

Exploiting the Reverse Vaccinology Approach to Design Novel Subunit Vaccine against Ebola Virus

Md. Asad Ullah ^{1*}, Bishajit Sarkar ¹, Syed Sajidul Islam¹

¹Department of Biotechnology and Genetic Engineering, Faculty of Biological Sciences, Jahangirnagar University, Savar, Dhaka, Bangladesh.

Corresponding author: Md. Asad Ullah *

Email of the corresponding author: ullah1194@gmail.com

Abstract

Ebola virus is a highly pathogenic RNA virus that causes haemorrhagic fever in human. With very high mortality rate, Ebola virus is considered as one of the dangerous viruses in the world. Although, the Ebola outbreaks claimed many lives in the past, no satisfactory treatment or vaccine have been discovered yet to fight against Ebola. For this reason, in this study, various tools of bioinformatics and immunoinformatics were used to design possible vaccines against Zaire Ebola virus strain Mayinga-76. To construct the vaccine, three potential antigenic proteins of the virus, matrix protein VP40, envelope glycoprotein and nucleoprotein were selected against which the vaccines would be designed. The MHC class-I, MHC class-II and B-cell epitopes were determined and after robust analysis through various tools and molecular docking analysis, three vaccine candidates, designated as EV-1, EV-2 and EV-3, were constructed. Since the highly conserved epitopes were used for vaccine construction, these vaccine constructs are also expected to be effective on other strains of Ebola virus like strain Gabon-94 and Kikwit-95. Next, the molecular docking study on these vaccine constructs were analyzed by molecular docking study and EV-1 emerged as the best vaccine construct. Later, molecular dynamics simulation study revealed the good performances as well as good stability of the vaccine protein. Finally, codon adaptation and in silico cloning were conducted to design a possible plasmid (pET-19b plasmid vector was used) for large scale, industrial production of the EV-1 vaccine.

Keywords: Ebola virus, vaccine, immunoinformatics, bioinformatics, epitopes.

1. Introduction

Highly pathogenic Ebola viruses are non-segmented RNA viruses with high mortality rates that are a distinguishing feature of this human pathogen associated with Ebola haemorrhagic fever [1, 2]. The first recorded Ebola outbreaks were observed in Africa continent between June and November of 1976 and 53% (150 of 284 victims) was the mortality rate [3, 4]. Currently, there is no effective anti-viral treatment for the rapid progression of Ebola that allows little opportunity to develop natural immunity [5]. Ebola virus was subsequently defined as the prototype viruses of a new taxonomic family, filoviridae [6]. Ebola viruses are subdivided into four different subtypes, Zaire, Sudan, Cote d'Ivoire, and Reston. The latest discovery includes fourth African species of human-pathogenic ebola virus is the Bundibugyo Ebola virus strain. Beside of this, there are many more pathogenic strains includes Zaire Mayinga-76, Gabon-96, Kikwit-95, Eckron-76 etc. [7, 8, 9]. Among them Zaire Mayinga-76 strain and its vaccination study is performed in this in silico research article. Ebola hemorrhagic viral fever is a severe, often deadly disease of humans and nonhuman primates cause by non-segmented, single stranded RNA molecules of Ebola virus [10]. After the attack of Ebola virus, the immune system causes a systemic inflammatory response that causes the impairment of the vascular, coagulation, asthenia and arthralgia systems, which leads to multi-organ failure, resembling septic shock are some of the indicators of ebola virus infections [7]. In 2014, World Health Organization (WHO) organized an emergency meeting attended by 60 researchers and public health administrators to assess and produce safe and effective Ebola vaccines as soon as possible [11]. A recent study shows that a trial vaccine called recombinant vesicular stomatitis virus–Zaire Ebola virus (rVSV-ZEBOV), which has shown to be safe and protective against the Zaire strain is recommended for use in Ebola outbreaks [12]. However, despite a huge number of research and investigation, there is no effective and

satisfactory therapeutic clarification or vaccination has been invented so far against the Mayinga-76 strain of Ebola virus. Even if an effective and safe vaccine can be produced, it is not likely to be hundred percent effective to succeed in slowing down and stopping the current outbreak [13]. Development of an effective and safe Ebola virus vaccine has been hindered by a lack of knowledge. In this study, an effort has been made to develop an effective Ebola vaccine using various tools of bioinformatics, immunoinformatics and reverse vaccinology (**Figure 01**). Reverse vaccinology is a process of vaccine development where the novel antigens are identified by analyzing the genomic information of an organism or a virus. In reverse vaccinology, various tools of in silico biology are used to discover the novel antigens by studying the genetic makeup of a pathogen and the genes that could lead to good epitopes are determined. This method is a quick easy and cost-effective way to design vaccine. Reverse vaccinology is successfully used for designing vaccines for many viruses like zika virus, chikungunya virus etc. [14].

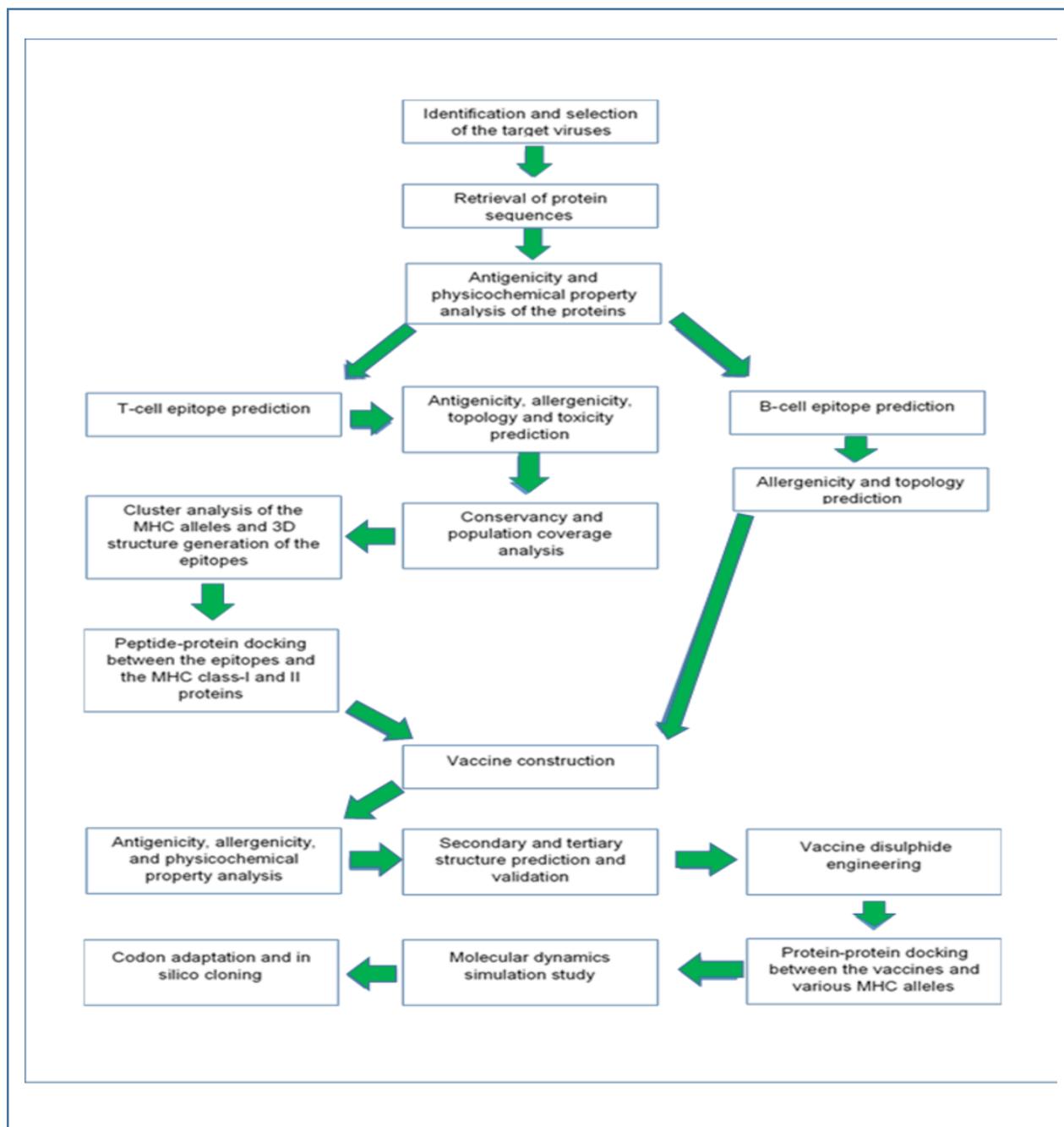


Figure 01. The flowchart of the procedures carried out in conducting the vaccine designing experiment.

2. Materials and methods:

The current experiment focuses on development of vaccines against the Zaire Ebola virus (strain Mayinga-76).

2.1. Strain Identification and Selection

The strain of the Ebola virus strain Mayinga-76 was identified and selected by analyzing different entries of the online server of National Center for Biotechnology Information or NCBI (<https://www.ncbi.nlm.nih.gov/>).

2.2. Retrieval of the Protein Sequences

The viral envelope glycoprotein (accession number: Q05320), matrix protein VP40 (accession number: Q05128) and nucleoprotein (accession number: P18272) were retrieved from the UniProt Knowledgebase (UniProtKB) tool of the online server UniProt (<https://www.uniprot.org/>).

2.3. Antigenicity Prediction and Physicochemical Property Analysis of the Protein Sequences

The antigenicity of the protein sequences were predicted by online server for antigenicity analysis, VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.htm>), keeping the threshold at 0.4 [15, 16, 17]. The various physicochemical properties of the selected protein sequences were determined by ExPASy's online tool ProtParam (<https://web.expasy.org/protparam/>) [18].

2.4. T-cell and B-cell Epitope Prediction

The T-cell and B-cell epitopes of the selected protein sequences were predicted using online epitope prediction server Immune Epitope Database or IEDB (<https://www.iedb.org/>). The IEDB database contains huge amount of experimental data on T-cell epitopes and antibodies. These data

are collected from various experiments that are carried on human, non-human primates and other animals. It is a server that allows robust analysis on various epitopes in the context of some tools: population coverage, conservation across antigens and clusters with similar sequences [19]. The MHC class-I restricted CD8+ cytotoxic T-lymphocyte (CTL) epitopes of the selected sequences were obtained using NetMHCpan EL 4.0 prediction method for HLA-A*11-01 allele. The MHC class-II restricted CD4+ helper T-lymphocyte (HTL) epitopes were obtained for HLA DRB1*04-01 allele using Sturniolo prediction method. Ten of the top twenty MHC class-I and MHC class-II epitopes were randomly selected based on their percentile scores and antigenicity scores (AS). Five random B-cell lymphocyte epitopes (BCL) were selected based on their length (the sequences that have ten amino acids or above were selected) and obtained using Bipipered linear epitope prediction method.

2.5. Transmembrane Topology and Antigenicity Prediction of the Selected Epitopes

The transmembrane topology of the selected epitopes were determined using the transmembrane topology of protein helices determinant, TMHMM v2.0 server (<http://www.cbs.dtu.dk/services/TMHMM/>). The server predicts whether the epitope would be transmembrane or it would remain inside or outside of the membrane [20]. The antigenicity of the selected epitopes were predicted using VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.htm>), using the tumor model and threshold of 0.4.

2.6. Allergenicity and Toxicity Prediction of the Epitopes

The allergenicity of the selected epitopes were predicted using two online tools, AllerTOP v2.0 (<https://www.ddg-pharmfac.net/AllerTOP/>) and AllergenFP v1.0 (<http://ddg-pharmfac.net/AllergenFP/>). However, the results predicted by AllerTOP were given priority since

the server has better accuracy of 88.7% than AllergenFP server (87.9%) [21, 22]. The toxicity prediction of the selected epitopes were carried out using ToxinPred server (<http://crdd.osdd.net/raghava/toxinpred/>), using SVM (Swiss-Prot) based method, keeping all the parameters default.

2.7. Conservancy Prediction of the Selected Epitopes

The conservancy analysis of the selected epitopes were performed using the epitope conservancy analysis tool of IEDB server (<https://www.iedb.org/conservancy/>) (Vita et al., 2018). The sequence identity threshold was kept ' ≥ 50 '. For the conservancy analysis of the selected epitopes of the Ebola virus strain Mayinga-76, the matrix protein VP40, envelope glycoprotein and nucleoprotein of Ebola virus strain Gabon-94 and Kikwit-95 were used for comparison along with the proteins of the Ebola virus strain Mayinga-76 itself (UniProt accession numbers: Q05128, Q2PDK5, Q77DJ6, Q05320, O11457, P87666, Q9QCE9, P18272 and O72142). Based on the antigenicity, allergenicity, toxicity and conservancy analysis, the best ligands were selected for the further analysis and vaccine construction. The T-cell epitopes that showed antigenicity, non-allergenicity, non-toxicity and high (more than 90%) conservancy and more than 50% minimum identity, were considered as the best epitopes. For B-cell epitope selection, only the antigenic and non-allergenic epitopes were taken for further analysis.

2.8. Cluster Analysis of the MHC Alleles

Cluster analysis of the MHC alleles helps to identify the alleles of the MHC class-I and class-II molecules that have similar binding specificities. The cluster analysis of the MHC alleles were carried out by online tool MHCcluster 2.0 (<http://www.cbs.dtu.dk/services/MHCcluster/>) [23]. During the analysis, the number of peptides to be included was kept 50,000, the number of

bootstrap calculations were kept 100 and all the HLA supertype representatives (MHC class-I) and HLA-DR representatives (MHC class-II) were selected. For analyzing the MHC class-I alleles, the NetMHCpan-2.8 prediction method was used. The output of the server generated results in the form of MHC specificity tree and MHC specificity heat-map.

2.9. Generation of the 3D Structures of the Selected Epitopes

The 3D structures of the selected best epitopes were generated using online 3D generating tool PEP-FOLD3 (<http://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3/>). The server is a tool for generating de novo peptide 3D structure [24, 25, 26].

2.10. Molecular Docking of the Selected Epitopes

The molecular docking of the selected epitopes were carried out by online docking tool PatchDock (<https://bioinfo3d.cs.tau.ac.il/PatchDock/php.php>). PatchDock tool has algorithms that divide the Connolly dot surface representation of the molecules into concave, convex and flat patches. After that the complementary patches are matched by the server for generating potential candidate transformations. Next, each of the candidate transformations is evaluated by scoring function and finally, an RMSD (root mean square deviation) clustering is applied to the candidate solutions for discarding the redundant solutions. The top scored solutions are made the top ranked solutions by the server. After docking by PatchDock, the docking results were refined and re-scored by FireDock server (<http://bioinfo3d.cs.tau.ac.il/FireDock/php.php>). The FireDock server generates global energies upon the refinement of the best solutions from the PatchDock server and ranks them based on the generated global energies and the lowest global energy is always appreciable and preferred [27, 28, 29, 30]. The molecular docking experiments were carried out using the HLA-A*11-01 allele (PDB ID: 5WJL) and HLA DRB1*04-01 (PDB ID: 5JLZ) as receptors and

the ligands were the best selected MHC-I and MHC-II epitopes, respectively. The receptors were downloaded from the RCSB- Protein Data Bank (<https://www.rcsb.org/>) server. The best results were visualized using Discovery Studio Visualizer (Dassault Systèmes BIOVIA, Discovery Studio Visualizer, v19.1.0.1828 (2019), San Diego: Dassault Systèmes).

2.11. Vaccine Construction

Three possible vaccines were constructed against the selected Ebola virus strain Mayinga-76. For successful vaccine construction, the predicted CTL, HTL and BCL epitopes were conjugated together. All the vaccines were constructed maintaining the sequence: adjuvant, PADRE sequence, CTL epitopes, HTL epitopes and BCL epitopes. Three different adjuvant sequences were used for vaccine construction: beta defensin, L7/L12 ribosomal protein and HABA protein (*M. tuberculosis*, accession number: AGV15514.1). By acting as agonist, beta-defensin adjuvants stimulate the activation of the toll like receptors (TLRs): 1, 2 and 4. The L7/L12 ribosomal protein and HABA protein have the ability to activate TLR-4. During the vaccine construction, various linkers were used: EAAAK linkers were used to conjugate the adjuvant and PADRE sequence, GGGG linkers were used to attach the PADRE sequence with the CTL epitopes and the CTL epitopes with the other CTL epitopes, GPGPG linkers were used to connect the CTL epitopes with the HTL epitopes and also the HTL epitopes among themselves. The KK linkers were used for conjugating the HTL and BCL epitopes as well as the BLC epitopes among themselves [31, 32, 33, 34, 35, 36, 37]. The PADRE sequence improves the CTL response of the vaccines that contain the PADRE sequence [38]. Total three vaccines were constructed in the experiment.

2.12. Antigenicity, Allergenicity and Physicochemical Property Analysis

The antigenicity of the constructed vaccines were determined by the online server VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.htm>). The threshold of the prediction was kept at 0.4 [15, 16, 17]. AlgPred (<http://crdd.osdd.net/raghava/algpred/>) and AllerTop v2.0 (<https://www.ddg-pharmfac.net/AllerTOP/>) were used for the prediction of the allergenicity of the vaccine constructs. The AlgPred server predicts the possible allergens based on similarity of known epitope of any of the known region of the protein [39]. MEME/MAST motif prediction approach is used in the allergenicity prediction of the vaccines by AlgPred. Moreover, various physicochemical properties of the vaccines were examined by the online server ProtParam (<https://web.expasy.org/protparam/>) [18].

2.13. Secondary and Tertiary Structure Prediction of the Vaccine Constructs

The secondary structures of the vaccine constructs were generated using online tool PRISPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>). PRISPRED is a simple secondary structure generator that can be used to predict the transmembrane topology, transmembrane helix, fold and domain recognition etc. along with the secondary structure prediction [40, 41]. The PRISPRED 4.0 prediction method was used to predict the secondary structures of the vaccine constructs. The β -sheet structure of the vaccines were determined by another online tool, NetTurnP v1.0 (<http://www.cbs.dtu.dk/services/NetTurnP/>) [42]. The tertiary or 3D structures of the vaccines were generated using online tool RaptorX (<http://raptorx.uchicago.edu/>) server. The server is a fully annotated tool for the prediction of protein structure, the property and contact prediction, sequence alignment etc. [43, 44, 45].

2.14. 3D Structure Refinement and Validation

The 3D structures of the constructed vaccines were refined using online refinement tool 3Drefine (<http://sysbio.rnet.missouri.edu/3Drefine/>). The server is a quick, easy and efficient tool for protein structure refinement [46]. For each of the vaccine, the refined model 1 was downloaded for validation. The refined vaccine proteins were then validated by analyzing the Ramachandran plots which were generated using the online tool, PROCHECK (<https://servicesn.mbi.ucla.edu/PROCHECK/>) [47, 48].

2.15. Vaccine Protein Disulfide Engineering

The vaccine protein disulfide engineering was carried out by online tool Disulfide by Design 2 v12.2 (<http://cptweb.cpt.wayne.edu/DbD2/>). The server predicts the possible sites within a protein structure that may undergo disulfide bond formation [49]. When engineering the disulfide bonds, the intra-chain, inter-chain and C_β for glycine residue were selected. The χ_3 angle was kept -87° or $+97^\circ \pm 5$ and C_α-C_β-S_γ angle was kept $114.6^\circ \pm 10$, during the experiment.

2.16. Protein-Protein Docking

In protein-protein docking, the constructed Ebola virus vaccines were analyzed by docking against various MHC alleles and toll like receptors (TLR). One best vaccine would be selected based on its superior performances in the docking experiment. When viral infections occur, the viral particles or antigens are recognized by the MHC complex. The various segments of the MHC molecules are encoded by different alleles. For this reason, the vaccines should have good binding affinity with these MHC portions that are encoded by different alleles [50]. All the vaccine constructs were docked against the selected MHC alleles to test their binding affinity. In this experiment, the vaccines constructs were docked against DRB1*0101 (PDB ID: 2FSE),

DRB3*0202 (PDB ID: 1A6A), DRB5*0101 (PDB ID: 1H15), DRB3*0101 (PDB ID: 2Q6W), DRB1*0401 (PDB ID: 2SEB), and DRB1*0301 (PDB ID: 3C5J). Moreover, studies have proved that TLR-8, present on the immune cells, are responsible for mediating the immune responses against the RNA viruses and TLR-3 of the immune cells mediates immune responses against the DNA viruses [51, 52]. The ebola virus is a RNA virus [53]. For this reason, the vaccine constructs of ebola virus were also docked against TLR-8 (PDB ID: 3W3M). The protein-protein docking was carried out using various online docking tools. The docking was carried out three times by three different online servers for improving the accuracy of the docking. First, the docking was carried out by ClusPro 2.0 (<https://cluspro.bu.edu/login.php>). The server ranks the clusters of docked complexes based on their center and lowest energy scores. However, these scores do not represent the actual binding affinity of the proteins with their targets [54, 55, 56]. The bonding affinity (ΔG in kcal mol⁻¹) of docked complexes were generated by PRODIGY tool of HADDOCK webserver (<https://haddock.science.uu.nl/>). The lower the binding energy generated by the server, the higher the binding affinity [57, 58, 59]. Moreover, the docking was again performed by PatchDock (<https://bioinfo3d.cs.tau.ac.il/PatchDock/php.php>) and later refined and re-scored by FireDock server (<http://bioinfo3d.cs.tau.ac.il/FireDock/php.php>). The FireDock server ranks the docked complexes based on their global energy and the lower the score, the better the result [27, 28, 29, 30]. Later, the docking was performed using HawkDock server (<http://cadd.zju.edu.cn/hawkdock/>). The Molecular Mechanics/Generalized Born Surface Area (MM-GBSA) study was also carried out using HawkDock server. According to the server, the lower scores and lower energy corresponds to better scores [60, 61, 62, 63]. The HawkDock server generates several models of docked complex and ranks them by assigning HawkDock scores in the ascending order. For each of the vaccines and their respective targets, the score of model 1 was

taken for analysis. Furthermore, the model 1 of every complex was analyzed for MM-GBSA study. From the docking experiment, one best vaccine was selected for further analysis. The docked structures were visualized by PyMol tool [64].

2.17. Molecular Dynamic Simulation

The molecular dynamics simulation study was conducted for the best selected vaccine, by the online server iMODS (<http://imods.chaconlab.org/>). iMODS server is a fast, online, user-friendly and effective molecular dynamics simulation server that can be used efficiently to investigate the structural dynamics of the protein complexes. The server provides the values of deformability, B-factor (mobility profiles), eigenvalues, variance, co-variance map and elastic network. For a protein complex, the deformability depends on the ability to deform at each of its amino acid. The eigenvalue is related with the energy that is required to deform the given structure and the lower the eigenvalue, the easier the deformability of the complex. The eigenvalue also represents the motion stiffness of the protein complex. The server is a fast and easy tool for determining and measuring the protein flexibility [65, 66, 67, 68, 69]. For analysing the molecular dynamics simulation of the EV-1-TLR-8 docked complex was used. The docked PDB files were uploaded to the iMODS server and the results were displayed keeping all the parameters as default.

2.18. Codon Adaptation and In Silico Cloning

In codon adaptation, the best selected vaccine protein was reverse transcribed to the possible DNA sequence. The DNA sequence should encode the target vaccine protein. After that, the reverse

transcribed DNA sequence was adapted according to the desired organism, so that the cellular mechanisms of that specific organism could use the codons of the adapted DNA sequences efficiently and provide better production of the desired product. Codon adaptation is a necessary step of in silico cloning because the same amino acid can be encoded by different codons in different organisms (codon biasness). Moreover, the cellular mechanisms of an organism may be different from another organism and a codon for a specific amino acid may not work in another organism. For this reason, codon adaptation can predict the suitable codon that can encode a specific amino acid in a specific organism [70, 71]. The predicted protein sequences of the best selected vaccines were used for codon adaptation by the Java Codon Adaptation Tool or JCat server (<http://www.jcat.de/>) [70]. Eukaryotic *E. coli* strain K12 was selected and rho-independent transcription terminators, prokaryotic ribosome binding sites and SgrA1 and SphI cleavage sites of restriction enzymes, were avoided. In the JCat server, the protein sequences were reverse translated to the optimized possible DNA sequences. The optimized DNA sequences were taken and SgrA1 and SphI restriction sites were conjugated at the N-terminal and C-terminal sites, respectively. Next, the SnapGene [72] restriction cloning module was used to insert the new adapted DNA sequences between SgrA1 and SphI of pET-19b vector.

3. Results

3.1. Identification, Selection and Retrieval of Viral Protein Sequences

The Zaire Ebola virus, strain Mayinga-76 was identified selected from the NCBI (<https://www.ncbi.nlm.nih.gov/>). Three proteins from the viral structures were selected for the possible vaccine construction. These proteins were: envelope glycoprotein (accession number: Q05320), matrix protein VP40 (accession number: Q05128) and nucleoprotein (accession number: P18272). The protein sequences were retrieved from the online server, UniProt (<https://www.uniprot.org/>). The protein sequences in fasta format:

```
>sp|Q05128|VP40_EBOZM Matrix protein VP40 OS=Zaire ebolavirus (strain Mayinga-76)  
OX=128952 GN=VP40 PE=1 SV=1
```

```
MRRVILPTAPPEYMEAIYPVRSNSTIARGGNSNTGFLTPESVNGDTPSNPLRPIADDTIDH  
ASHTPGSVSSAFILEAMVNVISGPKVLMKQIPIWLPLGVADQKTYSFDSTTAAIMLASYTI  
THFGKATNPLVRVNRLGPGIPDHPLRLLRIGNQAFLQEFVLPPVQLPQYFTFDLTALKLIT  
QPLPAATWTDDTPTGSNGALRPGISFHPKLRPILLPNKSGKKGNSADLTSPEKIQAIMTSL  
QDFKIVPIDPTKNIMGIEVPETLVHKLVTGKKVTSKNGQPIIPVLLPKYIGLDPVAPGDLTM  
VITQDCDTCHSPASLPAVIEK
```

```
>sp|Q05320|VGP_EBOZM Envelope glycoprotein OS=Zaire ebola virus (strain Mayinga-76)  
OX=128952 GN=GP PE=1 SV=1
```

```
MGV TGILQLPRDRFKRTSFFLWVILFQRTFSIPLGVIHNSTLQVSDVDKLVCRDKLSSTN  
QLRSVGLNLEGNGVATDVPSATKRWGFVSGVPPKVNYEAGEWAENCYNLEIKKPDG  
SECLPAAPDGIRGFPCRYVHKVSGTGPCAGDFAFHKEGAFFLYDRLASTVIYRGTTFAE  
GVVAFLILPQAKKDFSSHPLREPVNATEDPSSGYYSTTIRYQATGFGTNETEYLFVDN
```

LTYYVQLESRFTPQFLLQLNETIYTSGKRSNTTGKLIWKVNPEIDTTIGEWAFWETKKNLT
RKIRSEELSFTVVSNGAKNISGQSPARTSSDPGTNTTTEDHKIMASENSSAMVQVHSQGR
EAAVSHLTTLATISTSPQSLTTKPGPDNSTHNTVPYKLDISEATQVEQHHRRTDNDSTAS
DTPSATTAAGPPKAENTNTSKSTDFLDPATTTSPQNHSETAGNNNTHHQDTGEESASSG
KLGLITNTIAGVAGLITGRRTRREAIVNAQPKCNPNLHYWTTQDEGAAIGLAWIPYFGP
AAEGIYIEGLMHNQDGLICGLRQLANETTQALQLFLRATTELRTFSILNRKAIDFLLQRW
GGTCHILGPDCCIEPHDWTKNITDKIDQIIHDFVDKTLDPDQGDNDNWWTGWRQWIPAGI
GVTGVIIAVIALFCICKFVF

>sp|P18272|NCAP_EBOZM Nucleoprotein OS=Zaire ebola virus (strain Mayinga-76)

OX=128952 GN=NP PE=1 SV=2

MDSRPQKIWMAPSLTESDMDYHKILTAGLSVQQGIVRQRVIPVYQVNNLEEICQLIIQAF
EAGVDFQESADSFLMLCLHHAYQGDYKLFLESGAVKYLEGHGFRFEVKKRDGVKRLE
ELLPAVSSGKNIKRTLAAMPEEETTEANAGQFLSFASLFLPKLVVGEKACLEKVQRQIQV
HAEQGLIQYPTAWQSVGHMMVIFRLMRTNFLIKFLLIHQGMHMAVAGHDANDAVISNSV
AQARFSGLLIVKTVLDHILQKTERGVRLHPLARTAKVKNEVNSFKAALSSLAKHGEYAP
FARLLNLSGVNNLEHGLFPQLSAIALGVATAHGSTLAGVNVGEQYQQLREAATEAEKQ
LQYYAESRELDHLGLDDQEKKILMNFHQKKNEISFQQTNAMVTLRKERLAKLTEAITAA
SLPKTSGHYDDDDDDIPFPGPINDDDNPGHQDDDDPTDSQD TTIPDVVVD PDDGSYGEYQS
YSENGMNAPDDLVLFDLDEDEDTKPVPNRSTKGGQQKNSQKGQHIEGRQTQSRPIQN
VPGPHRTIHHASAPLTDNDRRNEPSGSTSPRMLTPINEEADPLDDADDETSSLPPLESDDDE
EQDRDGTSNRTPTVAPPAPVYRDHSEKKELPQDEQQDQDHTQEARNQDSDNTQSEHSF

EEMYRHILRSQGPFDVLYYHMMKDEPVVSTSDGKEYTYPDSLEEEYPPWLTEKEAM
NEENRFVTLDGQQFYWPVMNHKNKFMALQHHQ

3.2. Antigenicity Prediction and Physicochemical Property Analysis of the Protein Sequences

VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.htm>) and ProtParam (<https://web.expasy.org/protparam/>) servers were used for antigenicity analysis and physicochemical property prediction of the selected protein sequences (**Table 01**). In the physicochemical property analysis, the number of amino acids, the molecular weights, theoretical pI, extinction coefficients, half-lives, instability indexes, aliphatic indexes and grand average of hydropathicity (GRAVY) of the three proteins were predicted. All the three selected protein sequences showed good level of antigenicity. The matrix protein VP40 had the lowest molecular weight of 35182.83 and only envelope glycoprotein was stable according to the prediction. All of the three proteins had 30 hours of half-life in mammalian cells. The matrix protein VP40 also showed the highest theoretical pI, aliphatic index and GRAVY values (**Table 01**).

3.3. T-cell and B-cell Epitope Prediction and Topology Determination of the Epitopes

The T-cell epitopes of MHC class-I of the three proteins were determined by NetMHCpan EL 4.0 prediction method of the IEDB (<https://www.iedb.org/>) server keeping the sequence length 9. The server generated over 100 such epitopes. However, based on analyzing the antigenicity scores (AS) and percentile scores, for each epitope, ten potential epitopes from the top twenty epitopes were selected randomly for antigenicity, allergenicity, toxicity and conservancy tests. The server ranks the predicted epitopes based on the ascending order of percentile scores. The lower the percentile, the better the binding affinity. The T-cell epitopes of MHC class-II (HLA DRB1*04-01 allele) of

the proteins were also determined by IEDB (<https://www.iedb.org/>) server, where the Sturniolo prediction methods was used. For each protein, ten of the top twenty epitopes were selected randomly for further analysis. Moreover, the B-cell epitopes of the proteins were selected using Bipipered linear epitope prediction method of the IEDB (<https://www.iedb.org/>) server and epitopes were selected based on their higher lengths (**Figure 02**). The topology of the selected epitopes were determined by TMHMM v2.0 server (<http://www.cbs.dtu.dk/services/TMHMM/>). **Table 02** and **Table 03** list the potential T-cell epitopes of matrix protein VP40, **Table 04** and **Table 05** list the potential T-cell epitopes of envelope glycoprotein, **Table 06** and **Table 07** list the potential T-cell epitopes of nucleoprotein and **Table 08** list the potential B-cell epitopes with their respective topologies.

3.4. Antigenicity, allergenicity, toxicity and conservancy analysis

In the antigenicity, allergenicity, toxicity and conservancy analysis, the T-cell epitopes that were found to be highly antigenic as well as non-allergenic, non-toxic, had minimum identity of over 50% and had conservancy of over 90% as well as antigenic and non-allergenic B-cell epitopes were selected for further analysis and vaccine construction. Among the ten selected MHC class-I epitopes and ten selected MHC class-II epitopes of matrix protein VP40, total six epitopes (three epitopes from each of the category) were selected based on the mentioned criteria: MVNVISGPK, ILLPNKSGK, GISFHPKLR, QKTYSFDSTTAA, KTYSFDSTTAAI and WLPLGVADQKTY. On the other hand, among the ten selected MHC class-I epitopes and ten selected MHC class-II epitopes of envelope glycoprotein, total six epitopes (three epitopes from each of the category) were selected based on the mentioned criteria: STHNTPVYK, ATQVEQHHR, SGYYSTTIR, TPQFLLQLNETI, PQFLLQLNETIY and FLLQLNETIYTS. Moreover, like these proteins, six epitopes that obeyed the mentioned criteria, were selected for further analysis from the

nucleoprotein epitopes: TVLDHILQK, VVFSTSDGK, QTNAMVTLR, SLLMLCLHHAY, FLLMLCLHHAYQ and VVFSTSDGKEYT. For the selection of the B-cell epitopes, the highly antigenic and non-allergenic sequences were taken for vaccine construction. Three epitopes from each of the protein category were selected. For this reason, total nine epitopes B-cell epitopes were selected for vaccine construction, since they obeyed the selection criteria. (**Table 02, Table 03, Table 04, Table 05, Table 06, Table 07 and Table 08**).

Table 01. The antigenicity and physicochemical property analysis of the selected viral proteins.

Name of the protein sequence	Antigenicity (threshold 0.4)	Total amino acids	Molecular weight	Theoretical pI	Ext. coefficient (in $M^{-1} cm^{-1}$)	Est. half-life (in mammalian cell)	Instability index	Aliphatic index	Grand average of hydropathicity (GRAVY)
Matrix protein VP40	0.5692 (bacteria), 0.4598 (tumor), 0.5496 (parasite)	326	35182.83	8.76	20065	30 hours	40.39 (unstable)	96.32	-0.052
Envelope glycoprotein	0.5852 (bacteria), 0.4933 (tumor), 0.5364 (parasite)	676	74464.46	6.16	74464.46	30 hours	38.36 (stable)	75.77	-0.380
Nucleoprotein	0.5315 (bacteria), 0.5442 (tumor), 0.4692 (parasite)	739	83286.68	4.98	53415	30 hours	50.31 (unstable)	74.32	-0.691

Table 02. MHC class-I epitope prediction and topology, antigenicity, allergenicity, toxicity and conservancy analysis of the epitopes of matrix protein VP40.

Matrix protein VP40										
Epitope	Start	End	Topology	AS	Percentile scores	Antigenicity (tumor, threshold 0.4)	Allergenicity	Toxicity	Conservancy	Minimum identity
ASLPAVIEK	318	326	Outside	0.972	0.01	Antigen	Allergen	Non-toxic	100%	100%
FTFDLTALK	172	180	Outside	0.881	0.02	Non-antigen	Allergen	Non-toxic	100%	100%
MVNVISGPK	78	86	Inside	0.469	0.37	Antigen	Non-allergen	Non-toxic	100%	100%
SADLTSPEK	228	236	Outside	0.391	0.52	Non-antigen	Non-allergen	Non-toxic	88.89%	100%
TLVHKLTKG	266	274	Inside	0.363	0.57	Non-antigen	Allergen	Non-toxic	100%	100%
ILLPNKSGK	216	224	Outside	0.358	0.58	Antigen	Non-allergen	Non-toxic	100%	100%

IMTSLQDFK	240	258	Inside	0.356	0.58	Antigen	Allergen	Non-toxic	88.89%	100%
LVHKL TGKK	267	275	Inside	0.276	0.82	Antigen	Allergen	Non-toxic	100%	100%
MLASYTITH	116	124	Inside	0.131	1.6	Non-antigen	Non-allergen	Non-toxic	100%	100%
GISFHPKLR	206	214	Outside	0.127	1.7	Antigen	Non-allergen	Non-toxic	100%	100%

Table 03. MHC class-II epitope prediction and topology, antigenicity, allergenicity, toxicity and conservancy analysis of the epitopes of matrix protein VP40.

Matrix protein VP40										
Epitope	Start	End	Topology	AS	Percent ile scores	Antigenici ty (tumor, threshold 0.4)	Allergen icity	Toxicity	Conser vancy	Minimum identity
YFTFDLTALKLI	171	182	Outside	3.80	1.20	Non- antigen	Non- allergen	Non-toxic	100%	100%
QYFTFDLTALKL	170	181	Outside	3.80	1.20	Non- antigen	Non- allergen	Non-toxic	100%	100%
PQYFTFDLTALK	169	180	Outside	3.80	1.20	Antigen	Allergen	Non-toxic	100%	100%
QKTYSFDSTTAA	103	114	Inside	3.70	1.40	Antigen	Non- allergen	Non-toxic	100%	100%
KTYSFDSTTAAI	104	115	Inside	3.70	1.40	Antigen	Non- allergen	Non-toxic	100%	100%
PIWLPLGVADQK	93	104	Outside	3.50	1.90	Antigen	Allergen	Non-toxic	100%	100%
IWLPLGVADQKT	94	105	Outside	3.50	1.90	Antigen	Allergen	Non-toxic	100%	100%
WLPLGVADQKTY	95	106	Outside	3.50	1.90	Antigen	Non- allergen	Non-toxic	100%	100%
IYPVRSNSTIAR	17	28	Inside	3.28	2.40	Antigen	Allergen	Non-toxic	100%	100%
PVRSNSTIARGG	19	30	Inside	3.28	2.40	Antigen	Allergen	Non-toxic	100%	100%

Table 04. MHC class-I epitope prediction and topology, antigenicity, allergenicity, toxicity and conservancy analysis of the epitopes of envelope glycoprotein.

Envelope glycoprotein										
Epitope	Start	End	Topology	AS	Percentile scores	Antigenicity (tumor, threshold 0.4)	Allergenicity	Toxicity	Conservancy	Minimum identity
STHNTPVYK	387	395	Inside	0.964	0.01	Antigenic	Non-allergenic	Non-toxic	100%	100%
RTFSILNRK	580	588	Inside	0.947	0.01	Non-antigen	Non-allergen	Non-toxic	100%	100%
ATDVPSATK	76	84	Inside	0.872	0.03	Antigen	Allergen	Non-toxic	100%	100%
KAIDFLLQR	588	596	Inside	0.864	0.04	Non-antigen	Allergen	Non-toxic	100%	100%
QIIHDFVDK	625	633	Inside	0.465	0.37	Non-antigen	Non-allergen	Non-toxic	100%	100%
ATQVEQHHR	401	409	Inside	0.314	0.69	Antigenic	Non-allergenic	Non-toxic	100%	100%
CAGDFAFHK	147	155	Inside	0.311	0.70	Non-antigen	Allergen	Non-toxic	100%	100%
TSPQSLTTK	373	381	Inside	0.295	0.76	Antigen	Allergen	Non-toxic	66.67%	44.44%
SGYYSTTIR	211	219	Inside	0.224	1.00	Antigen	Non-allergen	Non-toxic	100%	100%
KLSSTNQLR	56	64	Inside	0.228	0.98	Antigen	Allergen	Non-toxic	100%	100%

Table 05. MHC class-II epitope prediction and topology, antigenicity, allergenicity, toxicity and conservancy analysis of the epitopes of envelope glycoprotein.

Envelope glycoprotein										
Epitope	Start	End	Topology	AS	Percent ile scores	Antigenicity (tumor, threshold 0.4)	Allergen icity	Toxicity	Conser vancy	Minimum identity
TPQFLQLNETI	249	260	Outside	4.00	0.93	Antigen	Non- allergen	Non-toxic	100%	100%
QFLLQLNETIYT	251	262	Outside	4.00	0.93	Antigen	Allergen	Non-toxic	100%	91.67%
VYKLDISEATQV	393	404	Inside	3.50	1.90	Antigen	Allergen	Non-toxic	100%	100%
VIHNSTLQVSDV	37	48	Inside	3.10	3.00	Antigen	Allergen	Non-toxic	100%	100%
PQFLQLNETIY	250	261	Outside	4.00	0.93	Antigen	Non- allergen	Non-toxic	100%	100%
FLLQLNETIYTS	252	263	Outside	4.00	0.93	Antigen	Non- allergen	Non-toxic	100%	100%
TPVYKLDISEAT	391	402	Outside	3.50	1.90	Non- antigen	Allergen	Non-toxic	100%	100%
LGVIHNSTLQVS	35	46	Outside	3.10	3.00	Antigen	Allergen	Non-toxic	100%	100%
GVIHNSTLQVSD	36	47	Outside	3.10	3.00	Antigen	Allergen	Non-toxic	100%	100%
QLFLRATTELRT	570	580	Inside	2.80	3.90	Non- antigen	Non- allergen	Non-toxic	100%	100%

Table 06. MHC class-I epitope prediction and topology, antigenicity, allergenicity, toxicity and conservancy analysis of the epitopes of nucleoprotein.

Nucleoprotein										
Epitope	Start	End	Topology	AS	Percent ile scores	Antigenicity (tumor, threshold 0.4)	Allergenicity	Toxicity	Conservancy	Minimum identity
TVLDHILQK	249	257	Inside	0.981	0.01	Antigen	Non-allergen	Non-toxic	100%	100%
AVLYYHMMK	664	672	Inside	0.865	0.03	Antigen	Allergen	Non-toxic	100%	100%
VVFSTSDGK	676	684	Outside	0.788	0.08	Antigen	Non-allergen	Non-toxic	100%	100%
KILMNFHQK	374	382	Inside	0.704	0.14	Non-antigen	Allergen	Non-toxic	100%	100%
AVSSGKNIK	123	131	Inside	0.643	0.02	Antigen	Allergen	Non-toxic	100%	100%
VTLDGQQFY	713	721	Outside	0.568	0.26	Non-antigen	Non-allergen	Non-toxic	100%	100%
QTNAMVTLR	390	398	Inside	0.512	0.32	Antigen	Non-allergen	Non-toxic	100%	100%
HGFRFEVKK	102	110	Inside	0.310	0.71	Antigen	Allergen	Non-toxic	100%	100%
SFASLFLPK	152	160	Outside	0.306	0.72	Non-antigen	Non-allergen	Non-toxic	100%	100%
VSSGKNIKR	124	132	Inside	0.264	0.85	Antigen	Allergen	Non-toxic	100%	100%

Table 07. MHC class-II epitope prediction and topology, antigenicity, allergenicity, toxicity and conservancy analysis of the epitopes of nucleoprotein.

Nucleoprotein										
Epitope	Start	End	Topology	AS	Percent ile scores	Antigenicity (tumor, threshold 0.4)	Allergen icity	Toxicity	Conser vancy	Minimum identity
VIFRLMRTNFLI	199	210	Inside	5.60	0.07	Non- antigen	Non- allergen	Non-toxic	100%	100%
IFRLMRTNFLIK	200	211	Inside	5.60	0.07	Non- antigen	Non- allergen	Non-toxic	100%	100%
FRLMRTNFLIKF	201	212	Inside	5.60	0.07	Antigen	Allergen	Non-toxic	100%	100%
SFLMLCLHHAY	72	83	Outside	3.68	1.50	Antigen	Non- allergen	Non-toxic	100%	100%
FLMLCLHHAYQ	73	84	Outside	3.68	1.50	Antigen	Non- allergen	Non-toxic	100%	100%
DSFLMLCLHHA	71	82	Outside	3.68	1.50	Antigen	Allergen	Non-toxic	100%	100%
VVFSTSDGKEYT	676	687	Outside	3.10	2.90	Antigen	Non- allergen	Non-toxic	100%	100%
DEPVVFSTSDGK	673	684	Outside	3.10	2.90	Antigen	Allergen	Non-toxic	100%	100%
PVYRDHSEKKEL	609	620	Inside	2.90	3.60	Non- antigen	Non- allergen	Non-toxic	100%	91.67%
PAPVYRDHSEKK	607	618	Inside	2.90	3.60	Non- antigen	Non- allergen	Non-toxic	100%	91.67%

Table 08. B-cell epitope prediction and antigenicity, allergenicity and topology prediction of the epitopes of the three selected proteins.

Matrix protein VP40				Envelope glycoprotein				Nucleoprotein			
Epitope	Antigenicity (tumor, threshold 0.4)	Allergy enitcit y	Topology	Epitope	Antigenicity (tumor, threshold 0.4)	Allergenicity	Topology	Epitope	Antigenicity (tumor, threshold 0.4)	Allergenicity	Topology
RSNSTIARGG NSNTGFLTPE SVNGDTPSN PLRPIADDTI DHASHTPGS V	Antigen	Non-allergen	Inside	NGVATD VPSATK RWGFRS GVPPKV VNYEAG EWAE	Antigen	Non-allergen	Outside	AMPEEETTE ANA	Antigen	Non-allergen	Inside
ADQKTYISFD S	Non-antigen	Allergen	Inside	YTSGKR SNTT	Antigen	Allergen	Inside	VGEQYQQLR EAATEAEKQ LQQYAESR	Antigen	Allergen	Inside
NKSGKKGNS ADLTSPEK	Antigen	Non-allergen	Inside	KKPDGS ECLPAA PDGIRG	Antigen	Non-allergen	Outside	APLTDNDRR NEPSGSTSPR ML	Antigen	Non-allergen	Inside
KKVTSKNGQ P	Antigen	Non-allergen	Inside	NGAKNI SGQSPA RTSSDP GTNTTT EDH	Antigen	Non-allergen	Inside	VVFSTSDGKE YTYPDSLEEE YPPWLTEKE AMNE	Antigen	Non-allergen	Outside
LPAATWTDD TPTGSNGAL R	Antigen	Allergen	Outside	KTLPDQ GDNDN WWT	Antigen	Allergen	Outside	SLPKTSGHYD DDDDIPFPGPI NDDDNPGHQ	Non-antigen	Allergen	Outside

								DDDP TDSQD TTIPDVVVDP DDGSYGEYQ SYSENGMNA P			
--	--	--	--	--	--	--	--	---------------------------------------------------------	--	--	--

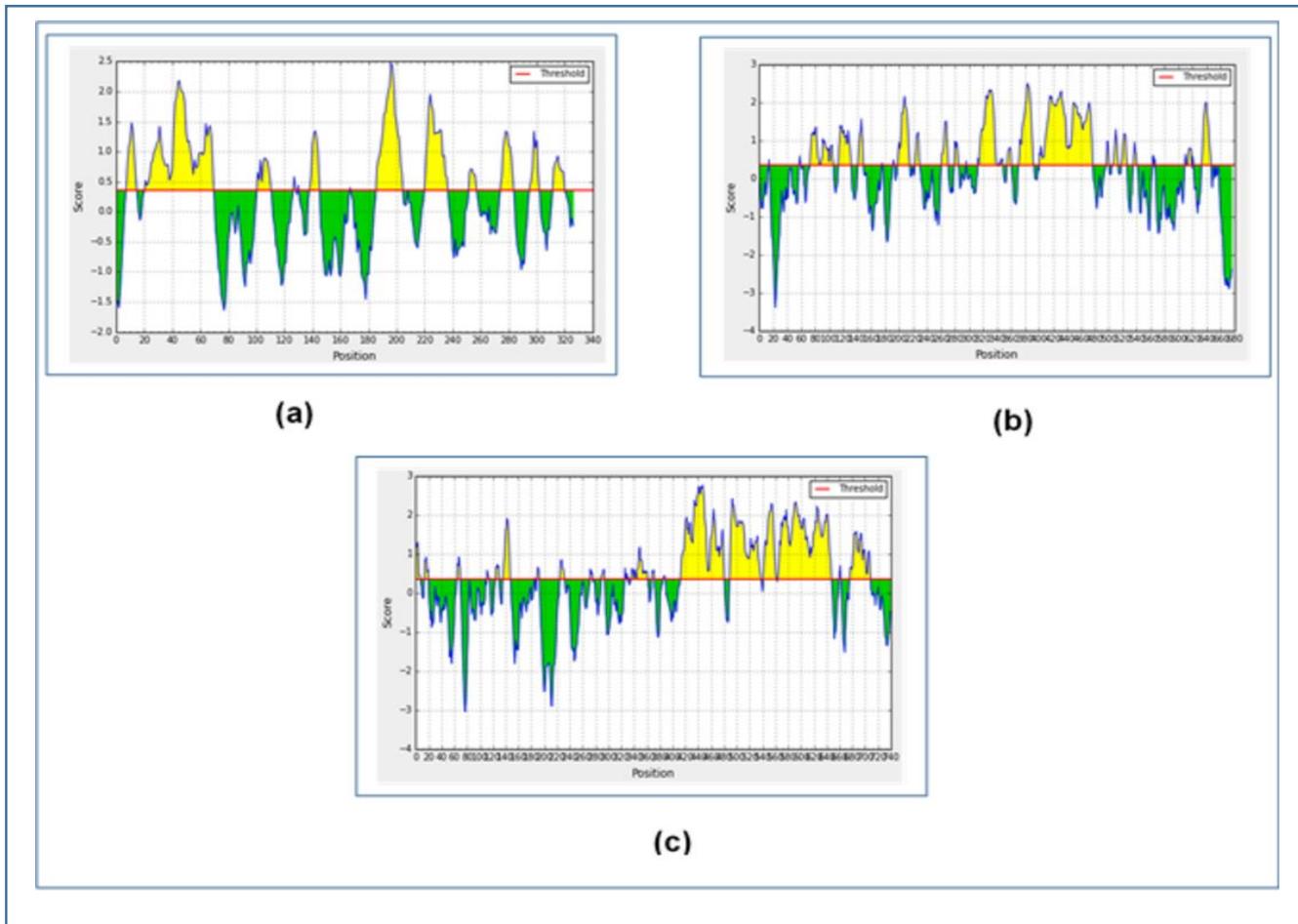


Figure 02. Figure showing the graphs of the B-cell epitope prediction of the three selected proteins of Ebola virus strain Mayinga-76, using Bipipered linear epitope prediction method. Here, (a) is the graph of epitope prediction for matrix protein VP40, (b) is the graph of epitope prediction for envelope glycoprotein and (c) is the graph of epitope prediction for nucleoprotein.

3.5. Cluster Analysis of the MHC Alleles

The cluster analysis of the possible MHC class-I and MHC class-II alleles that may interact with the predicted epitopes were performed by online tool MHCcluster 2.0 (<http://www.cbs.dtu.dk/services/MHCcluster/>). The tool generates the clusters of the alleles in phylogenetic manner. **Figure 03** illustrates the outcome of the experiment where the red zone indicates strong interaction and the yellow zone corresponds to weaker interaction.

3.6. Generation of the 3D Structures of the Epitopes and Peptide-Protein Docking

All the T-cell epitopes were subjected to 3D structure generation by the PEP-FOLD3 server. The 3D structures were generated for peptide-protein docking. The docking was performed to find out, whether all the epitopes had the ability to bind with the MHC class-I and MHC class-II molecule. The selected epitopes were docked against the HLA-A*11-01 allele (PDB ID: 5WJL) and HLA DRB1*04-01 (PDB ID: 5JLZ). The docking was performed using PatchDock online docking tool and then the results were refined by FireDock online server. Among the MHC class-I epitopes of matrix protein VP40, GISFHPKLR showed the best result with the lowest global energy of -31.68. Among the MHC class-I epitopes of envelope glycoprotein, SGYYSTTIR generated the lowest and best global energy score of -42.87. QTNAMVTLR generated the best global energy score of -45.40 of the MHC class-I epitopes of nucleoprotein. Among the MHC class-II epitopes of matrix protein VP40, KTYSFDSTTAAI generated the best global energy score of -24.32 and TPQFLLQLNETI generated the lowest global energy of -2.80 and VVFSTSDGKEYT generated the lowest global energy of -12.06, among the MHC class-II epitopes of envelope glycoprotein and nucleoprotein, respectively (**Table 09**). **Figure 04** illustrates the interaction between the best docked epitopes with their respective targets.

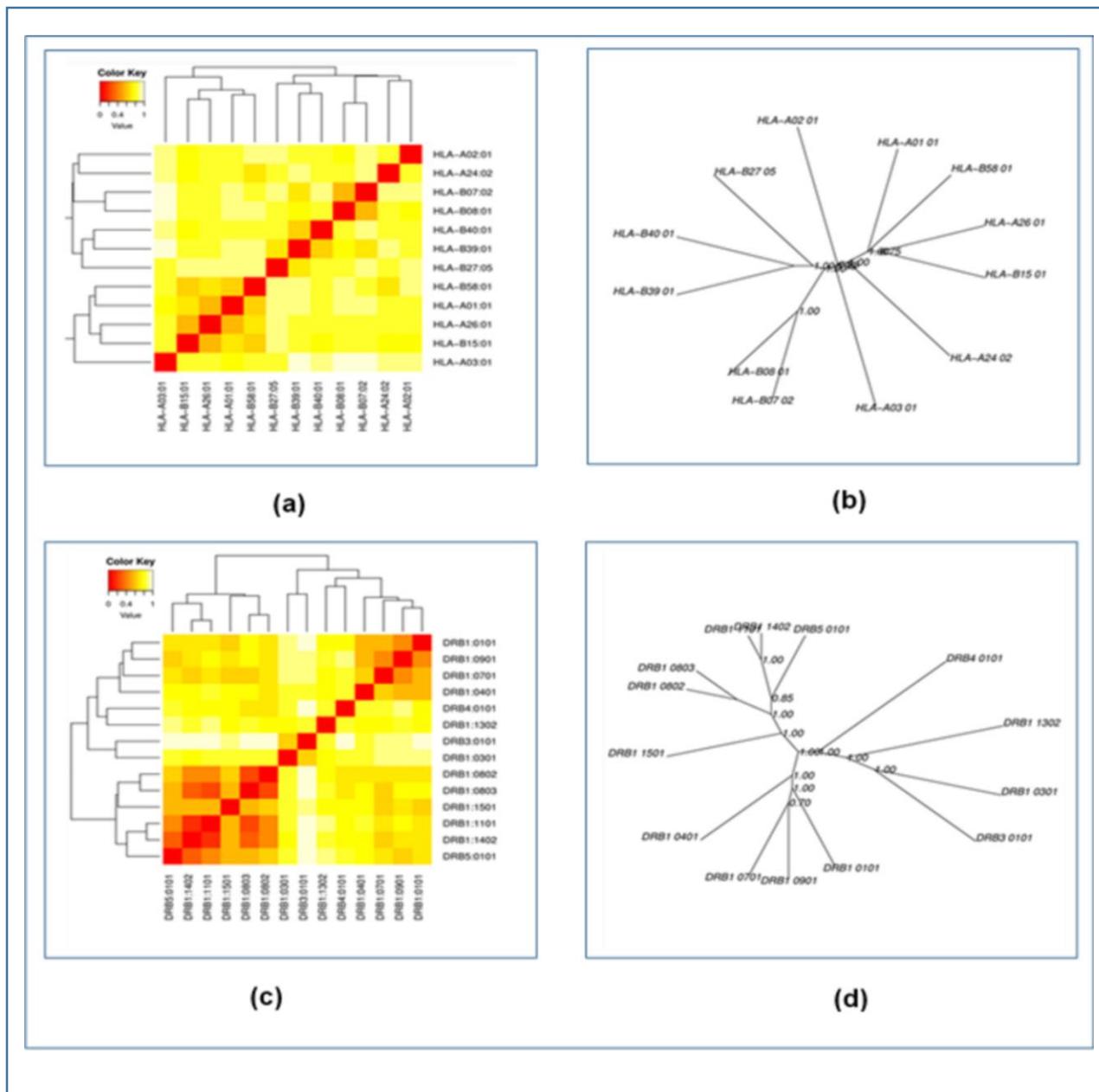


Figure 03. The results of the MHC cluster analysis. Here, (a) is the heat map of MHC class-I cluster analysis, (b) is the tree map of MHC class-I cluster analysis, (c) is the heat map of MHC class-II cluster analysis, (d) is the tree map of MHC class-II cluster analysis. The cluster analysis was carried out using online server MHCcluster 2.0 (<http://www.cbs.dtu.dk/services/MHCcluster/>).

Table 09. Results of molecular docking analysis of the selected epitopes.

Name of the protein	Epitope	MHC allele	Global energy	Hydrogen bond energy	Epitope	MHC allele	Global energy	Hydrogen bond energy
Matrix protein VP40	MVNVISGP K	HLA-A*11-01 allele (PDB ID: 5WJL)	-27.34	-0.63	QKTYSFDSTTAA	HLA DRB1*04-01 (PDB ID: 5JLZ)	-3.45	0.00
	ILLPNKSG K		-31.35	-0.94	KTYSFDSTTAAI		-24.32	-3.49
	GISFHPKL R		-31.68	-3.76	WLPLGVADQKTY		-17.07	-0.78
Envelope glycoprotein	STHNTPVY K	HLA-A*11-01 allele (PDB ID: 5WJL)	-37.49	-3.52	TPQFLLQLNETI	HLA DRB1*04-01 (PDB ID: 5JLZ)	-2.80	-4.87
	ATQVEQH HR		-30.49	-3.87	PQFLLQLNETIY		-0.39	0.00
	SGYYSTTI R		-42.87	-4.99	FLLQLNETIYTS		-1.78	-3.22
Nucleo protein	TVLDHILQ K	HLA-A*11-01 allele (PDB ID: 5WJL)	-29.99	-0.87	SFLLMLCLHHAY	HLA DRB1*04-01 (PDB ID: 5JLZ)	-8.47	-2.24
	VVFSTSDG K		-34.14	-3.19	FLLMLCLHHAYQ		-4.62	0.00
	QTNAMVT LR		-45.40	-3.31	VVFSTSDGKEYT		-12.06	-0.28

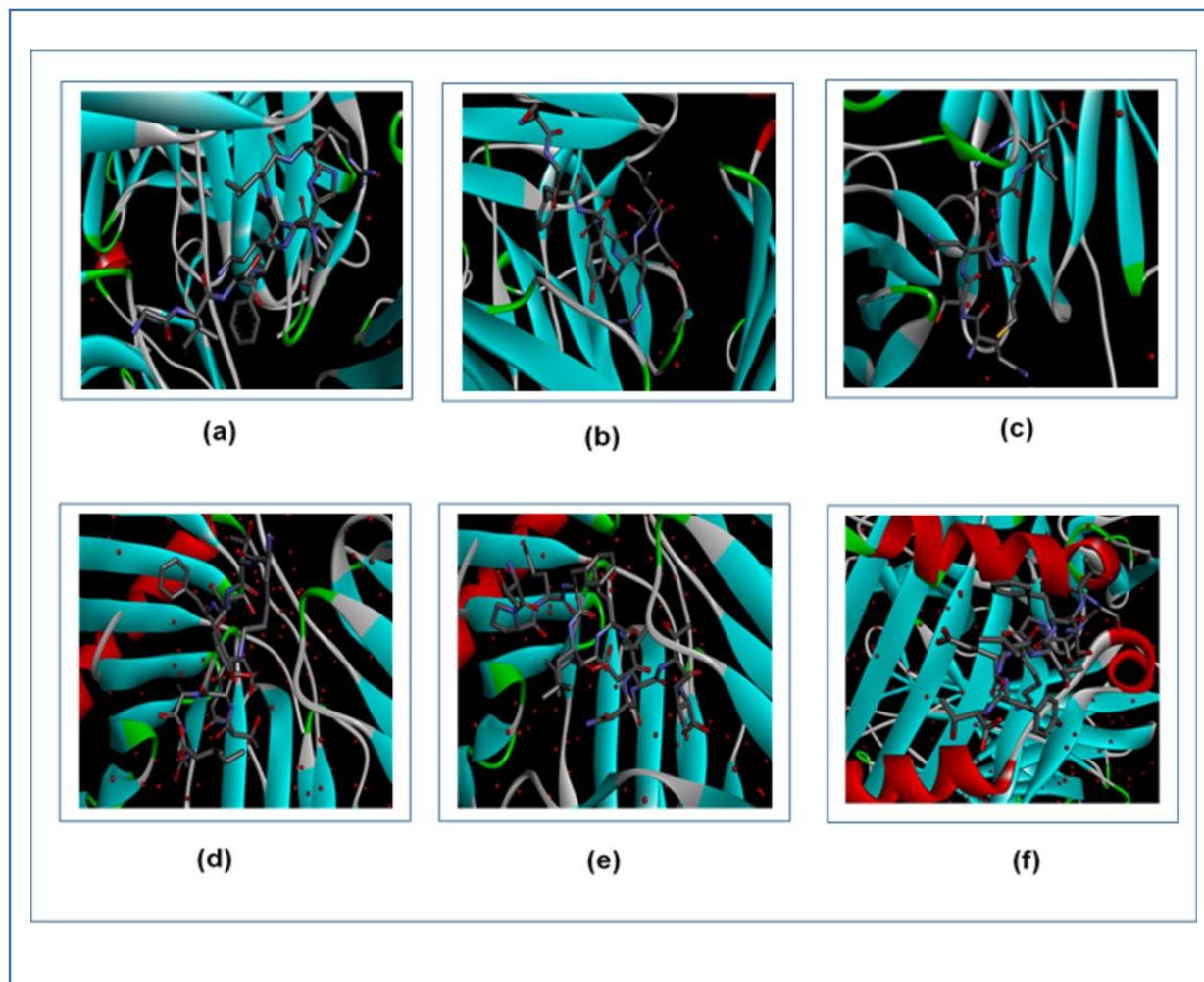


Figure 04. Figure showing the interactions between the best epitopes from the three proteins and their respective receptors. Here, (a) is the interaction between GISFHPKLR and MHC class-I, (b) is the interaction between SGYYSTTIR and MHC class-I, (c) is the interaction between QTNAMVTLR and MHC class-I molecule, (d) is the interaction between KTYSFDSTTAI and MHC class-II, (e) is the interaction between TPQFLLQLNETI and MHC class-II, (f) is the interaction between VVFSTSDGKEYT and MHC class-II molecule. The interactions were visualized by Discovery Studio Visualizer.

3.7. Vaccine Construction

After successful docking, three vaccines were constructed, that could be used effectively to fight against ebola virus strain Mayinga-76. For vaccine constructions, three different adjuvants were used to construct three different vaccine for each of the viruses: beta defensin, L7/L12 ribosomal protein and HABA protein, were used. PADRE sequence was also used for vaccine construction. Three different vaccine constructs differed from each other only in their adjuvant sequences. To construct a vaccine, first the adjuvant sequence was conjugated with the PADRE sequence by EAAAK linker, then the PADRE sequence was added to the CTL epitopes by GGGS linkers, the CTL epitopes were also conjugated with each other by the GGGS linkers. Next, the HTL epitopes were conjugated by GPGPG linkers and the BCL epitopes were linked by KK linkers. Each vaccine construct was ended by an additional GGGS linker. The newly constructed vaccines were designated as: EV-1, EV-2 and EV-3 (**Table 10**).

LHHAYQ GP PGVVFSTSDGKEYT KK RSNSTIARGGNSNTGFLTPESVNGDTPSNPLRPIADDTIDHASHTPGS V KK NKSGKKGNSADLTSPEK KKK KVTSKNGQP KK NGVATDVPSATKRWGFRSGVPPKVVNYEAGEWAEK KKK PDGSECLPAAPDGIRG KK NGAKNISGQSPARTSSDPGTNTTTEDH KK KAMPEEETTEANAK KK APLTDNDR RNEPSGSTSPRML KK VVFSTSDGKEYTYPDSLEEEYPPWLTEKEAMNE KK KAFVAAWTLKAAAG GG S	
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

3.8. Antigenicity, Allergenicity and Physicochemical Property Analysis of the Vaccine Constructs

The antigenicity prediction of the three vaccine constructs showed that all of the vaccine constructs were potential antigens. However, since all of the vaccine constructs were non-allergen, they are safe to use. In the physicochemical property analysis, the number of amino acids, molecular weight, extinction coefficient (in $M^{-1} cm^{-1}$), theoretical pI, half-life, instability index, aliphatic index and GRAVY were determined. All the vaccines quite similar theoretical pI and ext. coefficient (EV-3 had the lowest value of $55475 M^{-1} cm^{-1}$) and all of them were stable. All of the vaccine constructs had the similar half-life of 1 hour in the mammalian cells. EV-2 had the highest GRAVY value of -0.504. The antigenicity, allergenicity and physicochemical property analysis of the three vaccine constructs are listed in **Table 11**.

Table 11. Antigenicity, allergenicity and physicochemical property analysis of the three vaccine constructs.

Name of the vaccine	Antigenicity (on tumor, threshold 0.4)	Allergenicity	Total amino acids	Molecular weight	Theoretical pI	Ext. coefficient (in $M^{-1} cm^{-1}$)	Est. half-life (in mammalian cell)	Instability index	Aliphatic index	Grand average of hydrophobicity (GRAVY)
EV-1	Antigen (0.697)	Non-allergen	596	62136.84	9.62	58830	1 hour	37.98 (Stable)	55.70	-0.667
EV-2	Antigen (0.673)	Non-allergen	681	70416.13	9.01	55475	1 hour	34.14 (stable)	64.13	-0.504
EV-3	Antigen (0.709)	Non-allergen	710	74604.30	8.99	59945	1 hour	38.89 (stable)	62.90	-0.627

3.9. Secondary and Tertiary Structure Prediction of the Vaccine Constructs

The secondary structures of the three vaccine constructs were generated by two online tools, PRISPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>) and NetTurnP v1.0 (<http://www.cbs.dtu.dk/services/NetTurnP/>). From the secondary structure analysis, it was analyzed that, the EV-3 had the highest percentage of the amino acids (35.2%) in the coil formation and EV-2 the highest percentage of amino acids (53.2%) in the beta-strand formation. However, EV-1 had the highest percentage of 14.5% of amino acids in the alpha-helix formation (**Figure 05** and **Table 12**).

The 3D structures of the vaccine constructs were predicted by the online server RaptorX (<http://raptorx.uchicago.edu/>). All the three vaccines had 5 domains and EV-1 had the lowest p-value of $4.94e-12$. The homology modeling of the three dengue vaccine constructs were carried out using 5HJ3C as template from protein data bank (<https://www.rcsb.org/>). The results of the 3D structure analysis are listed in **Table 13** and illustrated in **Figure 06**.

Table 12. Results of the secondary structure analysis of the vaccine constructs.

Name of the vaccine	Alpha helix (percentage of amino acids)	Beta sheet (percentage of amino acids)	Coil structure (percentage of amino acids)
EV-1	19.8%	65.6%	14.5%
EV-2	27.8%	58.0%	14.2%
EV-3	35.2%	53.2%	11.5%

Table 13. Results of the tertiary structure analysis of the vaccine constructs.

Name of the vaccine	Number of the domains	p-value
EV-1	5	4.94e-12
EV-2	5	1.37e-07
EV-3	5	9.85e-12

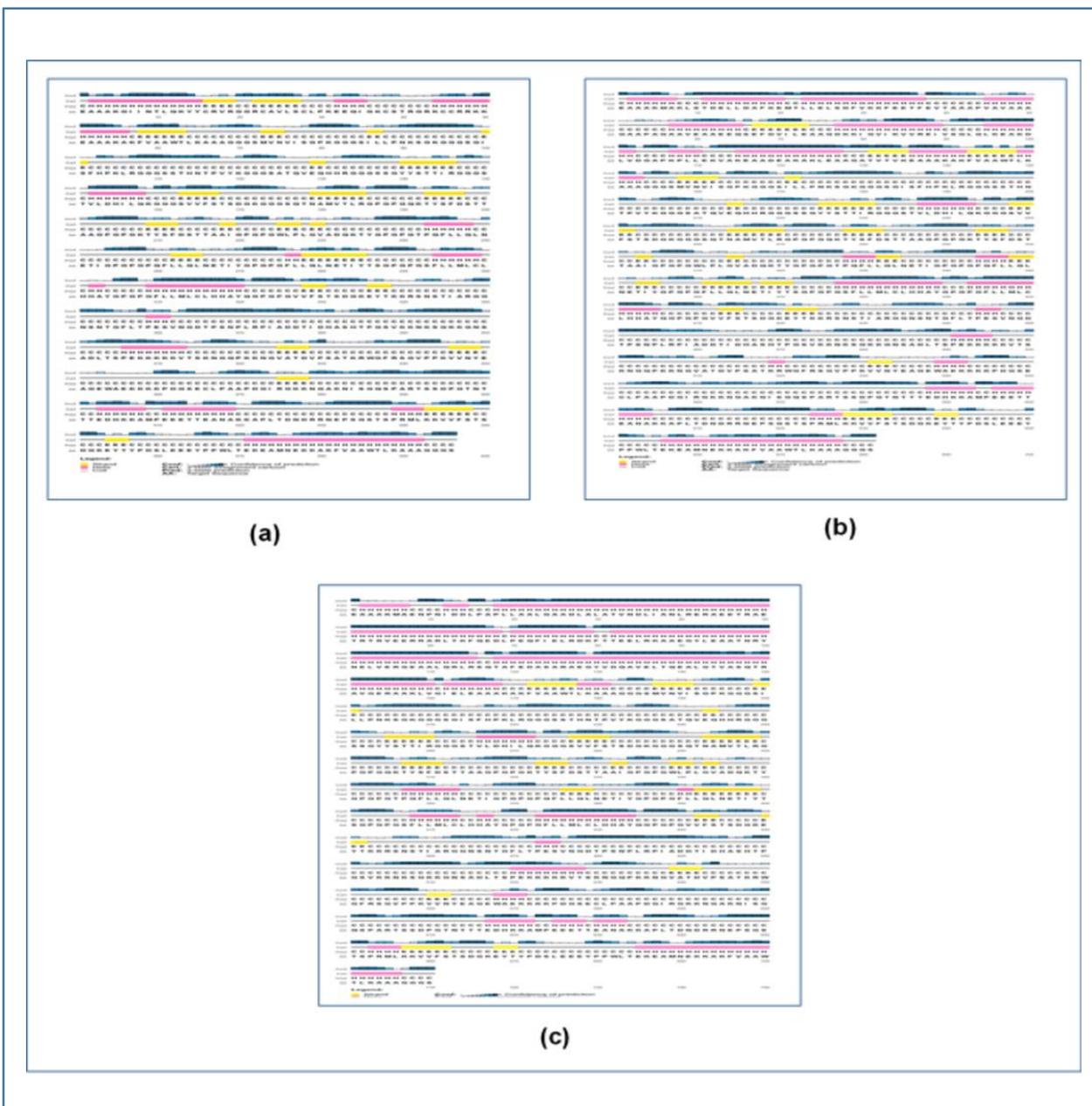


Figure 05. The results of the secondary structure prediction of the constructed HPV-16 vaccines. Here, (a) is EV-1, (b) is EV-2, (c) is EV-3. The secondary structures were predicted using online server PRISPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>).

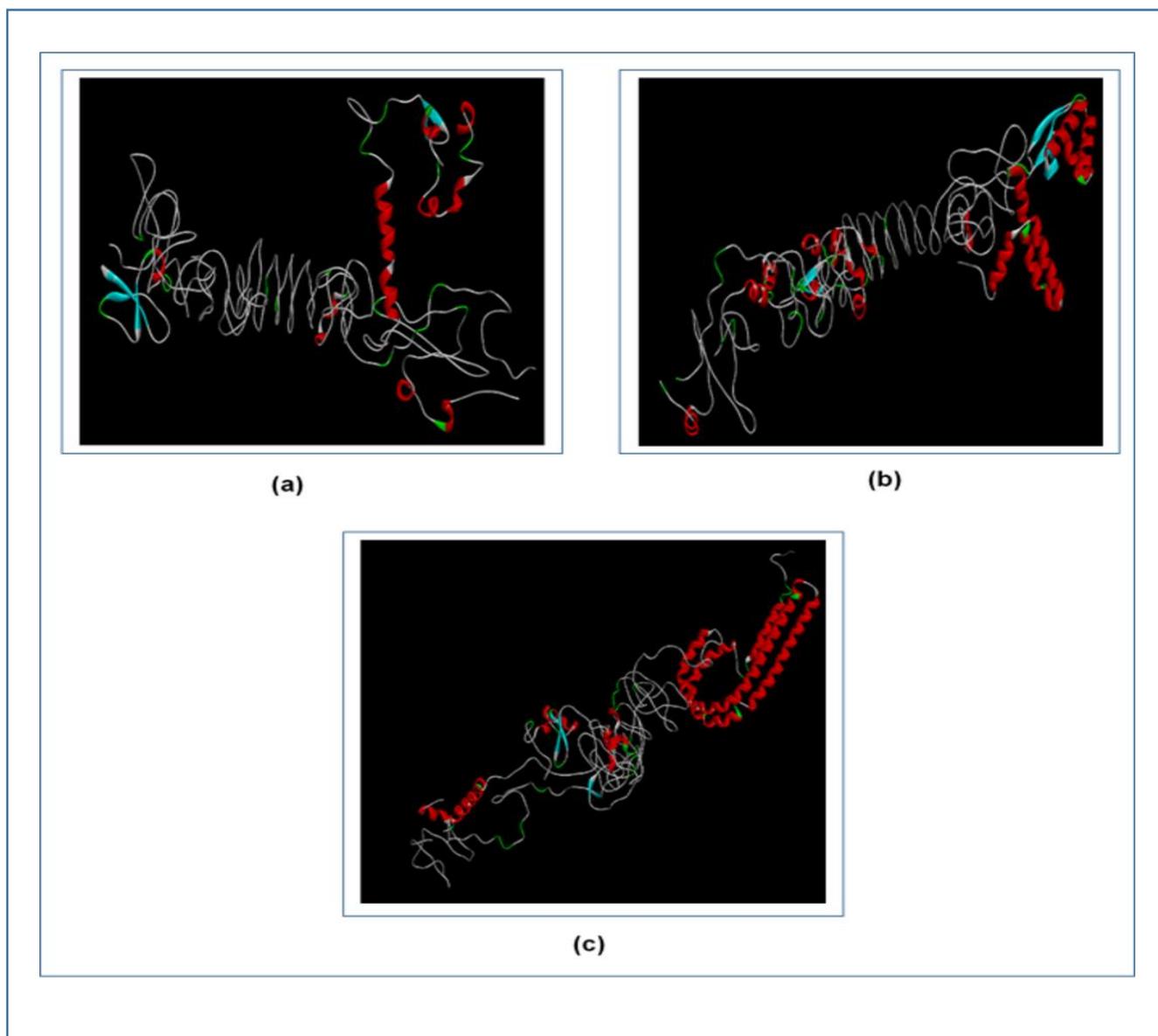


Figure 06. The tertiary structures of the three Ebola vaccines. Here, (a) is the EV-1, (b) is the EV-2 and (c) is the EV-3. The tertiary structures were predicted using the online server tool RaptorX (<http://raptorx.uchicago.edu/>) and visualized by Discovery Studio Visualizer.

3.10. 3D Structure Refinement and Validation

The protein structures generated by the RaptorX server (<http://raptorx.uchicago.edu/>) were refined using 3Drefine (<http://sysbio.rnet.missouri.edu/3Drefine/>) and then the refined structures were validated by Ramachandran plot generated by PROCHECK server (<https://servicesn.mbi.ucla.edu/PROCHECK/>). The analysis showed that EV-1 vaccine had 65.0% of the amino acids in the most favored region, 31.1% of the amino acids in the additional allowed regions, 3.3% of the amino acids in the generously allowed regions and 0.7% of the amino acids in the disallowed regions. The EV-2 vaccine had 66.0% of the amino acids in the most favored region, 29.9% of the amino acids in the additional allowed regions, 3.4% of the amino acids in the generously allowed regions and 0.7% of the amino acids in the disallowed regions. The EV-3 vaccine had 72.1% of the amino acids in the most favored regions, 24.0% of the amino acids in the additional allowed regions, 2.7% of the amino acids in the generously allowed regions and 1.2% of the amino acids in the disallowed regions (**Figure 07**).

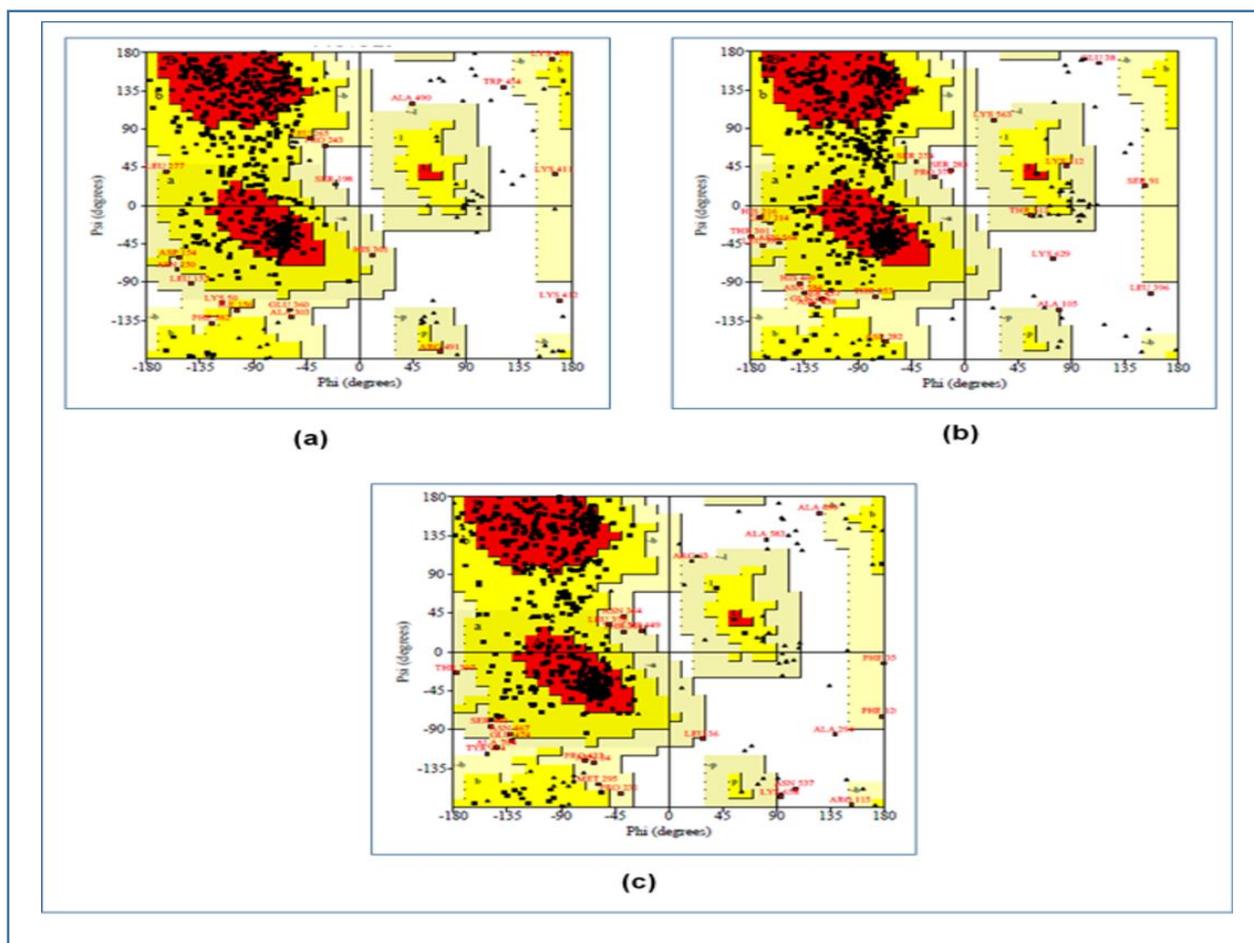


Figure 07. The Ramachandran plot analysis of the three vaccine constructs. (a) EV-1, (b) EV-2, (c) EV-3. The 3D structures of the constructed vaccines were refined using online refinement tool 3Drefine (<http://sysbio.rnet.missouri.edu/3Drefine/>) and validated by analyzing the Ramachandran plot, generated using the online tool PROCHECK (<https://servicesn.mbi.ucla.edu/PROCHECK/>).

3.11. Protein Disulfide Engineering

In protein disulfide engineering, disulfide bonds were generated for the 3D structures of the vaccine constructs. The DbD2 server identifies the pairs of amino acids that have the capability to form disulfide bonds based on the given selection criteria. In this experiment, we selected only those amino acid pairs that had bond energy value less than 2.00 kcal/mol. The EV-1 generated 28 amino acid pairs that had the capability to form disulfide bonds. However, only 6 pairs were selected since they had bond energy less than 2.00 kcal/mol: 20 Gly and 127 Gln, 59 Val and 79 Gly, 101 Ser and 114 His, 195 Ser and 238 Pro, 292 Gly and 293 Ser, 318 His and 364 Gly. EV-2 generated 29 pairs of amino acids that had the capability to form disulfide bonds, however, only 4 pairs were selected: 60 Glu and 67 Glu, 68 Phe and Ala, 240 His and 255 Asp, 498 Val and 560 Arg. EV-3 generated 28 pairs of amino acids capable of forming disulfide bonds and only 5 pairs of the amino acids were selected: 214 Ile and 261 Gly, 229 Asn and 293 Asn, 332 Ala and 344 Val, 438 Gly and 439 Pro, 459 Thr and 476 Val. The selected amino acid pairs formed the mutant version of the original vaccines in the DbD2 server (**Figure 08**).

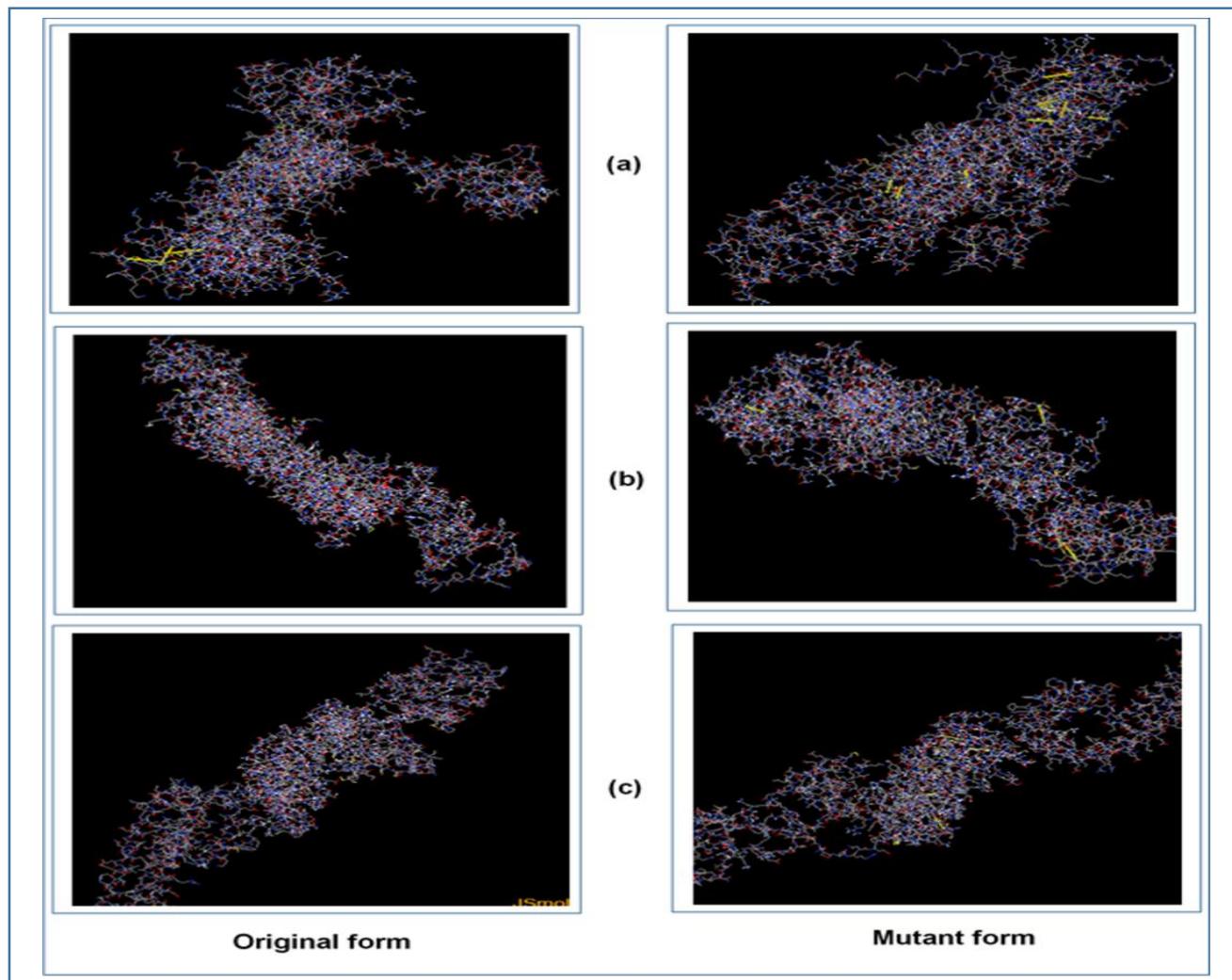


Figure 08. The disulfide engineering of the three vaccine constructs, both the original (left) and mutant (right) forms are shown. Here, (a) EV-1, (b) EV-2, (c) EV-3. The disulfide engineering was conducted using the online tool DbD2 server (<http://cptweb.cpt.wayne.edu/DbD2/>).

3.12. Protein-Protein Docking Study

The protein-protein docking of the vaccine constructs and the MHC alleles were carried out by several online tools for enhancing the accuracy of the prediction: ClusPro 2.0 (<https://cluspro.bu.edu/login.php>), PatchDock (<https://bioinfo3d.cs.tau.ac.il/PatchDock/php.php>) and HawkDock server (<http://cadd.zju.edu.cn/hawkdock/>). The docked complexes that were generated by ClusPro 2.0 and PatchDock were further analyzed by PRODIGY tool of HADDOCK webservice (<https://haddock.science.uu.nl/>) and FireDock server (<http://bioinfo3d.cs.tau.ac.il/FireDock/php.php>), respectively. The PRODIGY server generated binding affinity score (in kcal/mol) and FireDock server generated the global energy of the docked complexes. However, HawkDock generated ranking scores along with the binding free energy (in Kcal/mol). The binding free energy was generated after the MM-GBSA study in the HawkDock server. The vaccine constructs were also docked with the TLR-8.

EV-1 showed the best performances in the results generate by the ClusPro 2.0 and PRODIGY server. It had the lowest binding affinity when docked against DRB3*0202 (-17.3 kcal/mol), DRB1*0101 (-19.2 kcal/mol), DRB3*0101 (-19.2 kcal/mol), DRB1*0401 (-21.5 kcal/mol), DRB1*0301 (-17.9 kcal/mol) and TLR-8 (-21.9 kcal/mol). When the docking was performed by the PatchDock server, EV-3 showed the best performances with the lowest score while docking with DRB3*0202 (-19.59), DRB5*0101 (-23.35), DRB3*0101 (-9.89) and TLR-8 (-22.21). However, EV-1 showed the best performances and lowest scores, when docked against all the selected MHC alleles by the HawkDock server and also when analyzed by the MM-GBSA study. For this reason, EV-1 should be considered as the best vaccine construct among the three vaccine constructs, since it generated the best values among the three vaccines in most of the studies (**Figure 09**). **Table 14** lists the results of the docking study of the three Ebola vaccine constructs.

Table 14. Results of the docking study of all the vaccine constructs.

Name of the vaccines	Name of the Targets	PDB IDs of the targets	Binding affinity, ΔG (kcal mol ⁻¹)	Global energy	HawkDock score (the lowest score)	MM-GBSA (binding free energy, in kcal mol ⁻¹)
EV-1	DRB3*0202	1A6A	-17.3	-3.60	-6220.65	-58.62
	DRB5*0101	1H15	-19.8	1.34	-6003.66	-98.32
	DRB1*0101	2FSE	-19.3	-22.18	-7455.95	-126.19
	DRB3*0101	2Q6W	-19.2	8.86	-6347.92	-74.11
	DRB1*0401	2SEB	-21.5	-33.85	-5793.01	-60.95
	DRB1*0301	3C5J	-17.9	11.21	-6835.73	-69.93
	TLR8	3W3M	-21.9	-2.29	-6148.16	-66.70
EV-2	DRB3*0202	1A6A	-16.9	-18.38	-4926.81	-1.94
	DRB5*0101	1H15	-19.8	-13.53	-4935.45	-71.05
	DRB1*0101	2FSE	-19.2	-18.66	-5684.59	-52.92
	DRB3*0101	2Q6W	-18.9	-2.82	-6046.53	-42.83
	DRB1*0401	2SEB	-21.5	-40.90	-5243.99	-16.20
	DRB1*0301	3C5J	-17.5	-0.08	-4720.13	-36.24
	TLR8	3W3M	-20.7	-5.71	-5295.64	-27.63
EV-3	DRB3*0202	1A6A	-17.1	-19.59	-5205.55	-54.82
	DRB5*0101	1H15	-19.8	-23.35	-4481.78	-52.61
	DRB1*0101	2FSE	-19.3	-21.01	-5585.34	-85.44

	DRB3*0101	2Q6W	-19.0	-9.89	-4905.59	-23.38
	DRB1*0401	2SEB	-21.3	-10.08	-5048.95	-46.65
	DRB1*0301	3C5J	-17.0	9.69	-5153.60	-47.36
	TLR8	3W3M	-21.2	-22.21	-5773.78	-28.08

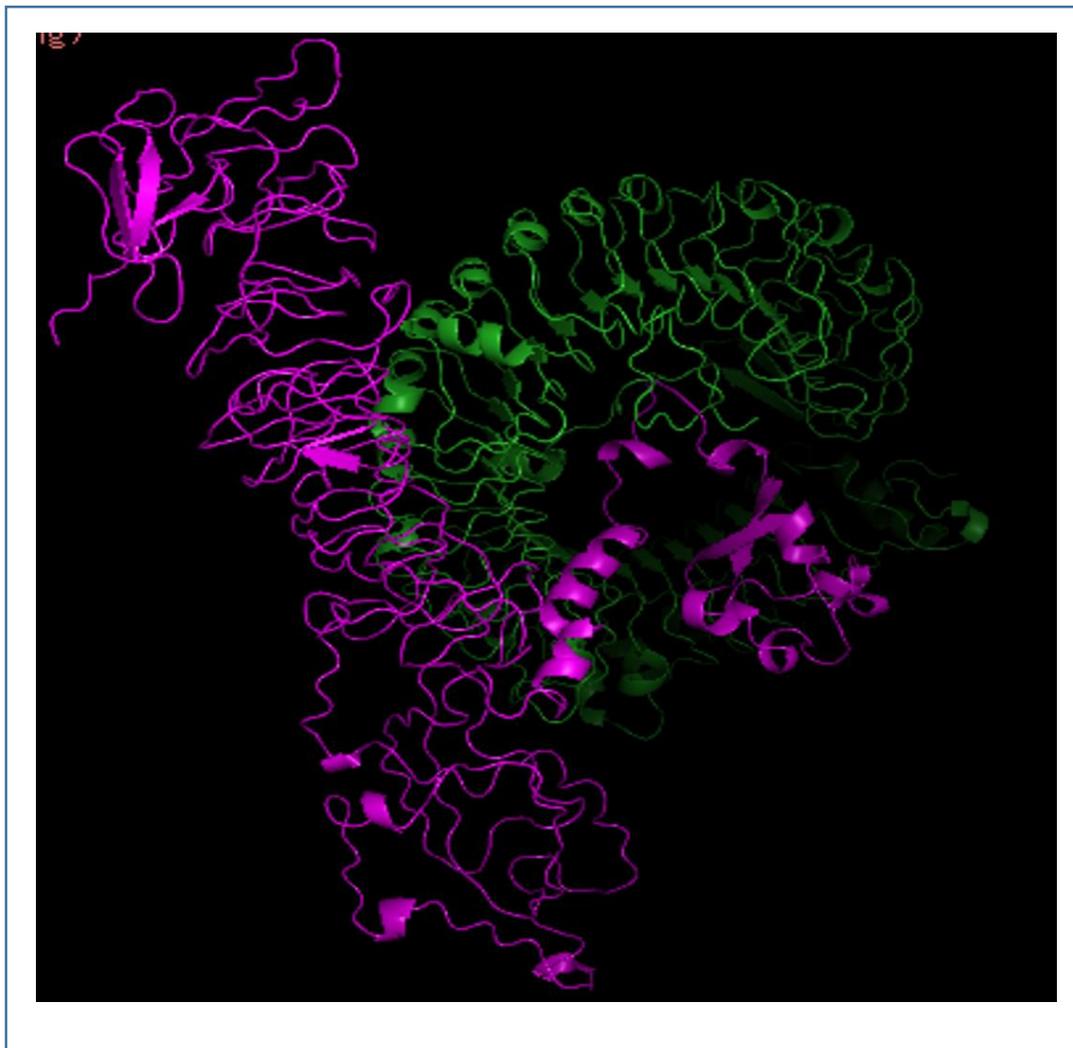


Figure 09. Figure showing the interaction between the ligand protein, EV-1 and receptor protein, TLR-8. The ligand protein is indicated by pink color and the receptor protein is indicated by green color. The visualization of the interactions of the ligands and their receptors were performed by PyMol tool.

3.13. Molecular Dynamics Simulation

Figure 10 illustrates the molecular dynamics simulation and normal mode analysis (NMA) of EV-1-TLR-8 docked complex. The deformability graphs of the complex illustrates the peaks in the graphs which represent the regions of the protein with deformability (**Figure 10b**). The B-factor graphs of the complexes give easy understanding and visualization of the comparison between the NMA and the PDB field of the docked complex (**Figure 10c**). The eigenvalues of the complex is illustrated in **Figure 10d**. EV-1 and TLR8 docked complex generated eigenvalue of 2.129193×10^{-5} . The variance graph indicates the individual variance by red colored bars and cumulative variance by green colored bars (**Figure 10e**). **Figure 10f** illustrates the co-variance map of the complex where the correlated motion between a pair of residues are indicated by red color, uncorrelated motion is indicated by white color and anti-correlated motion is indicated by blue color. The elastic map of the complex represents the connection between the atoms and darker gray regions indicate stiffer regions (**Figure 10g**) [67, 68, 69].

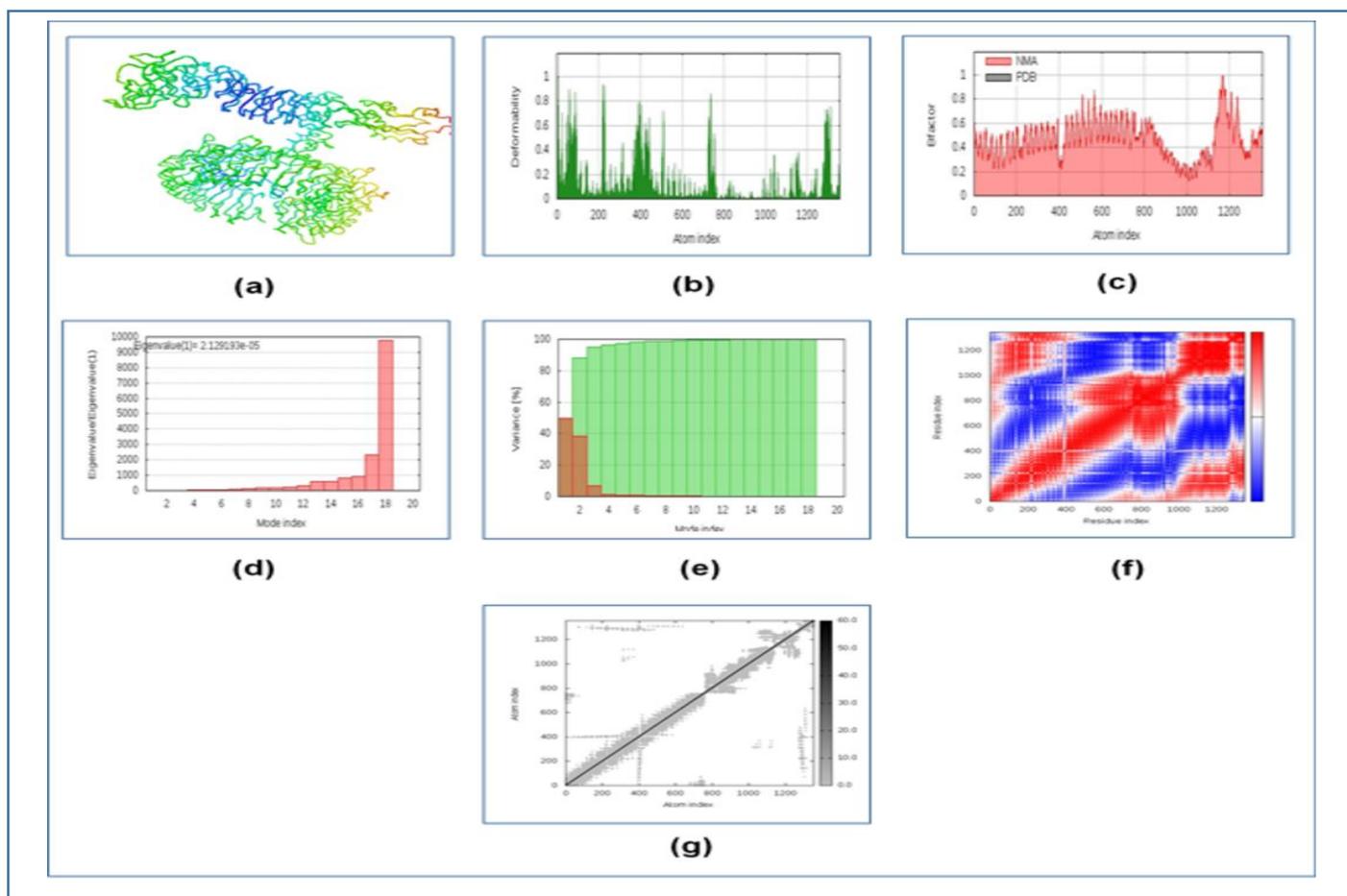


Figure 10. Figure displaying the results of molecular dynamics simulation study of EV-1 and TLR-8 docked complex. Here, (a) NMA mobility, (b) deformability, (c) B-factor, (d) eigenvalues, (e) variance (red color indicates individual variances and green color indicates cumulative variances), (f) co-variance map (correlated (red), uncorrelated (white) or anti-correlated (blue) motions) and (g) elastic network (darker gray regions indicate more stiffer regions).

3.14. Codon Adaptation and In Silico Cloning

For in silico cloning and plasmid construction, the protein sequences of the best selected vaccines were adapted by the JCat server (<http://www.jcat.de/>).

Since the EV-1 protein had 596 amino acids, after reverse translation, the number nucleotides of the probable DNA sequence of EV-1 would be 1788. The codon adaptation index (CAI) value of 0.943 of EV-1 indicated that the DNA sequences contained higher proportion of the codons that are most likely to be present and used in the cellular machinery of the target organism *E. coli* strain K12 (codon biasness). For this reason, the production of the EV-1 vaccine would be carried out efficiently [73, 74]. The GC content of the improved sequence was 50.73%. The predicted DNA sequence of EV-1 was inserted into the pET-19b vector plasmid between the SgrAI and SphI restriction sites. Since the DNA sequence did not have restriction sites for SgrAI and SphI restriction enzymes, SgrAI and SphI restriction sites were conjugated at the N-terminal and C-terminal sites, respectively, before inserting the sequence into the plasmid pET-19b vector. The newly constructed cloned plasmid would be 7360 base pair long, including the constructed DNA sequence of the EV-1 vaccine (the EV-1 vaccine DNA sequence also included the SgrAI and SphI restriction sites) (**Figure 11** and **Figure 12**).

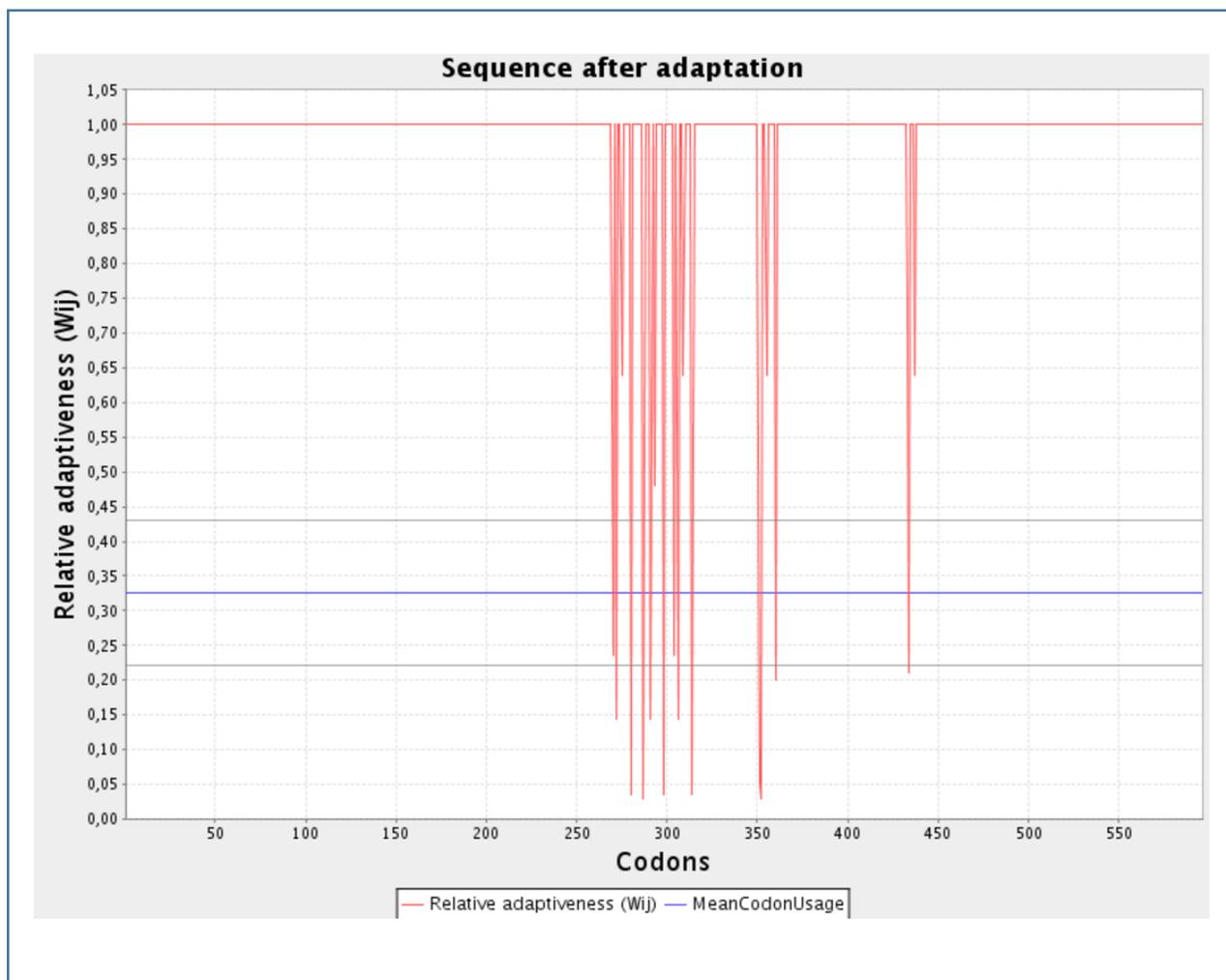


Figure 11. Figure showing the codon adaptation graphs of the EV-1 vaccine. The codon adaptation of the vaccine constructs were carried out using the server JCat (<https://jcatbeauty.com/>).

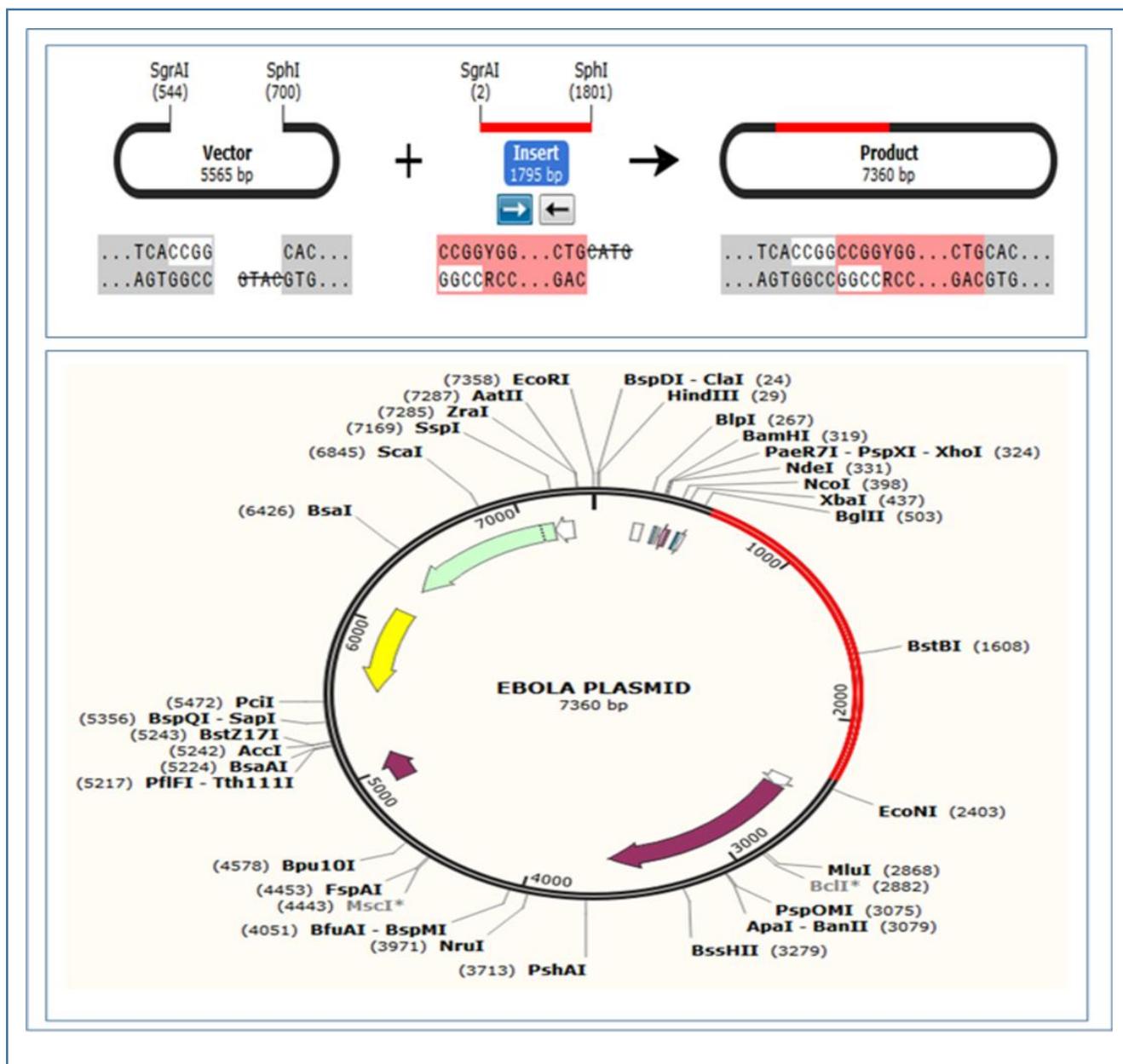


Figure 12. In silico restriction cloning of the EV-1 vaccine sequence in the pET-19b plasmid between the SgrAI and SphI restriction enzyme sites. The red colored marked sites contain the DNA inserts of the vaccines. The cloning was carried out using the SnapGene tool. The two newly constructed plasmids can be inserted into *E. coli* strain K12 for efficient vaccine production.

4. Discussion

Vaccine is one of the most important and widely produced pharmaceutical products. However, the development and production of vaccines, is a costly process. Sometimes, it takes many years to develop a proper vaccine candidate against a particular pathogen. In modern times, various methods and tools of bioinformatics, immunoinformatics and reverse vaccinomics are used for vaccine development, which save time and cost of the vaccine development process [75].

The current study was carried out to design possible vaccines against the Zaire Ebola virus, strain- Mayinga 76. The strain of the Ebola virus, against which the vaccines were designed, was selected by reviewing the NCBI database (<https://www.ncbi.nlm.nih.gov/>). Three proteins of the Zaire Ebola virus strain- Mayinga 76, were selected as the potential targets for the vaccines: matrix protein VP40, envelope glycoprotein and nucleoprotein. The matrix protein VP40 of Ebola virus plays very important roles in the life cycle of a virus, regulating the regulating the viral transcription, virion assembly and budding of viruses from the infected cells [76]. Moreover, the envelope glycoprotein of the Ebola virus mediates the proper viral attachment and entry into the target cell [77]. Studies have also revealed the essential role of nucleoprotein in the replication of Ebola virus [78]. Since these three proteins play very important role in the infection, proliferation and life cycle of the Ebola virus, these proteins could be potential target for vaccine development, thus inhibiting the viral reproduction and infection.

The three protein sequences were retrieved from the UniProt (<https://www.uniprot.org/>) database. The envelope glycoprotein had accession number: Q05320, matrix protein VP40 had accession number: Q05128 and nucleoprotein had accession number: P18272. Various physicochemical properties like number of amino acids, molecular weight, theoretical pI, extinction co-efficient, instability index, aliphatic index, GRAVY were determined by ProtParam

(<https://web.expasy.org/protparam/>) server. The envelope glycoprotein showed the best performances in the physicochemical property analysis. Envelope glycoprotein is the only stable protein among the three selected proteins of the virus, among the three viral proteins.

However, all the proteins were found to be quite antigenic in the bacteria, parasitic and tumor model, where 0.4 threshold was kept. Matrix protein VP40 generated 0.5692 (bacteria), 0.4598 (tumor), 0.5496 (parasite) in the antigenic test. Envelope glycoprotein showed 0.5852 (bacteria), 0.4933 (tumor), 0.5364 (parasite) scores and nucleoprotein generated 0.5315 (bacteria), 0.5442 (tumor), 0.4692 (parasite) scores in the antigenicity test.

The main cells that function in immunity are the T lymphocytic cell and B lymphocytic cell. After recognized by an antigen presenting cell (like macrophage, dendritic cell etc.), the antigen is presented through the MHC class-II molecule present on the surface of the antigen presenting cell, to the helper T cell. Since, the helper T cell contains CD4+ molecule on its surface, it is also known as CD4+ T cell. After activated by the antigen presenting cell, the T-helper cell then activates the B cell and cause the production of memory B cell and antibody producing plasma B cell. The plasma B cell produce a large number of antibodies and the memory B cell functions as the immunological memory. However, T-helper cell also activates macrophage and CD8+ cytotoxic T cell which destroy the target antigen [79, 80, 81, 82, 83].

The possible T cell and B cell epitopes of the selected viral proteins were determined by the IEDB (<https://www.iedb.org/>) server. These epitopes should be able to induce strong immune response against the vaccines. The IEDB server generates and ranks the T cell epitopes based on their antigenicity scores (AS) and percentile scores. Based on analyzing AS and percentile scores, ten of the top twenty MHC class-I epitopes or CTL epitopes were selected randomly for further analysis. On the other hand, MHC class-II epitopes or HTL epitopes were also selected based on

analyzing the AS or percentile scores. Like the MHC class-I epitopes, ten MHC class-II epitopes were randomly selected from the top twenty epitopes. However, the B cell epitopes that had sequence length of 10 amino acids or over 10 amino acids, were selected for further analysis. The transmembrane topology of the epitopes were determined to identify whether the epitopes would be present inside or outside of the cell membrane. The transmembrane topology of the selected epitopes were performed using TMHMM v2.0 server (<http://www.cbs.dtu.dk/services/TMHMM/>).

Antigenicity can be defined as the ability of a foreign substance to act as antigen and activate and stimulate the T cell and B cell responses, through their antigenic determinant portion or epitope [84]. The allergenicity of a substance refers to the ability of that substance to act as allergen and induce potential allergic reactions within the body [85]. When designing a vaccine, epitopes of a protein that remain conserved across various strain, are given much priority than genomic regions that are highly variable because the conserved epitopes of protein(s) provide broader protection across various strains and species [86]. When selecting the best epitopes for vaccine construction, some criteria were maintained: the epitopes should be highly antigenic, so that they can induce high antigenic response, the epitopes should be non-allergenic in nature, for this reason, they would not be able to induce any allergenic reaction in an individual and the epitopes should be non-toxic. The epitopes with 100% conservancy and over 50% minimum identity were selected for vaccine construction, so that, the conserved epitopes would be able to provide protection against various strains. The conservancy analysis was carried out using Zaire Ebola virus strains- Gabon 94 and Kikwit 95, for comparison. Since the selected epitopes were 100% conserved across these strains, for this reason, the vaccine that was constructed for ebola virus strain- Mayinga 76, should provide immunity against the Gabon 94 and Kikwit 95 strains too. Moreover, the cluster analysis of the

MHC class-I alleles and MHC class-II alleles were also carried out to determine their relationship with each other and cluster them functionally based on their predicted binding specificity [87].

In the next step, the protein-peptide docking was carried out between the epitopes and the MHC alleles. The MHC class-I epitopes were docked with the MHC class-I allele (PDB ID: 5WJL) and the MHC class-II epitopes were docked with the MHC class-II allele (PDB ID: 5JLZ). The protein-peptide docking was performed to determine the ability of the epitopes to bind with their respective MHC allele. The 3D structures of the selected epitopes were generated by PEP-FOLD3 (<http://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3/>) server. After that, the generated 3D structures were docked with HLA-A*11-01 allele (PDB ID: 5WJL) and HLA DRB1*04-01 (PDB ID: 5JLZ) using PatchDock (<https://bioinfo3d.cs.tau.ac.il/PatchDock/php.php>) server and refined by FireDock server (<http://bioinfo3d.cs.tau.ac.il/FireDock/php.php>). GISFHPKLR showed the best result with the lowest global energy of -31.68 among the MHC class-I epitopes of matrix protein VP40. SGYYSTTIR generated the lowest and best global energy score of -42.87 among the MHC class-I epitopes of envelope glycoprotein. QTNAMVTLR generated the best global energy score of -45.40 of the MHC class-I epitopes of nucleoprotein. On the other hand, among the MHC class-II epitopes of matrix protein VP40, KTYSFDSTTAAI generated the best global energy score of -24.32. TPQFLLQLNETI generated the lowest global energy of -2.80 among the MHC class-II epitopes of envelope glycoprotein. Moreover, among the MHC class-II epitopes of nucleoprotein, VVFSTSDGKEYT generated the lowest global energy of -12.06.

After the successful protein-peptide docking, the vaccine construction step was carried out. Three different vaccines were constructed and the three vaccines differed from each other only in their adjuvant sequence. The three vaccines were designated as EV-1, EV-2 and EV-3. EV-1 had beta

defensing adjuvant, EV-2 had L7/L12 ribosomal protein as adjuvant and EV-3 had HABA protein as adjuvant. The PADRE sequence was added after the adjuvant sequence and then the CTL, HTL and BCL epitopes were conjugated sequentially. EAAAK, GGGG, GPGPG, KK linkers were used to conjugate the epitopes as well as the adjuvants and PADRE sequence.

After the vaccine construction was performed, the antigenicity, allergenicity and physicochemical property analysis were carried out. All the vaccine constructs were proved to be potential antigenic, however, they were also non-allergenic, for this reason, they should not cause any allergenic reaction within the body. The instability index of a compound refers to the probability of the compound to be stable. IF a compound had instability index of over 40, then the compound is considered to be unstable [88]. The extinction coefficient refers to the amount of light, that is absorbed by a compound at a certain wavelength [89, 90]. EV-3 had the highest extinction coefficient of $59945 \text{ M}^{-1} \text{ cm}^{-1}$. The aliphatic index of a protein corresponds to the relative volume occupied by the aliphatic amino acids in the side chains like alanine, valine etc. [91]. EV-2 had the highest aliphatic index among the vaccine constructs. Since all the vaccine constructs had instability index of less than 40, for this reason, all of them were considered as stable compounds. All the three vaccine constructs had the similar half-life of 1 hour in the mammalian cell culture and all of them were stable. All of the constructs had quite similar theoretical pI. Quite similar performances were observed by the three vaccine constructs in the physicochemical property analysis.

The secondary structures of the vaccine constructs were generated by two online servers, PRISPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>) and NetTurnP v1.0 (<http://www.cbs.dtu.dk/services/NetTurnP/>) and the 3D structures were generated by the server RaptorX (<http://raptorx.uchicago.edu/>) server. EV-3 had the highest percentage of amino acids in

the alpha-helix conformation (35.2%). EV-1 had the highest percentage of amino acids in the beta-sheet conformation with 65.6% of the amino acids in the beta-sheet formation. EV-3 also had the lowest percentage of amino acids (11.5%) in the coil structure of the protein. All the vaccine constructs had 5 domains and EV-2 had the highest p-value of 1.37e-07, determined by the RaptorX server.

In the next step, the 3D structures were refined by online tool 3Drefine (<http://sysbio.rnet.missouri.edu/3Drefine/>) and validated by PROCHECK (<https://servicesn.mbi.ucla.edu/PROCHECK/>) server. EV-3 showed the best performances in the validation experiment with 72.1% of the amino acids in the most favored regions and 24.0% of the amino acids in the additional allowed regions. Moreover, EV-3 had 2.7% of the amino acids in the generously allowed regions and 1.2% of the amino acids in the disallowed regions. After validation of the 3D protein structures, the disulfide engineering of the vaccine constructs were performed using Disulfide by Design 2 v12.2 (<http://cptweb.cpt.wayne.edu/DbD2/>) server and the amino acid pairs with binding energy value less than 2.0 kcal/mol, were picked up for disulfide bond formation.

To find out the best vaccine construct, the protein-protein docking the vaccine constructs and several MHC alleles: DRB1*0101 (PDB ID: 2FSE), DRB3*0202 (PDB ID: 1A6A), DRB5*0101 (PDB ID: 1H15), DRB3*0101 (PDB ID: 2Q6W), DRB1*0401 (PDB ID: 2SEB), and DRB1*0301 (PDB ID: 3C5J) were conducted. The constructed vaccines were also docked against TLR-8 (PDB ID: 3W3M). The aim of the docking experiment is to find out the best vaccine construct among the constructed vaccines. For enhancing the accuracy of our prediction, three different online

servers, ClusPro 2.0 (<https://cluspro.bu.edu/login.php>), PatchDock (<https://bioinfo3d.cs.tau.ac.il/PatchDock/php.php>) and HawkDock server (<http://cadd.zju.edu.cn/hawkdock/>) were used for the docking study. The results of the ClusPro 2.0 and PatchDock were refined and analyzed by PRODIGY tool of HADDOCK webserver (<https://haddock.science.uu.nl/>) and FireDock server (<http://bioinfo3d.cs.tau.ac.il/FireDock/php.php>), respectively. EV-1 showed the best and excellent performances in the results generated by HADDOCK and HawkDock servers as well as in the MM-GBSA study. For this reason, EV-1 was considered the best vaccine construct which was selected for further analysis.

The molecular dynamics simulation study was carried out using the docked TLR-8 and EV-1 complex using the online tool iMODS (<http://imods.chaconlab.org/>). The complex showed quite good results in the molecular dynamics simulation study. The study showed that the complex had quite lesser chance of deformability with quite high eigenvalue of $2.129193e-05$. The deformability graph (**Figure 10b**) also confirmed that the location of the hinges in the structures were not quite significant and the complex is quite stable with lower degree of deforming for each individual amino acids residue. However, EV-1 had good number of correlated amino acids (marked by red color) and also showed a large number stiffer regions (marked by darker gray color) (**Figure 10f**, **Figure 10g**, **Figure 10f** and **Figure 10g**). Nevertheless, it can be concluded that, the complex showed good results in the molecular dynamics simulation study.

Finally, the designed EV-1 vaccine was adapted for in silico cloning. After reverse transcribed, the codon adaptation was carried for cloning in to *E. coli* strain K12. After codon adaptation, the CAI

value of 0.943 was obtained, which supported that, the codon adaptation was carried out perfectly and the DNA sequence contained the codons that are most likely used by the *E. coli* cells. The vaccine DNA sequence was inserted into the SgrAI and SphI restriction sites of the pET-19b plasmid vector for efficient expression and production of the desired vaccine, EV-1 in *E. coli* cells. This insertion produced 7360 base pair long plasmid. The plasmid could be used for producing EV-1 by inserting into the *E. coli* cells. However, more in vitro and in vivo studies should be conducted to finally confirm the safety and efficacy of the predicted best considered vaccines. Moreover, since other vaccines also showed quite good results, they can also be tested to finally confirm the best vaccines among the predicted three vaccines.

5. Conclusion

Ebola virus disease or EVD is a type of haemorrhagic fever that is caused by the infection of Ebola virus. With the high mortality rate, Ebola can be considered as one of the fatal viruses in the world. However, no vaccines are yet discovered with satisfactory results, to control the outbreak of Ebola. In this study, a possible subunit vaccine to fight against Ebola was designed using various tools of bioinformatics, immunoinformatics and vaccinomics. Bioinformatics and related technologies are heavily used in drug design and development since these technologies reduce the time and cost of drug R&D processes. In this study, first the potential proteins of the viral structure are identified. Next, the potential epitopes were identified through robust processes and these epitopes were used for vaccine construction. Three possible vaccines were constructed and by conducting the docking experiments, one best vaccine construct was identified. Later, the molecular dynamics simulation study and codon adaptation as well as in silico cloning were performed. However, wet lab researches should be carried out on the findings of this experiment to finally confirm the safety,

efficacy and potentiality of the vaccine constructs. Hopefully, this research will raise interests among the scientists of the respective field.

6. Acknowledgements

Authors are thankful to Swift Integrity Computational Lab, Dhaka, Bangladesh, a virtual platform of young researchers, for providing the tools.

7. Conflict of interest

Md. Asad Ullah declares that he has no conflict of interest. Bishajit Sarkar declares that he has no conflict of interest. Syed Sajidul Islam declares that he has no conflict of interest.

8. References

[1] Nabel GJ, Yang ZY, Sullivan N, Sanchez A, inventors; US Department of Health, Human Services (HHS), assignee. Development of a preventive vaccine for filovirus infection in primates. United States patent US 8,124,592. 2012.

[2] Sanchez A, Trappier SG, Mahy BW, Peters CJ, Nichol ST. The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing. Proc Natl Acad Sci 1996;16;93:3602-7. doi.org/10.1073/pnas.93.8.3602

- [3] Pourrut, X., Kumulungui, B., Wittmann, T., Moussavou, G., Délicat, A., Yaba, P., ... Leroy, E. M. (2005). The natural history of Ebola virus in Africa. *Microbes and Infection*, 7(7-8), 1005–1014. doi:10.1016/j.micinf.2005.04.006
- [4] Leroy EM, Kumulungui B, Pourrut X, Rouquet P, Hassanin A, Yaba P, Délicat A, Paweska JT, Gonzalez JP, Swanepoel R. Fruit bats as reservoirs of Ebola virus. *Nature* 2005;438:575. doi:10.1038/438575a
- [5] Sullivan NJ, Sanchez A, Rollin PE, Yang ZY, Nabel GJ. Development of a preventive vaccine for Ebola virus infection in primates. *Nature* 2000;408:605. doi:10.1038/35046108
- [6] Kiley MP, Bowen ET, Eddy GA, Isaäcson M, Johnson KM, McCormick JB, Murphy FA, Pattyn SR, Peters D, Prozesky OW, Regnery RL. Filoviridae: a taxonomic home for Marburg and Ebola viruses?. *Intervirology* 1982;18:24-32. doi.org/10.1159/000149300
- [7] Feldmann H, Geisbert TW. Ebola haemorrhagic fever. *The Lancet* 2011;377:849-62. doi.org/10.1016/S0140-6736(10)60667-8
- [8] Volchkov V, Volchkova V, Eckel C, Klenk HD, Bouloy M, LeGuénno B, Feldmann H. Emergence of subtype Zaire Ebola virus in Gabon. *Virology* 1997;232:139-44. doi.org/10.1006/viro.1997.8529
- [9] Leroy EM, Baize S, Volchkov VE, Fisher-Hoch SP, Georges-Courbot MC, Lansoud-Soukate J, Capron M, Debre P, Georges AJ, McCormick JB. Human asymptomatic Ebola infection and strong inflammatory response. *The Lancet* 2000;355:2210-5. doi.org/10.1016/S0140-6736(00)02405-3

- [10] Wamala JF, Lukwago L, Malimbo M, Nguku P, Yoti Z, Musenero M, Amone J, Mbabazi W, Nanyunja M, Zaramba S, Opio A. Ebola hemorrhagic fever associated with novel virus strain, Uganda, 2007–2008. *Emerg Infect Dis* 2010;16:1087. doi: 10.3201/eid1607.091525
- [11] World Health Organization, 2014. Experimental Ebola Vaccines: WHO Consultation on Ebola Vaccines.
- [12] Regules JA, Beigel JH, Paolino KM, Voell J, Castellano AR, Hu Z, Muñoz P, Moon JE, Ruck RC, Bennett JW, Twomey PS. A recombinant vesicular stomatitis virus Ebola vaccine. *N Engl J Med* 2017;376:330-41. DOI: 10.1056/NEJMoa1414216
- [13] Kanopathipillai R, Henao Restrepo AM, Fast P, Wood D, Dye C, Kieny MP, Moorthy V. Ebola vaccine—an urgent international priority. *N Engl J Med* 2014;371:2249-51. DOI: 10.1056/NEJMp1412166
- [14] Chong, L.C. and Khan, A.M., 2019. Vaccine Target Discovery.
- [15] Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* 2007a;8:4. doi.org/10.1186/1471-2105-8-4
- [16] Doytchinova IA, Flower DR. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. *Vaccine* 2007b;25:856-66. doi.org/10.1016/j.vaccine.2006.09.032
- [17] Doytchinova IA, Flower DR. Bioinformatic approach for identifying parasite and fungal candidate subunit vaccines. *Open Vaccine J* 2008;1:4.

- [18] Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook* 2005;571-607. Humana press. doi.org/10.1385/1-59259-890-0:571
- [19] Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B. The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res* 2018;47:D339-43. doi.org/10.1093/nar/gky1006
- [20] Möller S, Croning MD, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* 2001;17:646-53. doi.org/10.1093/bioinformatics/17.7.646
- [21] Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v. 2—a server for in silico prediction of allergens. *J Mol Model* 2014;20:2278. doi.org/10.1007/s00894-014-2278-5
- [22] Dimitrov I, Naneva L, Doytchinova I, Bangov I. AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics* 2013;30:846-51. doi.org/10.1093/bioinformatics/btt619
- [23] Thomsen M, Lundegaard C, Buus S, Lund O, Nielsen M. MHCcluster, a method for functional clustering of MHC molecules. *Immunogenetics* 2013;65:655-65. doi.org/10.1007/s00251-013-0714-9
- [24] Lamiable A, Thévenet P, Rey J, Vavrusa M, Derreumaux P, Tufféry P. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res* 2016;44:W449-54. doi.org/10.1093/nar/gkw329
- [25] Shen Y, Maupetit J, Derreumaux P, Tufféry P. Improved PEP-FOLD approach for peptide and miniprotein structure prediction. *J Chem Theory Comput* 2014;10:4745-58. doi.org/10.1021/ct500592m

- [26] Thevenet P, Shen Y, Maupetit J, Guyon F, Derreumaux P, Tuffery P. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Res* 2012;40:W288-93. doi.org/10.1093/nar/gks419
- [27] Duhovny D, Nussinov R, Wolfson HJ. Efficient unbound docking of rigid molecules. In *International workshop on algorithms in bioinformatics 2002 Sep 17* (pp. 185-200). Springer, Berlin, Heidelberg. doi.org/10.1007/3-540-45784-4_14
- [28] Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res* 2005;33:W363-7. doi.org/10.1093/nar/gki481
- [29] Gujjula KR. *Prediction and comparison of hiv-1 protease inhibitor binding energies by various molecular docking methods* (Doctoral dissertation).
- [30] Atapour A, Mokarram P, MostafaviPour Z, Hosseini SY, Ghasemi Y, Mohammadi S, Nezafat N. Designing a fusion protein vaccine against HCV: an in silico approach. *Int J Pept Res Ther* 2019;25:861-72. doi.org/10.1007/s10989-018-9735-4
- [31] Rana A, Akhter Y. A multi-subunit based, thermodynamically stable model vaccine using combined immunoinformatics and protein structure based approach. *Immunobiology* 2016;221:544-57. doi.org/10.1016/j.imbio.2015.12.004
- [32] Hancock RE, Nijnik A, Philpott DJ. Modulating immunity as a therapy for bacterial infections. *Nat. Rev. Microbiol* 2012;10:243. doi.org/10.1038/nrmicro2745
- [33] Toussi DN, Massari P. Immune adjuvant effect of molecularly-defined toll-like receptor ligands. *Vaccines* 2014;2:323-53. doi.org/10.3390/vaccines2020323

- [34] Funderburg N, Lederman MM, Feng Z, Drage MG, Jadowsky J, Harding CV, Weinberg A, Sieg SF. Human β -defensin-3 activates professional antigen-presenting cells via Toll-like receptors 1 and 2. *Proc Natl Acad Sci* 2007;104:18631-5. doi.org/10.1073/pnas.0702130104
- [35] Lee SJ, Shin SJ, Lee MH, Lee MG, Kang TH, Park WS, Soh BY, Park JH, Shin YK, Kim HW, Yun CH. A potential protein adjuvant derived from *Mycobacterium tuberculosis* Rv0652 enhances dendritic cells-based tumor immunotherapy. *PloS one* 2014;9:e104351. doi.org/10.1371/journal.pone.0104351
- [36] Hajighahramani N, Nezafat N, Eslami M, Negahdaripour M, Rahmatabadi SS, Ghasemi Y. Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against *Staphylococcus aureus*. *Infect Genet Evol* 2017;48:83-94. doi.org/10.1016/j.meegid.2016.12.010
- [37] Pandey RK, Sundar S, Prajapati VK. Differential expression of miRNA regulates T cell differentiation and plasticity during visceral leishmaniasis infection. *Front Microbiol* 2016;7:206. doi.org/10.3389/fmicb.2016.00206
- [38] Wu CY, Monie A, Pang X, Hung CF, Wu TC. Improving therapeutic HPV peptide-based vaccine potency by enhancing CD4+ T help and dendritic cell activation. *J Biomed Sci* 2010;17:88. doi.org/10.1186/1423-0127-17-88
- [39] Saha S, Raghava GP. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res* 2006;34:W202-9. doi.org/10.1093/nar/gkl343
- [40] Buchan DW, Jones DT. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res* 2019;47:W402-7. doi.org/10.1093/nar/gkz297

- [41] Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 1999;292:195-202. doi.org/10.1006/jmbi.1999.3091
- [42] Petersen B, Lundegaard C, Petersen TN. NetTurnP—neural network prediction of beta-turns by use of evolutionary information and predicted protein sequence features. *PLoS One* 2010;5:e15079. doi.org/10.1371/journal.pone.0015079
- [43] Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* 2012;7:1511. doi.org/10.1038/nprot.2012.085
- [44] Ma J, Wang S, Zhao F, Xu J. Protein threading using context-specific alignment potential. *Bioinformatics* 2013;29:i257-65. doi.org/10.1093/bioinformatics/btt210
- [45] Peng J, Xu J. RaptorX: exploiting structure information for protein alignment by statistical inference. *Proteins: Struct, Funct, Bioinf* 2011;79:161-71. doi.org/10.1002/prot.23175
- [46] Bhattacharya D, Nowotny J, Cao R, Cheng J. 3Drefine: an interactive web server for efficient protein structure refinement. *Nucleic Acids Res* 2016;44:W406-9. doi.org/10.1093/nar/gkw336
- [47] Laskowski RA, MacArthur MW, Thornton JM. PROCHECK: validation of protein-structure coordinates. *Int Tab Crystallography* 2006;F:722-725.
- [48] Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. *Proteins: Struct, Funct, Bioinf* 1992;12:345-64. doi.org/10.1002/prot.340120407
- [49] Craig DB, Dombkowski AA. Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinformatics* 2013;14:346. doi.org/10.1186/1471-2105-14-346

[50] Stern LJ, Calvo-Calle JM. HLA-DR: molecular insights and vaccine design. *Curr Pharm Des* 2009;15:3249-61.

[51] Thompson JM, Iwasaki A. Toll-like receptors regulation of viral infection and disease. *Adv. Drug Deliv Rev* 2008;60:786-94. doi.org/10.1016/j.addr.2007.11.003

[52] Lester SN, Li K. Toll-like receptors in antiviral innate immunity. *J Mol Biol* 2014;426:1246-64. doi.org/10.1016/j.jmb.2013.11.024

[53] Regnery RL, Johnson KM, Kiley MP. Virion nucleic acid of Ebola virus. *J Virol* 1980;36:465-469.

[54] Vajda S, Yueh C, Beglov D, Bohnuud T, Mottarella SE, Xia B, Hall DR, Kozakov D. New additions to the ClusPro server motivated by CAPRI. *Proteins: Struct, Funct, Bioinf* 2017;85:435-44. doi.org/10.1002/prot.25219

[55] Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, Beglov D, Vajda S. The ClusPro web server for protein–protein docking. *Nat Protoc* 2017;12:255. doi.org/10.1038/nprot.2016.169

[56] Kozakov D, Beglov D, Bohnuud T, Mottarella SE, Xia B, Hall DR, Vajda S. How good is automated protein docking?. *Proteins: Struct, Funct, Bioinf* 2013;81:2159-66. doi.org/10.1002/prot.24403

[57] Vangone A, Bonvin AM. Contacts-based prediction of binding affinity in protein–protein complexes. *Elife* 2015;4:e07454.

- [58] Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM, Vangone A. PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics* 2016;32:3676-8. doi.org/10.1093/bioinformatics/btw514
- [59] Yuriev E, Ramsland PA. Latest developments in molecular docking: 2010–2011 in review. *J Mo Recog* 2013;26:215-39. doi.org/10.1002/jmr.2266
- [60] Weng G, Wang E, Wang Z, Liu H, Zhu F, Li D, Hou T. HawkDock: a web server to predict and analyze the protein–protein complex based on computational docking and MM/GBSA. *Nucleic Acids Res* 2019. doi.org/10.1093/nar/gkz397
- [61] Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J Chem Inf Model* 2010;51:69-82.
- [62] Feng T, Chen F, Kang Y, Sun H, Liu H, Li D, Zhu F, Hou T. HawkRank: a new scoring function for protein–protein docking based on weighted energy terms. *J Cheminformatics* 2017;9:66. doi.org/10.1186/s13321-017-0254-7
- [63] Sun H, Li Y, Tian S, Xu L, Hou T. Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. *Phys Chem Chem Phys* 2014;16:16719-29. DOI: 10.1039/C4CP01388C
- [64] The *PyMOL* Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.
- [65] Awan F, Obaid A, Ikram A, Janjua H. Mutation-structure-function relationship based integrated strategy reveals the potential impact of deleterious missense mutations in autophagy

related proteins on hepatocellular carcinoma (HCC): A comprehensive informatics approach. *Int J Mol Sci* 2017;18:139. doi.org/10.3390/ijms18010139

[66] Prabhakar PK, Srivastava A, Rao KK, Balaji PV. Monomerization alters the dynamics of the lid region in *Campylobacter jejuni* CstII: an MD simulation study. *J Biomol Struct Dyn* 2016;34:778-91. doi.org/10.1080/07391102.2015.1054430

[67] López-Blanco JR, Aliaga JI, Quintana-Ortí ES, Chacón P. iMODS: internal coordinates normal mode analysis server. *Nucleic Acids Res* 2014;42:W271-6. doi.org/10.1093/nar/gku339

[68] López-Blanco JR, Garzón JI, Chacón P. iMod: multipurpose normal mode analysis in internal coordinates. *Bioinformatics* 2011;27:2843-50. doi.org/10.1093/bioinformatics/btr497

[69] Kovacs JA, Chacón P, Abagyan R. Predictions of protein flexibility: first-order measures. *Proteins: Struct, Funct, Bioinf* 2004;56:661-8. doi.org/10.1002/prot.20151

[70] Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, Jahn D. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res* 2005;33:W526-31. doi.org/10.1093/nar/gki376

[71] Angov E. Codon usage: nature's roadmap to expression and folding of proteins. *Biotechnol J* 2011;6:650-9. doi.org/10.1002/biot.201000332

[72] Solanki V, Tiwari V. Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against *Acinetobacter baumannii*. *Sci Rep* 2018;8:9044. doi.org/10.1038/s41598-018-26689-7

[73] Carbone A, Zinovyev A, Képes F. Codon adaptation index as a measure of dominating codon bias. *Bioinformatics* 2003;19:2005-15. doi.org/10.1093/bioinformatics/btg272

- [74] Sharp PM, Li WH. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 1987;15:1281-95. doi.org/10.1093/nar/15.3.1281
- [75] María RA, Arturo CV, Alicia JA, Paulina ML, Gerardo AO. The impact of bioinformatics on vaccine design and development. InTech, Rijeka, Croatia; 2017.
- [76] Madara JJ, Han Z, Ruthel G, Freedman BD, Harty RN. The multifunctional Ebola virus VP40 matrix protein is a promising therapeutic target. *Future Virol* 2015;10:537-46. doi.org/10.2217/fvl.15.6
- [77] Martinez O, Ndungo E, Tantral L, Miller EH, Leung LW, Chandran K, Basler CF. A mutation in the Ebola virus envelope glycoprotein restricts viral entry in a host species-and cell-type-specific manner. *J Virol*. 2013;87:3324-34. doi.org/10.1128/jvi.01598-12
- [78] Watanabe S, Noda T, Kawaoka Y. Functional mapping of the nucleoprotein of Ebola virus. *J Virol*. 2006 Apr 15;80(8):3743-51. doi.org/10.1128/jvi.80.8.3743-3751.2006
- [79] Goerdt S, Orfanos CE. Other functions, other genes: alternative activation of antigen-presenting cells. *Immunity* 1999;10:137-42.
- [80] Tanchot C, Rocha B. CD8 and B cell memory: same strategy, same signals. *Nat Immunol* 2003;4:431. doi.org/10.1038/ni0503-431
- [81] Pavli P, Hume DA, Van De Pol E, Doe WF. Dendritic cells, the major antigen-presenting cells of the human colonic lamina propria. *Immunology* 1993;78:132.
- [82] Arpin C, Dechanet J, Van Kooten C, Merville P, Grouard G, Briere F, Banchereau J, Liu YJ. Generation of memory B cells and plasma cells in vitro. *Science* 1995;268:720-2. DOI: 10.1126/science.7537388

- [83] Cano RL, Lopera HD. Introduction to T and B lymphocytes. In *Autoimmunity: From Bench to Bedside* [Internet] 2013. El Rosario University Press.
- [84] Fishman JM, Wiles K, Wood KJ. The acquired immune system response to biomaterials, including both naturally occurring and synthetic biomaterials. In *Host Response to Biomaterials* 2015;151-187. Academic Press. doi.org/10.1016/B978-0-12-800196-7.00008-6
- [85] Andrae D, Nowak-Węgrzyn A. The Effect of Infant Allergen/Immunogen Exposure on Long-Term Health. In *Early Nutrition and Long-Term Health* 2017;131-173. Woodhead Publishing. doi.org/10.1016/B978-0-08-100168-4.00006-9
- [86] Bui HH, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics* 2007;8:361. doi.org/10.1186/1471-2105-8-361
- [87] Thomsen M, Lundegaard C, Buus S, Lund O, Nielsen M. MHCcluster, a method for functional clustering of MHC molecules. *Immunogenetics* 2013;65:655-65. doi.org/10.1007/s00251-013-0714-9
- [88] Guruprasad K, Reddy BB, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng Des Sel* 1990;4:155-61. doi.org/10.1093/protein/4.2.155
- [89] Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* 1995;4:2411-23. doi.org/10.1002/pro.5560041120
- [90] Ikai A. Thermostability and aliphatic index of globular proteins. *J Biochem* 1980;88:1895-8.

[91] Gill SC, Von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* 1989;182:319-26. doi.org/10.1093/oxfordjournals.jbchem.a133168