

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Predicting Candidate Epitopes on Ebola Virus for Possible Vaccine Development

Maryam Hemmati, Ehsan Raoufi and Hossein Fallahi

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.72413>

Abstract

Zaire ebolavirus, a member of family *Filoviridae* is the cause of hemorrhagic fever. Due to lack of appropriate antiviral or vaccine, this disease is very lethal. In this study, we tried to find epitopes for superficial glycoprotein and nucleoprotein of *Zaire ebolavirus* (that have high antigenicity for MHC I, II and B cells) by using *in silico* methods and immunoinformatics approach. By using CTLPred, SYFPEITHI and ProPred web applications for MHC class I and SYFPEITHI and ProPred1 web applications for MHC class II, we had been able to find epitopes (peptides) that have the highest score. Also ElliPro, IgPred and DiscoTope web tools had been performed to predict B cells conformational epitopes. Linear epitope prediction for B cell was performed with six methods from IEDB. All of the results that including candidate epitopes for T cells and B cells were reported. It was expected that these peptides could be stimulated immune response and used for designing the multi-peptide vaccine against ZEV but these results should be reliable with experimental analysis.

Keywords: epitopes, glycoprotein, immune response, immunoinformatics approach, multi-peptide vaccine, nucleoprotein, *Zaire ebolavirus*

1. Introduction

Zaire ebolavirus, a member of genus *Ebola virus*, family *Filoviridae*, and order *Mononegavirales*, is enveloped, RNA negative strand genome and filamentous virus. This virus is the cause of serious hemorrhagic fever (HF) in human and the mortality rate is 50–90% [1, 2]. There have been several outbreaks of *Ebola virus* up to now and the last one was on 2014. This outbreak started in Guinea and spread to other countries in West Africa. The 2014 outbreak was the largest one which had the highest mortality rate and risk of spreading to different parts of the world [3, 4].

Experimental work with EBOV is very difficult and currently, there is no effective and licensed vaccine available for Ebola. Vaccination is a good approach for the prevention and treatment against many types of diseases like the viral infection. There are several types of vaccines, and the goal of all of them is presenting antigenic fragment to the immune system to induce the adaptive immune response. Several types of antiviral vaccines consist of inactivated, live attenuated, virus-like particles and DNA vaccines. Another type of vaccine is peptide-based vaccines. These vaccines are based on designing different epitopes for B cells and T cells. These vaccines, in comparison with other vaccines, have fewer side effects and are safe and easy to prepare [4–6]. Obviously, T cells have an important role in stimulating immune responses, but for having a better and accurate response, first of all, antigenic fragments should be attached to major histocompatibility complex (MHC) molecules. MHC molecules process and present antigenic peptides to T cells. These peptide epitopes must be linear for attaching to MHC molecules [7]. The most of T cells are belonging to two groups namely CD8+ and CD4+. The difference between two groups refers to different glycoproteins in the surface of T cells. CD8+ T cells are cytotoxic T cells (CTL) that bind to MHC class I molecules and CD4+ T cells are T helper cells that attach to MHC class II [8, 9]. B cells are another part of immune systems, which have receptors and secrete antibodies. Epitopes can be discontinuous or continuous for B cells. These epitopes can also bind to lipids, carbohydrates and peptides but 3D structure of antigens has an important role in stimulating humoral immune response by B cells [10]. With the advances in in silico method and bioinformatics, immunoinformatics, a branch of science was progressed. Immunoinformatics is an interdisciplinary science that emanates from immunology and bioinformatics and generates meaningful immunological data. With the help of immunoinformatics approaches, we can find epitopes for B cells and T cells, and also we can design vaccines based on peptide or multi-peptide vaccine [4, 11].

In this chapter, we tried to use immunoinformatics tools and in silico method to predict MHCs linear epitopes and B cells discontinuous and linear epitopes (peptides) for glycoprotein and nucleoprotein of *Zaire ebolavirus* (as antigen). These results can be useful for finding epitopes that can be the candidate for designing vaccine and therapeutic strategies for fighting with HF (hemorrhagic fever) of Ebola virus.

2. The best candidate for producing vaccines against *Ebola virus*

The ZEOBV genome has seven ORF, including NP-VP35-VP40-GP-VP30-VP24-L. Nucleoprotein (NP) causes encapsulation of the genome of Ebola. It has been aggregated with VP30 (transcription factor) and VP35 (polymerase cofactor). L is RNA-dependent RNA polymerase. VP24 is a minor matrix protein that associates with the membrane. VP40 is a major matrix protein that can mediate virus particles creation [1, 4, 12]. Glycoprotein (GP) is present on the surface of Ebola. Secreted nonstructural GP, structural GP 1 and GP 2 are the results of mRNA editing during transcription of GP of ZEBOV. The N-terminal sequence between the GP and sGP is the same but the C-terminal of them is different [4, 13]. The ratio between sGP and GP during infection is 80–20%. Superficial GP is cleaved by furin in Golgi to GP1 and GP2 that form homotrimeric proteins on the surface of EBOV

[5, 10]. The GP1 is a soluble protein on the surface of virus, which has a mucin-like domain and is highly glycosylated. GP2 is a membrane-spanning subunit and smaller than GP1, which is connected to GP1 by disulfide bonds [14]. GP has a major role in attaching to host cells and has cytotoxicity effects. This polycistronic GP is the main difference between ZEBOV and other mononegavirus. Because of the GP on the surface, this protein has the most antigenicity for the immune system and so it can be the best candidate for producing vaccines against EBOV [2, 4]. Immunoinformatic studies on Ebola virus proteins also have shown that there is a high rate of immune response for epitopes of Ebola nucleoprotein (NP), in which, it can be even more than the response for epitopes of Ebola glycoprotein [15–18]. Therefore, it can be another target for peptide-based vaccine design against Ebola virus.

Bioinformatics evaluation on the superficial glycoprotein showed less than 30% identity with family and less than 50% with species of identity in the this protein, this glycoprotein is also glycosylated. Studies have been shown that the structure and sequencing of GP are constantly changing and these changes are the major causes of the weakness of the immune system against the virus and, consequently, its pathogenicity. For these reasons, designing and development of the vaccine against the Ebola virus is difficult and complex.

3. Epitope mapping for the purpose of peptide-based vaccine design

The most commonly used assimilation and the concept of the term epitope mapping is the mapping of the antigenic regions detected by antibodies. This term is ambiguous because the purpose of using this term in research is not clear, and this mapping can be done with a variety of goals. Therefore, researchers have tried to use alternative terms for epitope mapping in articles in order to clarify the purpose of this bioinformatics process such as immunological analysis, microstructure analysis and epitope determination. Despite all these efforts and including a wide range of goals, scientists are still advised to use this term in articles and resources.

A number of important applications of epitope mapping including:

- Determination of biological process mechanism,
- Recognition of an epitope of practical value,
- Connection of any type of polymorphism such as SNP to protein structure or Ab binding,
- Characterization of Ab binding in patients,
- Evaluation of vaccine design,
- Identification of autoimmune diseases,
- Qualification of an Ab for diagnostic use such as Western blot analysis, trans-species assays, finding isoforms of Ag and allergen characterization,
- Distinguishing of antigen peptide mimic,

- Establishing of antigen structure and
- Finding an antigen which it can make different between immunization antibodies and infection antibodies.

Immunogenicity studies have shown epitope mapping (the branch of bioinformatics) is based on mathematical modeling and evolutionary algorithms. Epitope mapping methods can be used to molecular modeling behaviors in nature, so the data obtained from these methods can predict molecular interactions, such as binding of antigen to antibody and peptide to MHC, with high probability [19, 20].

Designing of vaccines based on peptides or epitopes is the goal that we have been considering in this study. They are within the epitope mapping compass and to achieve these goals, we identify and predict the epitopic areas by using approved and high-performance software.

Mathematically modeled methods for epitope mapping are artificial neural networks (ANN), quantization matrices, decision trees, HMM (Markov secret models), SMM is a stabilized matrix method (SMM), among which ANN, SVM and HMM are capable of analyzing linear and nonlinear data. In many applications that analyze epitope prediction, these three methods are used to identify the sequential linear epitopes of the T lymphocytes and the nonlinear (spatial) epitopes of the B lymphocytes [4, 21, 22].

There are other predictive methods for detecting spatial epitopes of the B cell lymphocytes, which include the following: homology modeling, docking, 3D- and threading techniques. These methods with the ANN, SMM and HMM methods are most used in the study of the spatial epitope and are among the software main methods used in this research [4, 23]. In sum, all methods for identifying epitopes that have the best antigenic properties but the epitopes should have important characteristics, including (1) structural flexibility, (2) in the surface of the protein, (3) exposure to solvent, (4) containing charged amino acids and (5) contains hydrophilic amino acids.

4. Candidate for designing peptide-based vaccines against *Ebolavirus*

Conventional vaccines are prepared from an attenuated or inactive version of the pathogen. However, often the antigen to which the immune system responds is of small number of amino acids or peptide of antigen. The alternative approach to stimulate the immune response, such as humoral and cell-mediated immune response, is the identification of peptide sequences or epitopes that have a protective immune response. Therefore, these epitopes can have no risk of mutation and they would be more stable, and also, there can be fewer side effects [24, 25]. This approach is called peptide-based vaccine designing.

4.1. Determine the T cell antigenic fragment

Epitope analysis was performed for MHC classes I and II.

4.1.1. Epitopes analysis for MHC class I

By using CTLPred (<http://www.imtech.res.in/raghava/ctlpred/index.html>), SYFPEITHI (<http://www.syfpeithi.de/bin/MHCServer.dll/EpitopePrediction.htm>) and ProPred1 (<http://www.imtech.res.in/raghava/propred/>) web tools, we had been able to find some peptidic epitopes for MHC class I.

- A. Finding epitopes for MHC class I by using CTLPred web tool and sequence of GP and NP of ZEBOV. In **Tables 1** and **2**, three peptides that have higher score are shown for each protein. Positions 19, 245 and 130 for GP and positions 292, 263 [4] and 397 for NP have higher scores, so these peptides have been suggested by the application to be epitopes.
- B. Using SYFPEITHI web tool for predicting epitopes. We investigated some alleles of the MHCI for the protein of interest. According to these results, as is shown in **Table 3**, for GP 564, 246 and 205 positions have a higher score in this prediction web tools [4]. But when we investigated other alleles, we had observed that the position 246 was the most repetitive of all and had a high score. For NP, results are shown in **Table 4**, position 266 has the highest score.
- C. ProPred1 was used to predict epitopes of GP and NP, which bind with the highest score to MHC class I alleles. The peptide with highest score was selected and is shown in

Strat position	Sequence	Score(ANN\SVM)
130	RGFPRCRYV	0.97/0.99110928
245	ESRFTPQFL	0.98/0.867525
19	FFLWVILF	0.97/078863896

Table 1. Epitope prediction of GP based on CTLPred web server.

Strat position	Sequence	Score(ANN\SVM)
292	EYAPFARLL	0.96/1.5915619
263	RLHPLARTA	0.89/1.2678105
397	LRKERRLAKL	0.84/1.0652382

Table 2. Epitope prediction of NP based on CTLPred web server.

Allele	Sequence	Position	score
HLA-A*26	ETTQALQLF	564	31
HLA-A*26	GVIIAVIAL	660	28
HLA-A*01	ATEDPSSGY	205	30
HLA-A*01	LFEVDNLTY	233	27
HLA-B*08	TRKIRSEEL	298	29
HLA-B*1402	SRFTPQFLL	246	30
HLA-B*1402	NRKAIDFLL	586	27
HLA-B*1516	YFGPAAEGI	534	27
HLA-B*2705	LRTFSILNR	579	27
HLA-B*37	RDRFKRTSF	11	27
HLA-B*40:01	NETTQALQL	563	27
HLA-B*58:02	RATTELRTF	574	29
HLA-A*02:01	GLICGLRQL	553	29
HLA-A*03	TVIYRGTTF	168	27
HLA-A*03	FLILPQAKK	183	27
HLA-A*03	QIIHDFVDK	625	27
HLA-A*11:01	STHNTPVYK	387	28

Table 3. Epitopes with scores above 27 of GP for MHC class I according to SYFPEITHI web server.

Tables 5 and 6. According to this result, the highest score belonged to position 264 for the HLA-B*2705 allele for GP, and as shown in **Table 6** for NP, positions 273 and 109 for HLA-A20 allele have higher scores.

Finally with revision of all data that are achieved for MHC class I, we are able to conclude that sequences: “TRKIRSEEL” (with position 298 for T residue), “SRFTPQFLL” (with position 246 for S residue) and “ETTQALQLF” (with position 561 for E residue) for GP [4], and sequences: “RLHPLARTA” (with position 263 for first residue), “SRELDHLGL” (with position 360 for

Allele	Sequence	Position	score
HLA-A*01	TSDGKEYTY	680	29
HLA-A*01	YPDSLEEEY	688	28
HLA-A*03	PLARTAKVK	266	31
HLA-A*03	PVYRDHSEK	609	30
HLA-A*03	TVLDHILQK	249	29
HLA-A*03	EVKKRDGVK	107	28
HLA-A*03	TLRKERLAK	396	28
HLA-A*03	TVAPPAPVY	603	27
HLA-A*1101	TVLDHILQK	249	28
HLA-A*26	EVNSFKAAL	276	30
HLA-B*08	LRKERRLAKL	397	30
HLA-B*08	NHKNKFMAI	726	27
HLA-B*1402	ARFSGLLIV	239	30
HLA-B*1402	SRELDHLGL	360	29
HLA-B*4001	NEENRFVTL	707	27
HLA-B*0201	RLEELLPAV	116	28
HLA-B*0201	GLFPQLSAI	311	27

Table 4. Epitopes with scores above 27 of NP for MHC class I according to SYFPEITHI web server.

first residue) and “VKNEVNSFK” (with position 273 for first residue), have the highest score and the most frequent within these analysis. Therefore, it can be a better candidate peptide epitopes than other sequences.

4.1.2. Epitopes analysis for MHC class II

By using SYFPEITHI and ProPred web tools, some peptidic epitopes for MHC class II were predicted.

Allele	Sequence	Position	Real score
HLA-B*2705	TRKIRSEEL	298	10000
HLA-B*2705	GREAAVSHL	357	9000
HLA-B*2705	FQRTFSIPL	27	6000

Table 5. Epitope prediction from GP for MHC class I by ProPred1 web tools (only top score peptides are shown).

Allele	Sequence	Position	score
HLA-A20 cattle	VKNEVNSFK	273	4000
HLA-A20 cattle	KKRDGVKRL	109	4000
HLA-B2705	SRELDHLGL	360	3000

Table 6. NP epitope prediction of MHC class I according to ProPred1 web server.

Allele	Sequence	Position	score
HLA_DRB1*0101	CRYVHKVSGTGPCAG	135	35
HLA_DRB1*0101	FFLYDRLASTVIYRG	159	33
HLA_DRB1*0701	IILFQRTFSIPLGVI	24	34
HLA_DRB1*0101	LRTFSILNRKAIDFL	579	30

Table 7. GP epitope prediction for MHC class II based on SYFPEITHI web server.

- A. Using SYFPEITHI web tools, epitope mapping of the amino acid sequence of the GP and NP of ZEBOV was performed. Accordingly, we investigated some alleles of MHC class II for the protein of interest. Peptides that have higher score were selected and are shown in **Table 7** for GP and in **Table 8** for NP. According to this result for GP, the highest score belongs to position 135 with the score of 35 and after that, the position of 159 has a higher

Allele	Sequence	Position	score
DRB1*0301	NRFVTLDGQQFYWPV	710	38
DRB1*0101	SGAVKYLEGHGFRFE	93	36
DRB1*0101	ENRFVTLDGQQFYWP	709	35
DRB1*0101	DMDYHKILTAGLSVQ	18	34
DRB1*0101	HGLFPQLSAIALGVA	310	34
DRB1*0101	MVIFRLMRTNFIKF	198	33
DRB1*0101	HQGMHMOVAGHDANDA	216	31
DRB1*0101	RHILRSQGPDAVLY	653	31
DRB1*0701	DMDYHKILTAGLSVQ	18	32
DRB1*0107	GVDFQESADSFLML	63	30
DRB1*1501	AGQFLSFASLFLPKL	147	32
DRB1*1501	GHMMVIFRLMRTNFL	195	30
DRB1*1501	TNFLKFLIHQGMH	206	30
DRB1*1101	EEMYRHILRSQGPFD	649	30
DRB1*1101	AVLYYHMMKDEPVVF	664	30

Table 8. NP epitope prediction for MHC class II based on SYFPEITHI web server.

score for the HLA-DRB1*0101 allele [4]. For NP, position 710 with score 38 have the highest score.

- B.** ProPred web tool was used to predict epitopes. Peptide with the highest score was selected and is shown in **Table 9** for GP and in **Table 10** for NP. This result for GP has

Allele	Sequence	Position	score
DRB1_0703	FQRTFSIPL	26	9.7000
DRB1_0817	FFLYDRLAS	158	7.2000
DRB1_0701	FQRTFSIPL	26	9.7000

Table 9. Epitope prediction of GP for MHC class II based on ProPred web server.

Allele	Sequence	Position	score
DRB1_0701	FLSFASLFL	149	7.6000
DRB1_0701	FRLMRTNFL	200	7.5200
DRB1_0701	LNLSGVNNL	299	7.2000
DRB1_0703	FLSFASLFL	149	7.6000
DRB1_0703	FRLMRTNFL	200	7.5200
DRB1_0703	LNLSGVNNL	299	7.2000
DRB1_0817	FLIKFLLIH	207	7.8000
DRB1_1501	MVIFRLMRT	197	7.7000
DRB1_1506	MVIFRLMRT	197	7.7000
DRB1_0405	FRLMRTNFL	200	7.6000

Table 10. MHC class II epitope prediction for NP of according to ProPred web server.

been shown that the positions 158 and 26 have the highest score. These positions are very frequent in reviewing other results. For NP positions, 149 and 200 have higher score according to ProPred web server.

As a result for MHC class II for GP, sequences: “IILFQRTFSIPLGVI” (with position 24 for first residue), “CRYVHKVSGTGPCAG” (with position 135 for the first residue) and “FFLYDTLAS” (with position 158 for first the residue) [4] and for NP sequences: “FLSFASLFL” (with position 149 for the first residue), “NRFVTLDGQQFYWPV” (with position 710 for first residue) and “FRLMRTNFL” (with position 200 for first residue) have the highest scores and the most frequent within these analysis.

By considering these results for MHC classes I and II, we think these epitopes can activate the cell-mediated immune response; therefore they can be used for producing peptide-based vaccines.

4.2. Determine the B cell antigenic fragment

4.2.1. Prediction of linear (sequential) epitopes

In this section, six methods from IEDB were used for the prediction of linear epitopes. A collection of methods to predict linear B cell epitopes based on sequence characteristics of the antigen using amino acid scales and HMMs. These methods include the following:

1. BepiPred Linear Epitope Prediction: BepiPred predicts the location of linear B cell epitopes using a combination of a HMM and a propensity scale method [26].
2. Chou & Fasman Beta-Turn Prediction: This method is commonly used to predict beta turns to the prediction of antibody epitopes.
3. Emini Surface Accessibility Prediction: The computation was based on surface accessibility scale on antibodies. The accessibility profile was achieved using the formulae $S_n = (n + 4 + i) (0.37) - 6$, where S_n is the surface probability, d_n is the fractional surface probability value, and i vary from 1 to 6. A hexapeptide sequence with S_n greater than 1.0 indicates an increased probability of existing on the surface [27].
4. Karplus & Schulz Flexibility Prediction: In this method, flexibility scale based on ability to move protein segments. The calculation based on a flexibility scale is similar to classical calculation, except that the center is the first amino acid of the six amino acids window length, and there are three scales for describing flexibility instead of a single one [28].
5. Kolaskar & Tongaonkar Antigenicity: A semiempirical method that makes use of physicochemical properties of amino acid residues and their frequencies of occurrence in experimentally known segmental epitopes was developed to predict antigenic peptide established on proteins. This method can predict antigenic peptide with about 75% accuracy [29].
6. Parker Hydrophilicity Prediction: In this method, hydrophilic scale based on peptide retention times during high-performance liquid chromatography (HPLC) on a reversed-phase column was formulated [30].

4.2.1.1. Linear epitopes prediction for GP

We performed all of six methods for the GP protein sequence and summarized its results in **Table 11**.

Method number	Threshold	Max score	The best candidate peptide sequences
1	0.271	2.51 for 384 position	370 to 394
2	1.007	1.437 for 384 position	381 to 387
3	1	7.532 for 499 position	Scattered areas
4	1.008	1.119 for 321 position	318 to 324
5	1.015	1.225 for 183 position	180 to 186
6	1.793	7.957 for 640 position	637 to 643

Table 11. Conclusion results for B cell linear epitope prediction from GP protein sequence.

4.2.1.2. Linear epitopes prediction for NP

We performed all of six methods for the NP protein sequence and summarized its results in **Table 12**.

4.2.2. Prediction of discontinuous (conformational) epitopes

4.2.2.1. Discontinuous epitopes prediction for GP

For B cell epitope prediction, 3D structure of antigen is more important than linear sequence, therefore, in this study, we used PDB ID of GP of ZEBOV.

By considering the structure of GP in PDB, we comprehended that “I” chain of protein has a maximum length than other chains. Also, with investigating two epitopes that predicted for MHCs molecules, we understood both of them are on the “I” chain. Therefore, this chain was selected in this analysis.

- A. Using ElliPro web tool, epitope mapping of the amino acid sequence of the GP of ZEBOV was performed. PDB Id (3csy) and “I” chain was used for ElliPro tool. Results from this prediction for linear epitopes have been shown that the sequence of the peptides from 31 to 64 and 255 to 310 had a higher score than the other part of the protein and are illustrated in **Figures 1–3** and **Table 13** [4].
- B. By using DiscoTope and 3D structure of GP of ZEV, we could find discontinuous epitopes. These sequences may be near to each other in 3D conformation but far from each other in the amino acid sequence or the first structure. In this study, we analyze the “I” chain from GP1 and we set the threshold on -7.7 , upstream regions of the threshold have the positive prediction. Two regions that are shown in **Figure 4** have positive predictions but 261–310 regions have a more score and these scores for each amino acid are illustrated in **Table 14** [4].

In **Table 14**, contact number indicates the number of amino acids that are next to each amino acid. As much as the contact number is lower, it shows that our amino acids are

Method number	Thereshold	Max score	The best candidate peptide sequences
1	0.359	2.777 for 477 position	Scattered areas
2	0.991	1.511 for 437 position	434 to 440
3	1	5.323 for 371 position	369 to 374
4	1.007	1.141 for 507 and 508 positions	504 to 511
5	1.016	1.22 for 43 position	40 to 46
6	2.101	8.686 for 492 position	489 to 495

Table 12. Conclusion results for B cell linear epitope prediction from NP protein sequence.

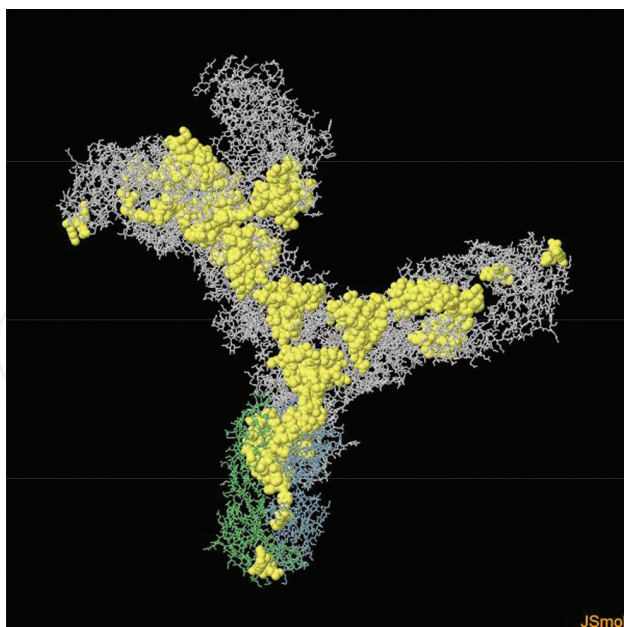


Figure 1. Discontinuous epitope(s) number one 3D structure for GP.

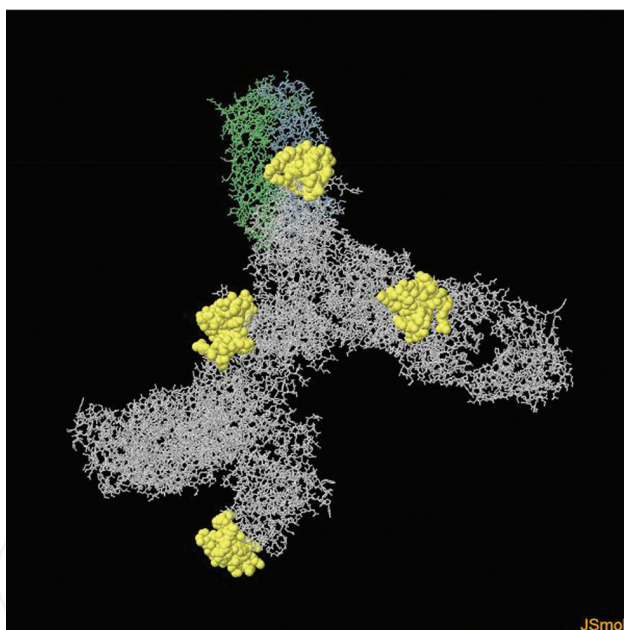


Figure 2. Discontinuous epitope(s) number two 3D structure for GP.

more external. For example, in this study, VAL residue in the 310th position had the lowest contact number and was more external.

- C. Using IgPred web tool, epitope mapping of the amino acid sequence of the GP of ZEBOV was performed. The sequence of the chain “I” of glycoprotein was selected to study the interaction with the antibodies IgG, IgE and IgA.

As illustrated in **Table 15**, different regions in this chain had a different score for IgG but the end of the sequence had the highest score and was approximately 255–310 regions.

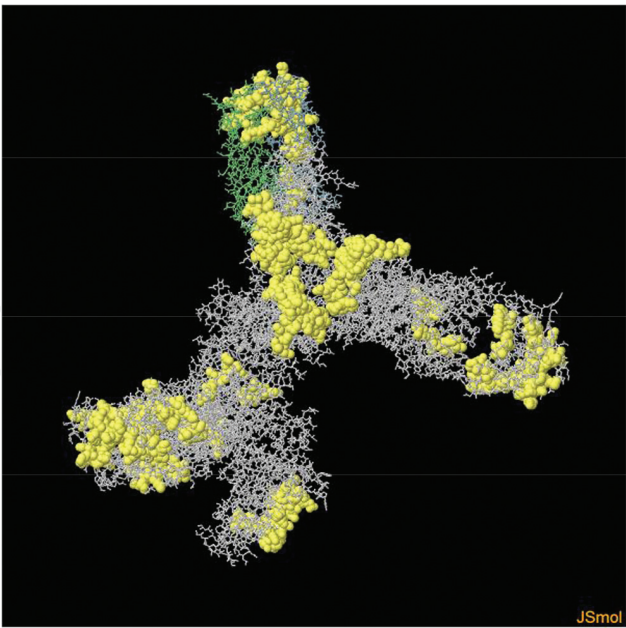


Figure 3. Discontinuous epitope(s) number three 3D structure for GP.

No.	Residues	Number of residues	Score
1	IR31, IS32, IY33, IP34, IL35, IG36, IV37, IY38, IH39, IN40, IS41, IV42, IL43, IQ44, IV45, IS46, ID47, IV48, ID49, IK50, IL51, IV52, IC53, IR54, ID55, IK56, IL57, IS58, IS59, IT60, IN61, IQ62, IL63, IR64, IE100, IH185, IL186, IP187, IQ188, IA189	40	0.796
2	IE235, IV236, ID237, IN238, IL239, IT240, IQ243, IF248, IF252, IQ255, IL256, IN257, IE258, IT259, ID260, IV261, IT262, IS263, IG264, IK265, IR266, IS267, IN268, IT269, IT270, IG271, IK272, IL273, IT274, IN275, IK276, IV277, IN278, IR299, IK300, IR301, IR302, IS303, IE304, IE305, IL306, IS307, IF308, IT309, IV310	45	0.701
3	IR89, IS90, IG91, IV92, IP93, IP94, IK95, IK114, IK115, IP116, ID117, IG118, IS119, IE120, IC121, IL122, IP123, IA124, IA125, IP126, ID127, IG128, IH129, IR130, IG131, IG145, IP146, IC147, IA148, IG149, ID150, IF151, ID163, IV169, IV171, IR172, IG173, IG224, IG226, IT227, IN228, IE229, IV230	43	0.614

Table 13. Three top scores predicted discontinuous epitope for B cell according to ElliPro web server.

Therefore, according to prediction with these web tools, 255–310 regions are the proper candidate for being epitope [4].

4.2.2.2. *Discontinuous epitopes prediction for NP*

No specific structure was found in the PDB for NP. Therefore, homology modeling of these types of proteins is needed to determine their structures. This goal has been achieved with the help of homology modeling, and its stages have been described below:

- A. Template Selection:** The template to be used in homology modeling should be based on target-template alignment, and the template that most closely resembles with our protein is selected as the template.
- B. Model Building:** The models are constructed using target-template alignment by ProMod3, and the areas that are shared and conserved between the template and the target are copied from the template on the model. The areas that are added or removed are rebuilt using fragment library. Side chains are then remodeled. Finally, its overall geometry is determined, but the loop areas are building with PROMOD-II.

Chain ID	Residue ID	Residue Name	Contact number	Disco Tope Score
I	261	TYR	13	-7.447
I	262	THR	11	-5.510
I	263	SER	11	-5.510
I	264	GLY	9	-2.952
I	265	LYS	15	-6.372
I	267	SER	15	-7.348
I	268	ASN	10	-3.780
I	269	THR	10	-4.544
I	270	THR	8	-3.423
I	271	GLY	13	-6.966
I	277	VAL	10	-5.284
I	278	ASN	8	-3.870
I	299	ARG	7	-2.755
I	300	LYS	5	-0.883
I	301	ILE	11	-5.702
I	302	ARG	8	-2.804
I	303	SER	12	-4.847
I	304	GLU	10	-4.323
I	306	LEU	11	-7.687
I	309	THR	7	-6.575
I	310	VAL	3	-3.807

Table 14. The predicted epitope sequence from 261 to 310 discontinuous amino acids for B cell according to DiscoTope web server.

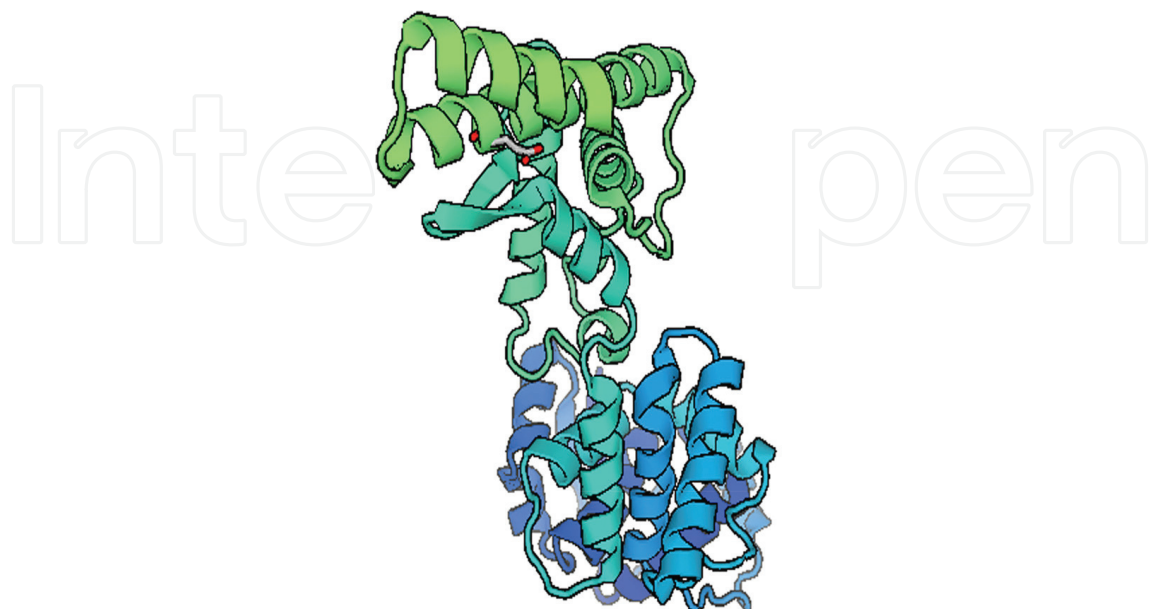


Figure 4. The model 3D structure.

Peptide seq	IgG Score	IgE Score	IgA Score	Prediction
YPYDVPDYAIEGRGARSIPLGVIHNSVLQ VSDVDKLVCRDKLSSTNQLRSVGLNLEG NGVATDVPSATKRWCFRSGVPPK	1.215	-0.148	0.546	IgG Epitope
VVNYEAGEWAENCYNLEIKKPDGSECLP AAPDGIKGFPRCRYVHKVSGTGPCAGDF AFHKEGAFFLYDRLASTVIYRGTT	1.019	-0.029	0.380	IgG Epitope
FAEGVVAFLILPQAKKDFSSHPLREPVN ATEDPSSGYSTTIRYQATGFGTNEVEYL FEVDNLTYVQLESRFTPQFLQ	0.921	-0.238	0.459	IgG Epitope
LNETHYTSGKRSNTTCKLIWKVNPEIDTT IGEWAFWETKKNLTRKIRSELSFTVVT HHQDTGEESASSGKGLIITNTIA	1.518	-0.844	1.009	IgG Epitope
GVAGLITGGRTRR	0.691	-0.912	-0.156	Non-Epitope

Table 15. Antibodies scores for sequence of “I” chain of GP according to IgPred web server.

- C. **Results:** The SWISS-MODEL template library (SMTL version 2017-09-21, PDB release 2017-09-15) was searched with BLAST and HHblits for evolutionarily related structures matching the target sequence is presented in **Table 16** and the model 3D structure is also shown in **Figure 4**.
- D. After building homology modeling, ElliPro web tool was performed to predict conformational epitopes of NP. These results are shown in **Table 17** and (**Figures 5** and **6**).

Template	Seq Identity	Oligo-state	Found by	Method	Resolution	Seq Similarity	Range	Coverage	Description
4ypi.2.A	100.00	hetero-oligomer	HHblits	X-ray	3.71Å	0.60	39 - 384	0.47	Nucleoprotein

Table 16. Details on the template search.

No.	Residues	Number of residues	Score
1	A:H265, A:P266, A:L267, A:A268, A:R269, A:T270, A:A271, A:K272, A:V273, A:K274, A:N275, A:E276, A:V277, A:N278, A:S279, A:K281, A:A282, A:S285, A:K289, A:H290, A:G311, A:F313, A:P314, A:Q315, A:G333, A:V334, A:N335, A:Y336, A:G337, A:E338, A:Q339, A:Y340, A:Q341, A:Q342, A:L343, A:R344, A:E345, A:A346, A:A347, A:T348, A:E349, A:A350, A:K352, A:Q353, A:L354, A:Q355, A:Q356, A:Y357, A:A358, A:E359, A:S360, A:R361, A:E362, A:L363, A:D364, A:H365, A:L366, A:G367, A:L368, A:D369, A:D370, A:Q371, A:E372, A:K373, A:K374, A:I375, A:L376, A:M377, A:N378, A:F379, A:H380, A:Q381, A:K382, A:K383, A:N384	75	0.752
2	A:T258, A:E259, A:R260, A:G261	4	0.75

Table 17. Top scores predicted discontinuous epitope for B cell according to ElliPro web server.

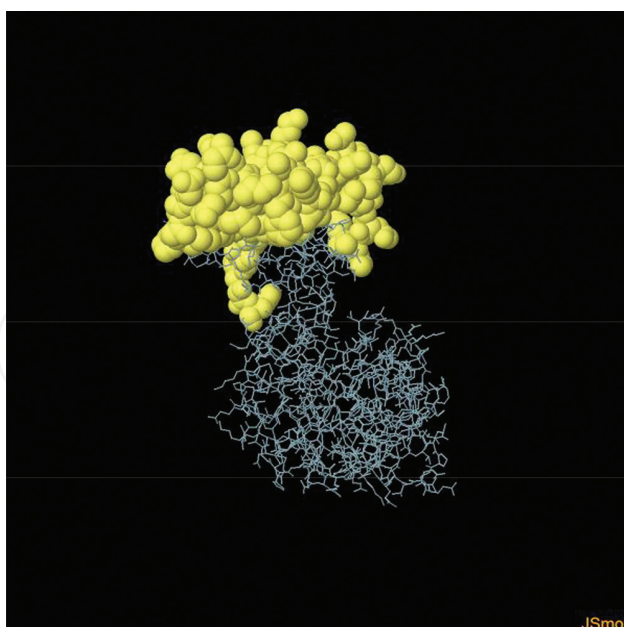


Figure 5. Discontinuous epitope(s) number one 3D structure for NP.

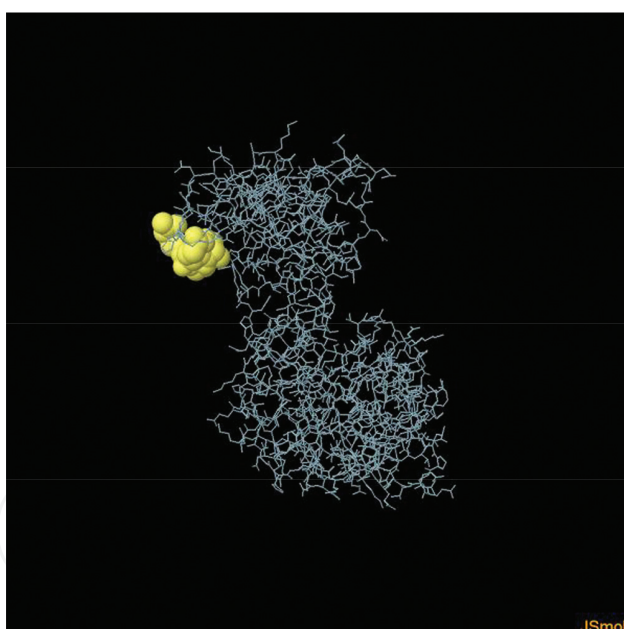


Figure 6. Discontinuous epitope(s) number two 3D structure for NP.

5. Conclusion

The hemorrhagic fever (HF) is a lethal disease from ZEBOV that caused the death of many people in Africa. The glycoprotein of ZEBOV is the only protein on the surface of the virus, and has severe cytotoxicity effects, also nucleoprotein can stimulate the immune response, stronger than GP, therefore, it was offered to design the vaccine against both of them [4, 13,

16, 17]. Vaccination is a good idea to prevent infection and limit the spreading of Ebola virus. So far, several vaccines against Ebola glycoprotein have been tested on animals, but no licensed vaccine has been reported in humans [31, 32]. There are several types of vaccine, in which one of them is a peptide vaccine, and it is the goal of this analysis. These vaccines are designed based on the epitopes fragment of antigens and they are safer and easy to prepare. For T cells, these epitopes are linear and they must bind to MHC I and MHC II at first, but for B cells, 3D conformation of antigens is very important so it can be discontinuous epitopes [6]. These discontinuous epitopes may be near to each other in the 3D conformation but far from each other in the amino acid sequence or first structure [4, 10]. The experimental work on this virus is very difficult and biosafety level is 4. With the help of immunoinformatics tools for vaccine design, the use of laboratory work is reduced, and hence we save more expense and time. Immunoinformatics approach can help us to predict T cell and B cell epitopes, and also have application in *in silico* vaccination. By predicting T cell and B cell epitopes, we can find the peptide that is useful for *in silico* vaccination or for designing multi-peptide vaccine [11]. By using ProPred, CTLPred, SYFPEITHI and ProPred1 web tools for predicting the epitopes for T cells, we had been able to introduce peptides that can be the candidate for designing multi-peptide vaccines. With the use of ElliPro, DiscoTope, IgPred web servers and linear epitope prediction methods from IEDB predicting epitopes for B cells, and these peptides can induce the immune response and design for the peptide-based vaccine. With the help of immunoinformatics tools for predicting epitopes for T cells and B cells, we can design the multi-peptide vaccine, and this vaccine can include both epitopes from GP and NP, which is useful for increasing immune response against Ebola virus. As conclusion for linear epitopes that bind to MHC I “TRKIRSEEL,” “SRFTPQFLL” and “ETTQALQLF” peptides for GP [4], and “RLHPLA-RTA,” “SRELDHLGL” and “VKNEVNSFK” peptides for NP are candidate epitopes. Also for MHC class II: “IILFQRTFSIPLGVI,” “CRYVHKVSGTGPCAG” and “FFLYDTLAS” peptides for GP and “FLSFASLFL,” “NRFVTLDGQQFYWPV” and “FRLMRTNFL” peptides for NP have the highest scores and the most frequent within this analysis. For B cell linear epitope prediction, results are shown in **Tables 11** and **12**. As a final result for conformational epitopes, we can only say, peptide sequence from 255 to 310 amino acids for GP has a higher score. These peptides are able to be the candidate for the vaccine against Ebola virus. It should be noted that these results just in *in silico* are valid and need laboratory (in vivo and in vitro) confirmation.

Author details

Maryam Hemmati^{1†}, Ehsan Raoufi^{1*†} and Hossein Fallahi²

*Address all correspondence to: ehsan.raoufi@gmail.com

1 Department of Medical Biotechnology, School of Allied medicine, Iran University of Medical Sciences, Tehran, Iran

2 Department of Biology, Faculty of Science, Razi University, Kermanshah, Iran

[†] These authors contributed equally.

References

- [1] Lee JE et al. Structure of the Ebola virus glycoprotein bound to a human survivor antibody. *Nature*. 2008;**454**(7201):177
- [2] Saeed MF et al. Cellular entry of Ebola virus involves uptake by a macropinocytosis-like mechanism and subsequent trafficking through early and late endosomes. *PLoS Pathogens*. 2010;**6**(9):e1001110
- [3] Gire SK et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science*. 2014;**345**(6202):1369-1372
- [4] Raoufi E, Hemmati M, EinAbadi H, Fallahi H. Predicting candidate epitopes on Ebolaviruse for possible vaccine development. In: *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. New York, NY, USA: ACM; Aug 2015. pp. 1083-1088
- [5] Chakraborty S. Ebola vaccine: Multiple peptide-epitope loaded vaccine formulation from proteome using reverse vaccinology approach. *International Journal of Pharmacy and Pharmaceutical Sciences*. 2014;**6**:407-412
- [6] Patronov A, Doytchinova I. T-cell epitope vaccine design by immunoinformatics. *Open Biology*. 2013;**3**(1):120139
- [7] Martin JE et al. A DNA vaccine for Ebola virus is safe and immunogenic in a phase I clinical trial. *Clinical and Vaccine Immunology*. 2006;**13**(11):1267-1277
- [8] Korber B, LaBute M, Yusim K. Immunoinformatics comes of age. *PLoS Computational Biology*. 2006;**2**(6):e71
- [9] Moise L, De Groot AS. Putting immunoinformatics to the test. *Nature Biotechnology*. 2006;**24**(7):791-792
- [10] Sharon J et al. Discovery of protective B-cell epitopes for development of antimicrobial vaccines and antibody therapeutics. *Immunology*. 2014;**142**(1):1-23
- [11] Tomar N, De Rajat K. Immunoinformatics: An integrated scenario. *Immunology*. 2010;**131**(2):153-168
- [12] Dowling W et al. Influences of glycosylation on antigenicity, immunogenicity, and protective efficacy of ebola virus GP DNA vaccines. *Journal of Virology*. 2007;**81**(4):1821-1837
- [13] Qiu X et al. Characterization of *Zaire ebolavirus* glycoprotein-specific monoclonal antibodies. *Clinical Immunology*. 2011;**141**(2):218-227
- [14] Volchkova VA et al. Genomic RNA editing and its impact on Ebola virus adaptation during serial passages in cell culture and infection of guinea pigs. *The Journal of Infectious Diseases*. 2011;**204**(Suppl_3):S941-S946
- [15] Ali MT, Islam MO. A highly conserved GEQYQQLR epitope has been identified in the nucleoprotein of Ebola virus by using an in silico approach. *Advances in Bioinformatics*. Feb 1 2015;**2015**

- [16] McElroy AK et al. Human Ebola virus infection results in substantial immune activation. *Proceedings of the National Academy of Sciences*. 2015;**112**(15):4719-4724
- [17] Patronov A, Doytchinova I. T-cell epitope vaccine design by immunoinformatics. *Open Biology*. Jan 1 2013;**3**(1):120139
- [18] Sundar K, Boesen A, Coico R. Computational prediction and identification of HLA-A2. 1-specific Ebola virus CTL epitopes. *Virology*. 2007;**360**(2):257-263
- [19] Haste Andersen P, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Science*. 2006;**15**(11):2558-2567
- [20] Li Pira G, Ivaldi F, Moretti P, Manca F. High throughput T epitope mapping and vaccine development. *BioMed Research International*. Jun 15 2010;**2010**
- [21] Kringelum JV et al. Reliable B cell epitope predictions: Impacts of method development and improved benchmarking. *PLoS Computational Biology*. 2012;**8**(12):e1002829
- [22] Tsurui H, Takahashi T. Prediction of T-cell epitope. *Journal of Pharmacological Sciences*. 2007;**105**(4):299-316
- [23] Tong JC, Tan TW, Ranganathan S. Methods and protocols for prediction of immunogenic epitopes. *Briefings in Bioinformatics*. 2006;**8**(2):96-108
- [24] Khan M et al. Epitope-based peptide vaccine design and target site depiction against Ebola viruses: An immunoinformatics study. *Scandinavian Journal of Immunology*. 2015;**82**(1):25-34
- [25] Lapelosa M et al. In silico vaccine design based on molecular simulations of rhinovirus chimeras presenting HIV-1 gp41 epitopes. *Journal of Molecular Biology*. 2009;**385**(2):675-691
- [26] Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Research*. 2006;**2**(1):2
- [27] Emini EA et al. Induction of hepatitis a virus-neutralizing antibody by a virus-specific synthetic peptide. *Journal of Virology*. 1985;**55**(3):836-839
- [28] Karplus P, Schulz G. Prediction of chain flexibility in proteins. *Naturwissenschaften*. 1985;**72**(4):212-213
- [29] Kolaskar A, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Letters*. 1990;**276**(1-2):172-174
- [30] Parker J, Guo D, Hodges R. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: Correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry*. 1986;**25**(19):5425-5432
- [31] Wu S et al. Prediction and identification of mouse cytotoxic T lymphocyte epitopes in Ebola virus glycoproteins. *Virology Journal*. 2012;**9**(1):111.4
- [32] Atanas P, Irini D. T-cell epitope vaccine design by immunoinformatics. *Open Biology* 2013;**3**:120139