5.
[60] M.S. Rahman, et al., DPP-PseAAC: a DNA-binding protein prediction model using Chou's general PseAAC, J. Theor. Biol. 452 (2018) 22–34.
[61] K.-C. Chou, X. Cheng, X. Xiao, pLoc_bal-mHum: predict subcellular localization of human proteins by PseAAC and quasi-balancing training dataset, Genomics (2018).
[62] J. Jia, et al., iPPI-PseAAC (CGR): Identify protein-protein interactions by incorporating chaos game representation into PseAAC, J. Theor. Biol. 460 (2019) 195–203.
[63] Z. Hajisharifi, et al., Predicting anticancer peptides with Chou′ s pseudo amino acid composition and investigating their mutagenicity via Ames test, J. Theor. Biol. 341 (2014) 34–40.
[64] K.C. Chou, Prediction of protein cellular attributes using pseudo-amino acid composition, Proteins Struct. Funct. Bioinf. 43 (3) (2001) 246–255.
[65] M. Ali, et al., Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection, Sci. Rep. 7 (1) (2017) 9232.
[66] S. Khan, et al., Immunoinformatics and structural vaccinology driven prediction of multi-epitope vaccine against Mayaro virus and validation through in-silico expression, Infect. Genet. Evol. (2019).
[67] S.S. Mohamed, et al., Immunoinformatics approach for designing epitope-based peptides vaccine of L1 major capsid protein against HPV Type 16, Int. J. Multidisc. Curr. Res. (2016).