In silico-based vaccine design against Ebola virus glycoprotein

Raju Dash¹, Rasel Das², Md Junaid³, Md Forhad Chowdhury Akash⁴, Ashekul Islam⁵, SM Zahid Hosen¹

¹Molecular Modeling and Drug Design Laboratory (MMDL), Pharmacology Research Division, Bangladesh Council of Scientific and Industrial Research (BCSIR), Chittagong, Bangladesh; ²Nano-technology and Catalysis Research Center, University of Malaya, Kuala Lumpur, Malaysia; ³Department of Pharmaceutical Sciences, North South University, Dhaka, Bangladesh; ⁴Department of Pharmacy, BGC Trust University Bangladesh, Chittagong, Bangladesh; ⁵Department of Biochemistry and Molecular Biology, University of Chittagong, Chittagong, Bangladesh

Abstract: Ebola virus (EBOV) is one of the lethal viruses, causing more than 24 epidemic outbreaks to date. Despite having available molecular knowledge of this virus, no definite vaccine or other remedial agents have been developed yet for the management and avoidance of EBOV infections in humans. Disclosing this, the present study described an epitope-based peptide vaccine against EBOV, using a combination of B-cell and T-cell epitope predictions, followed by molecular docking and molecular dynamics simulation approach. Here, protein sequences of all glycoproteins of EBOV were collected and examined via silico methods to determine the most immunogenic protein. From the identified antigenic protein, the peptide region ranging from 186 to 220 and the sequence HKEGAFFLY from the positions of 154–162 were considered the most potential B-cell and T-cell epitopes, correspondingly. Moreover, this peptide (HKEGAFFLY) interacted with HLA-A*32:15 with the highest binding energy and stability, and also a good conservancy of 83.85% with maximum population coverage. The results imply that the designed epitopes could manifest vigorous enduring defensive immunity against EBOV.

Keywords: Ebola virus, epitope, glycoprotein, vaccine design

Introduction

Ebola virus (EBOV) is an antisense-strand RNA virus from the Filoviridae family, and it is structurally filamentous.¹ Although the initial discovery of EBOV was in 1976, till now more than 24 epidemics have been reported from Africa, mostly with the Zaire species (http://who.int/mediacentre/factsheets/fs103/en/).²–⁴ The genome of EBOV enciphers the seven structural proteins, ie, nucleoprotein (NP), viral structural proteins (VP35, VP40, VP30, and VP24), glycoprotein (GP), and RNA-dependent RNA polymerase (L).⁵

Among these, three different versions of glycoprotein are transcribed by the GP gene.⁶–⁸ Both attachment protein (GP1) and entry/fusion protein (GP2) are expressed from the full length of the GP chain, which are synthesized from messenger RNAs (mRNAs), containing an additional nontemplated adenosine. The soluble GP (sGP) is synthesized from the unedited RNA transcript. On the contrary, small soluble GP (ssGP) is translated during this process by adding two additional adenosine residues.¹⁰ The GPs are expressed virally on the virion surface, which plays a crucial role in the catalysis of membrane fusion and amalgamation to host cells. As a result, it is considered not only a crucial component for vaccines but also an essential target for developing inhibitors and antibodies of attachment and fusion.¹¹–¹³
With the advances in genomics, proteomics, and the understanding of pathogens, the field of viral vaccine preparation has been recently expanded by a most promising approach, known as epitope-based vaccine design. Epitope represents the negligible immunogenic region of a protein sequence, which specifically elicits accurate immune responses. Various studies recently reported that the vaccination process based on epitope efficiently educes defensive immune responses against diverse pathogens. In this context, prediction of epitopes via in silico tools in vaccine designing process can significantly minimize the time and cost required in the development process.

Thereby, based on the available GP sequences of EBOV, this study attempted to design effective epitope-based peptide vaccines (T-cell and B-cell epitope) using various in silico tools. These results offer new epitope vaccine candidates for vaccine development against EBOV.

Materials and methods

The methodologies used for peptide vaccine development are shown in Figure 1.

Protein sequence retrieval, evaluation analysis, and antigenic protein identification

All available sequences of the GP of EBOV were extracted from the UniProt database. After that, multiple sequence alignment was performed by using the ClustalW2 tool, and a phylogenetic tree was assembled by MEGA 6.0 software. And then, VaxiJen v2.0 was used to predict most efficient antigenic protein from the available protein sequences.

Figure 1 Graphical depiction of the methodologies used in peptide vaccine design.

Abbreviations: EBOV, Ebola virus; GP, glycoprotein; RMSF, root mean square function; RMSD, root mean square deviation; MD, molecular dynamics; HLA, human leukocyte antigen; IEDB, immune epitope database.
T-cell epitope identification and conservancy analysis

T-cell identification was done using the NetCTL 1.2 server,\textsuperscript{22} setting thresholds at 0.5, 0.89, and 0.94 for sensitivity and accuracy. MHC-I binding of the identified epitopes and epitope conservancy were then calculated using tools from the immune epitope database (IEDB).\textsuperscript{24–26} These tools calculate the half maximal inhibitory concentration ($IC_{50}$) value of epitope binding to human leukocyte antigen (HLA) molecules using the stabilized matrix base method.\textsuperscript{26,27} The restriction for epitope identification was set to 12 MHC-I supertypes. Prior to the run, all the alleles were considered, and the length of the peptides was set at 9.0.

Prediction of population coverage and allergenicity assessment

The population coverage tool from IEDB was applied to determine the population coverage for every single epitope by selecting HLA alleles of the corresponding epitope. Allergenicity of the predicted epitope was calculated using AllerHunter,\textsuperscript{27} which can predict both nonallergens and allergens with a high level of accuracy, by comparing the input sequence with the sequence of known allergens.\textsuperscript{29}

Molecular simulation analysis of HLA allele interaction

Design of the three-dimensional structure of epitope and HLA protein

The three-dimensional structures of all the five epitopes were predicted by a PEP-FOLD web-based server.\textsuperscript{30} For each sequence, this server predicted the five most probable structures, the best of which, having the lowest energy model, was chosen for further analysis.

To validate the binding of identified epitope and HLA molecule, we considered the homology modeling as there is no relevant structure available in the protein data bank. We selected homology modeling using the most popular online tools. Molecules were then analyzed using the stabilizer matrix base method.\textsuperscript{26,27} The restriction for epitope identification was set to 12 MHC-I supertypes. Prior to the run, all the alleles were considered, and the length of the peptides was set at 9.0.

Docking analysis

Molecular docking analysis was performed using AutoDock Vina,\textsuperscript{39} by considering HLA molecule as a protein and identified epitopes as ligands. First, we used the protein preparation wizard of UCSF Chimera\textsuperscript{40} to prepare the hypothetical protein for docking analysis by adding hydrogens and Gasteiger–Marsili charges.\textsuperscript{41} The prepared file was then converted into pdbqt format. The parameters used for the docking simulation were set to default. The size of the grid box in AutoDock Vina was kept at 36.3095, 54.3374, and 48.025, respectively, for X, Y, and Z. The energy range was kept at 4, according to the default setting. AutoDock Vina was implemented via the shell script offered by AutoDock Vina developers. Docking results were observed by negative score in kcal/mol, as binding affinity of ligands.\textsuperscript{39}

Binding energy estimation and molecular dynamics (MD) simulation

The binding free energy of HLA-epitope complexes were calculated by using MM (CHARMM)\textsuperscript{42} – Generalized Born Surface Area (GBSA) and Poisson–Boltzmann Surface Area (PBSA) protocols, implemented in Accelrys Discovery Studio 2.5. Using implicit solvent models of GBSA and PBSA, the binding free energy ($\Delta G_{\text{bind}}$) for each epitope was calculated by maintaining salt concentration of 0.15 M. Default value was set for conformational entropy and ligand minimization. The distance cutoff value was set to 14.0 Å. The binding energy was calculated by using following equation:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - (G_{\text{HLA}} + G_{\text{epitope}})$$

where $G = <G_{\text{int}} > + <G_{\text{intr}} > + <G_{\text{prot}} > + <G_{\text{w}} > - T\Delta S$.

The entire dynamics simulation study for the HLA-epitope complex was accomplished in YASARA Dynamics software. Prior to simulation, the complex was cleaned and optimized the hydrogen bond network.\textsuperscript{43} After that, a cubic simulation cell was created with a periodic boundary condition, and the atoms of the complex were typed using the AMBER14\textsuperscript{44} force field. The pKa (acid dissociation constant) values of protein titratable amino acids were calculated and solvated the simulation box using the transferable intermolecular potential3 points (TIP3P) water model (density: 0.997 g/L). The system consistent with 46406 atoms was energy minimized using the steepest gradient approach (5000 cycles) followed by simulated annealing method. Restrained and unrestrained all-atom molecular dynamics simulation were performed in solvent using the PME method to describe long-range electrostatic interactions at a cut off distance of 8 Å at physiological conditions (298 K, pH 7.4, 0.9% NaCl).\textsuperscript{45} A multiple time step algorithm together with a simulation time step interval of 2.50 fs was chosen.\textsuperscript{46} Molecular dynamics simulations of 100 ns long were performed at constant temperature using a Berendsen thermostat and constant
pressure. The MD trajectories were saved every 250 ps for analysis. The trajectories generated from the simulation were analyzed for the stability by various evaluative measures viz. RMSD, RMSF (RMS fluctuations), and initial and final protein backbone comparisons using YASARA structure built in macros and VMD software.

**Identification of the B-cell epitope**

To detect B-cell epitope, various tools from IEDB were used to identify the B-cell antigenicity, together with the Emini surface accessibility prediction,47 Kolaskar and Tongaonkar antigenicity scale,48 Karplus and Schulz flexibility prediction,49 and Bepipred linear epitope prediction analysis.50 Since antigenic parts of a protein belong to the beta turn regions,51 the Chou and Fasman beta turn prediction tool52 was also used.

**Results**

**Evolutionary analysis of the GP sequences**

A total of 46 GP sequences from the different variants of EBOV were collected from the UniProtKB database. Multiple sequence alignment analysis was then performed, and a phylogenetic tree (Figure 2) was constructed thereby. Using the unweighted pair-group method with arithmetic mean, a phylogram was constructed using the bootstrap with 1,000 replications in MEGA.53 From the multiple sequence alignment analysis, it is clearly seen that protein sequences that isolated from various strains were having a close relationship. Also, from the multiple comparison result, the selected sequences of EBOV of the same subtype have 78%–99% similarity. This result also confers the possibilities of mutation in glycoprotein of all strains, which demonstrates a good agreement with the results from Veljkovic et al.54

**Antigenic protein prediction**

Protein sequences in this study were considered to screen out using VaxiJen web server for the identification of potent antigenic protein. As a corollary, UniProtKB id: Q9YMG2 was identified as the most potent antigenic protein having a maximum total prediction score of 0.5390. Here the threshold of 0.5 is considered as the potent antigenicity.55 This sequence was used for further analysis.

**Prediction of potent T-cell epitope**

On the basis of the high combinatorial score, the five best epitopes were predicted by the NetCTL server from the selected protein sequence in a preselected environment. The identified epitopes are represented in Table 1. In combination with several methods such as proteasomal cleavage/transporter associated with antigen processing (TAP)/MHC-I combined predictor, MHC-I processing of the NetCTL server calculates an overall score for each peptide's intrinsic potential from a protein for the designing of T-cell epitope. Peptides with a higher score represent higher processing capabilities.

The five T-cell epitopes were subjected to MHC-I binding prediction, using the stabilized matrix base method. The epitopes that elicited higher affinity (IC$_{50}$ <200 nM) were subjected to afterward analysis (Table 2). Notably, proteins are transformed into peptides by proteasome complex, which cleaved the peptide bonds. By combining with Class I MHC molecules, these peptides were deported to the cell membrane, where they were introduced to T helper cells.

As shown in Table 2, a 9-mer T-cell epitope HKEGAFFLY was detected to interact with most MHC-I alleles, including HLA-B*15:02; HLA-C*03:03; HLA-A*32:15; HLA-B*15:03; HLA-B*27:20; HLA-C*12:03; HLA-A*68:23; HLA-B*40:13; and HLA-A*32:07 with higher affinity, among the five T-cell epitopes.

Furthermore, this epitope retained the highest conservancy of 83.85%, according to the IEDB conservancy analysis, as tabulated in Table 2. As population coverage in vaccine design generally plays a crucial role, it was calculated in this study.

**Population coverage and allergenicity analysis**

The cumulative percentage of population coverage was obtained for the predicted epitope HKEGAFFLY. As shown in Table 3, the population coverage for East Africa was found to be 66.98%; in West and North Africa, it was 69.50% and 63.89%, respectively; and for Central Africa it was observed to be 75.93%. The population coverage was recorded at 55.88% for the East Asian region, which was a major hotspot for viral infection. For North America, the population coverage was found to be 58.69%. In current vaccine design pipeline, allergenicity is considered the most prominent barrier in vaccine designing, since most vaccines convert the immune system into an “allergic” reaction by inducting Type 2 T helper cells and immunoglobulin E. That is why we predicted allergenicity of the selected epitope by the AllerHunter web server, where the probability is >0.06. The epitope HKEGAFFLY was scored 0.00 (sensitivity =91.6%, specificity =89.3%), and was thus considered a nonallergen, according to the Food and Agriculture Organization/World Health Organization evaluation system of allergenicity prediction.
In silico-based vaccine design against EBOV GP

Q9YM42 Super small secreted glycoprotein Zaire EBOV (strain Mayinga-76)
L7QWH1 Second secreted glycoprotein Zaire EBOV
Q0532D Envelope glycoprotein Zaire EBOV (strain Mayinga-76)
P60170 Pre-small/secreted glycoprotein Zaire EBOV (strain Mayinga-76)
A9QPM0 sGP Zaire EBOV
P0C773 Super small secreted glycoprotein Zaire EBOV (strain Kikwit-95)
P87666 Envelope glycoprotein Zaire EBOV (strain Kikwit-95)
A9QPL9 GP12 Zaire EBOV
Q6V1QS Second nonstructural secreted glycoprotein Zaire EBOV (strain Kikwit-95)
P87670 Pre-small/secreted glycoprotein Zaire EBOV (strain Eckron-76)
P87671 Envelope glycoprotein Zaire EBOV (strain Eckron-76)
O11457 Envelope glycoprotein Zaire EBOV (strain Gabon-94)
O11458 Pre-small/secreted glycoprotein Zaire EBOV (strain Gabon-94)
A9A066JJ419 GP12 Zaire EBOV
A9A01EDTO Virion spike glycoprotein Zaire EBOV
X5HM49 Virion spike glycoprotein Zaire EBOV
G8DB50 sGP Zaire EBOV
G8DB49 Surface glycoprotein Zaire EBOV
G8DB42 Second nonstructural secreted glycoprotein Zaire EBOV
X5HMX4 Virion spike glycoprotein Zaire EBOV
Q68811 Pre-small/secreted glycoprotein Tai Forest EBOV (strain Cote dIvoire-94)
B8XCPO Soluble secreted glycoprotein Tai Forest EBOV
B8XCN9 Spike glycoprotein Tai Forest EBOV
B8XCP1 Putative second secreted glycoprotein Tai Forest
POC772 Super small secreted glycoprotein Sudan EBOV (strain Uganda-00)
Q7T9E0 Pre-small/secreted glycoprotein Sudan EBOV (strain Uganda-00)
Q7T9D9 Envelope glycoprotein Sudan EBOV (strain Uganda-00)
I7FLM2 Secreted glycoprotein Sudan EBOV - Nakisamata
I7F2J9 Structural glycoprotein Sudan EBOV - Nakisamata
P60173 Pre-small/secreted glycoprotein Sudan EBOV (strain Maleo-79)
P66814 Envelope glycoprotein Sudan EBOV (strain Boniface-76)
BOLPL7 Structural glycoprotein Sudan EBOV
BOLPL8 Structural soluble secreted glycoprotein Sudan EBOV
C6G8E4 Structural glycoprotein OSReston EBOV (strain Philippines-96)
C6G8E5 Small/secreted nonstructural glycoprotein Reston EBOV (strain Philippines-96)
C6G8D6 Structural glycoprotein Reston EBOV (strain Philippines-96)
C6G8F3 Small/secreted nonstructural glycoprotein Reston EBOV (strain Philippines-96)
C6G8D7 Small/secreted nonstructural glycoprotein Reston EBOV (strain Philippines-96)
C6G8F2 Structural glycoprotein Reston EBOV (strain Philippines-96)
Q91DD8 Envelope glycoprotein Reston EBOV (strain Philippines-96)
POC770 Super small secreted glycoprotein Reston EBOV (strain Philippines-96)
P60172 Pre-small/secreted glycoprotein Sudan EBOV (strain Boniface-76)
Q89569 Pre-small/secreted glycoprotein Reston EBOV (strain Siena/Philippine-92)
Q66800 Pre-small/secreted glycoprotein Reston EBOV (strain Reston-89)
POC771 Super small secreted glycoprotein Reston EBOV (strain Reston-89)
P66799 Envelope glycoprotein Reston EBOV (strain Reston-89)

Figure 2 Evolutionary divergence analysis of available glycoproteins of different strains of EBOV; results are represented in a phylogenetic tree.

Abbreviations: EBOV, Ebola virus; GP, glycoprotein; sGP, soluble GP.
Table 1 Five most potent T-cell epitopes, according to the overall score predicted by the NetCTL server

<table>
<thead>
<tr>
<th>Number</th>
<th>Epitopes</th>
<th>Overall score (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ATEDPSSGY</td>
<td>2.8069</td>
</tr>
<tr>
<td>2</td>
<td>TEDPSSGYY</td>
<td>1.5503</td>
</tr>
<tr>
<td>3</td>
<td>LFEVDNLTY</td>
<td>1.1224</td>
</tr>
<tr>
<td>4</td>
<td>HKEGAFFLY</td>
<td>1.0887</td>
</tr>
<tr>
<td>5</td>
<td>LLQLNETIY</td>
<td>0.9753</td>
</tr>
</tbody>
</table>

Table 2 Interaction, binding, and conservancy of identified T-cell epitopes

<table>
<thead>
<tr>
<th>Epitope</th>
<th>Interacting MHC-I allele</th>
<th>Epitope conservancy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATEDPSSGY</td>
<td>HLA-B<em>15:02 (0.59), HLA-A</em>26:02 (0.59)</td>
<td>66.23</td>
</tr>
<tr>
<td>TEDPSSGY</td>
<td>HLA-B<em>40:13 (0.52), HLA-B</em>15:02 (0.64), HLA-A*32:15 (0.69)</td>
<td>66.23</td>
</tr>
<tr>
<td>LFEVDNLTY</td>
<td>HLA-B<em>40:13 (0.6), HLA-B</em>32:15 (0.63), HLA-C*03:01 (0.67)</td>
<td>66.28</td>
</tr>
<tr>
<td>HKEGAFFLY</td>
<td>HLA-B<em>15:02 (0.23), HLA-C</em>03:03 (0.31), HLA-A<em>32:15 (0.72), HLA-B</em>15:03 (0.73), HLA-B<em>27:20 (0.91), HLA-C</em>12:03 (1.08)</td>
<td>83.85</td>
</tr>
<tr>
<td>LLQLNETIY</td>
<td>HLA-B<em>15:02 (0.45), HLA-B</em>15:03 (0.7), HLA-C<em>03:03 (0.71), HLA-C</em>14:02 (0.71), HLA-A<em>32:15 (0.74), HLA-A</em>68:23 (1.11), HLA-C<em>12:03 (1.17), HLA-A</em>32:07 (1.26)</td>
<td>60.38</td>
</tr>
</tbody>
</table>

Table 3 Analysis of the population coverage for the proposed epitope against EBOV

<table>
<thead>
<tr>
<th>Population</th>
<th>Coverage (%)a</th>
<th>Average hitb</th>
<th>PC90c</th>
</tr>
</thead>
<tbody>
<tr>
<td>East Africa</td>
<td>66.98</td>
<td>1.00</td>
<td>0.23</td>
</tr>
<tr>
<td>West Africa</td>
<td>69.50</td>
<td>0.89</td>
<td>0.25</td>
</tr>
<tr>
<td>North Africa</td>
<td>63.89</td>
<td>1.13</td>
<td>0.26</td>
</tr>
<tr>
<td>Central Africa</td>
<td>75.93</td>
<td>1.04</td>
<td>0.23</td>
</tr>
<tr>
<td>North America</td>
<td>58.69</td>
<td>0.98</td>
<td>0.21</td>
</tr>
<tr>
<td>East Asia</td>
<td>55.88</td>
<td>2.11</td>
<td>0.29</td>
</tr>
<tr>
<td>Europe</td>
<td>67.57</td>
<td>1.00</td>
<td>0.23</td>
</tr>
<tr>
<td>Southeast Asia</td>
<td>63.00</td>
<td>1.45</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Notes: aProjected population coverage. bAverage number of epitope hits/HLA combinations recognized by the population. cMinimum number of epitope hits/HLA combinations recognized by 90% of the population.

Abbreviations: EBOV, Ebola virus; HLA, human leukocyte antigen.

Validation of predicted T-cell epitope

As described in the “Materials and methods” section, the hypothetical structure of HLA-A*32:15 protein was generated using the homology technique. The structure was then analyzed through various web-based protein validation software. As shown in Figure 3, the Ramachandran plot generated by the PROCHECK server showed that about 98.9% of the residues of protein are located in the most favored region, as against 0% in the outlier region and 1.1% in the generously allowed region. It should be noted that the protein model having >90% of the residues in the core and allowed regions can be considered a high-quality model. The hypothetical model was further analyzed using ERRAT and Verify 3D. For a good model, structure should retain an ERRAT score >80.00, against which the model in this study obtained an ERRAT score of 89.859. Verify 3D graph indicates that 100.00% of residues of this model had an averaged 3D-1D score of 0.2, which is good. Along with the QMEAN analysis, the protein model in our interest resulted in a Z-score of −1.33, and the total score was 0.636. This value denotes a higher quality of the model, where the acceptable score ranges between 0 and 1 (Figure 3B). On the basis of the results obtained from the aforementioned structural validation programs, the model (Figure 3C) showed much reliability and was considered for further study.

Molecular docking simulation revealed that the proposed epitopes bound in the cleft of the HLA-A*32:15 (Figure S2), where the highest binding affinity was −7.6 kcal/mol (Table S2, observed for the HKEGAFFLY epitope). The Chimera program was used to visualize the interactions of docked HLA-A–epitope complexes, as shown in Figures 4 and S2. Then, binding energy calculation was carried out to understand the binding of HLA with epitopes. Here the binding free energies of MM-GBSA and MM-PBSA are approximate free energies of binding, so a more negative value denotes stronger binding. From MM-GBSA analysis, the highest binding free energy was observed for HLA-A*32:15 with epitope (HKEGAFFLY) of −63.89 kJ/mol (Table S1). On the contrary, the lowest binding free energy was obtained for ATEDPSSGY epitope, i.e. −44.86 kJ/mol. In contrast of MM-GBSA, the HKEGAFFLY epitope was also resulted the highest binding energy of −38.48 kJ/mol, while the lowest binding free energy was seen for TEDPSSGY epitope, −20.98 kJ/mol. Since HKEGAFFLY epitope obtained the highest docking affinity and binding free energy, its complex subjected for molecular dynamics simulation.
The 100 ns MD simulation of HLA-epitope (HLA-A*32:15-HKEGAFFLY) complex was carried out using AMBER14 force field, following the energy minimization protocol. The stability of the HLA-epitope complex by means of RMSD was calculated and rendered in Figure 5A. From the results, it is revealed that the HLA molecule was stabilized after 5 ns simulation and tended to remain in plateau phase thereafter for rest of the period. The RMSD value of HLA was observed to grow up quickly from 0.48 Å to 2.214 Å and remained stable in the range from 2.0 Å to 3 Å. In case of epitope, similar RMSD pattern was observed, where the order of magnitude was seen to fluctuate in some range. The average energy of the simulation was \(-578125.270\) kJ/mol; the average Coulombic charge and van der Waals interactions was \(-694749.662\) kJ/mol, 77122.511 kJ/mol, respectively. We also calculated the contribution of each residue for both HLA and epitope in the simulation, in terms of RMSF and RMSD. As seen in Figure 6A, highest RMSD was observed
for ARG residue at the position of 180 in HLA, while lowest RMSD observed for CYS100. However, this residue was also resulted highest RMSF value of 7.181 Å, while the rests of the residues were in lowest fluctuation. In case of epitope, the histidine residue at the first position and the tyrosine residue in 9th position were seen to be very much flexible, as these residues were resulted with highest RMSD and RMSF (Figure 6B). In the meanwhile, we calculated the number of hydrogen bond formed between the epitope and HLA molecule during the simulation. The results represented in Figure 5B, showed that hydrogen bond at initial stage was 236, and the range decreased to 160. During the simulation, the number of hydrogen bond was at a range of 160–210,
which indicates the strong binding of epitope-HLA complex. Hence, all analyses lead to the conclusion that HKEGAFFLY is one of the most prominent T-cell epitopes for GP based designing of vaccine.

B-cell epitope identification

For the identification of potential B-cell epitopes, amino acid scale base methods have been used in this study. Consistent with this protocol, we used diverse investigation processes for the calculation of an incessant B-cell epitope.

According to the analysis of Kolaskar and Tongaonkar’s antigenicity prediction method, the average antigenicity was 1.028, while 1.225 and 0.894 were the maximum and minimum, respectively. The Kolaskar and Tongaonkar antigenicity prediction uses a semiempirical method to predict antigenicity on the basis of physicochemical properties of the residues in a protein and their diversity in experimentally known epitopes, where values >1.00 were considered to denote a potential antigen. As summarized in Table 4 and Figure S3, 14 epitopes have been found to satisfy the threshold value, and also have the potentiality to express the B-cell response. Furthermore, the surface accessibility of the protein was also analyzed using the Emini surface accessibility prediction methods, since a potent B-cell epitope should be accessible through the surface. As shown in Figure S4 and Table 5, higher accessibility was found in regions 9–17 and 186–223 amino acid residues. Figure S5 represents the β-turns region identified by Chou and Fasman β-turn methods. According to the result, the region from 200 to 220 (in the region of 200–220 and 105–150) is regarded as β-turns as well as hydrophilic in nature. These are two properties required to be a potent B-cell epitope. Experimentally, antigenicity is related to the protein flexibility. That is why we implemented the Karplus and Schulz flexibility prediction method, where it was evident that the regions of 255–280 and 200–220 were regarded as the most flexible (Figure S6). Finally, based on the Hidden Markov model, the Bepipred linear epitope prediction tool was utilized to predict linear B-cell epitopes. The predicted result is rendered and tabulated in Figure S7 and Table 6. Hence, by comparing the foregoing results, the peptide sequences ranging from 186 to 220 are
of GP of EBOV therefore makes an appropriate antigenic target for vaccine development; as a result, several reports have been published on this perspective. However, the information representing the population coverage in the worldwide are still limited. In such case, computational based epitope screening is very much efficient in context of HLA class I molecules, and also much safe, high specificity and cost effective. Therefore, this study incorporated various immunoinformatics and molecular modelling tools to identify potential epitopes present in EBOV GPs.

Initially, a set of 46 glycoprotein sequences from the different strains of EBOV has been subjected to perform multiple sequence alignment. Previous GP sequences analysis of different strains of each EBOV species revealed a high degree of sequence similarity, and thereby, it is believed that targeting GP from old strain could provide strong and cross reactive immunity against the new strain and previous outbreaks in 2014. Interestingly, in our molecular analysis, we have found ~98–99% conservation for the amino acid sequences of different strains within the species, which confers the degree of able to provoke the immune response as B-cell epitope for GP-based designing of vaccine.

**Discussion**

In recent trends, the primary focus of vaccine development is very much rely on GPs, as they are involved in cell attachment, fusion and entry as well as assist in invasion; and thus plays the role of pathogenesis of disease. The central role of GP of EBOV therefore makes an appropriate antigenic target for vaccine development; as a result, several reports have been published on this perspective. However, the information representing the population coverage in the worldwide are still limited. In such case, computational based epitope screening is very much efficient in context of HLA class I molecules, and also much safe, high specificity and cost effective. Therefore, this study incorporated various immunoinformatics and molecular modelling tools to identify potential epitopes present in EBOV GPs.

Initially, a set of 46 glycoprotein sequences from the different strains of EBOV has been subjected to perform multiple sequence alignment. Previous GP sequences analysis of different strains of each EBOV species revealed a high degree of sequence similarity, and thereby, it is believed that targeting GP from old strain could provide strong and cross reactive immunity against the new strain and previous outbreaks in 2014. Interestingly, in our molecular analysis, we have found ~98–99% conservation for the amino acid sequences of different strains within the species, which confers the degree of

**Table 4** Prediction of antigenic region of the protein by Kolaskar and Tongaonkar antigenicity prediction method

<table>
<thead>
<tr>
<th>Number</th>
<th>Start</th>
<th>End</th>
<th>Peptide</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>11</td>
<td>TGILQLPR</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>56</td>
<td>TSFFLVWVIIFQRTFSIPLGVIHNSLPQVSDDKLKLVR</td>
<td>40</td>
</tr>
<tr>
<td>3</td>
<td>63</td>
<td>69</td>
<td>LRSGVLN</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>76</td>
<td>82</td>
<td>ATDVPASA</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>89</td>
<td>99</td>
<td>RSGVPKVVNY</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>118</td>
<td>126</td>
<td>GSECLPAAP</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>132</td>
<td>154</td>
<td>FPRCRYVKVSGTPCAGDFAFH</td>
<td>23</td>
</tr>
<tr>
<td>8</td>
<td>156</td>
<td>172</td>
<td>EGAFILYKLRALTIVYR</td>
<td>17</td>
</tr>
<tr>
<td>9</td>
<td>177</td>
<td>189</td>
<td>AEGVAFILIPOA</td>
<td>13</td>
</tr>
<tr>
<td>10</td>
<td>194</td>
<td>202</td>
<td>FSSHPREP</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>211</td>
<td>221</td>
<td>SGYSSATTIRYQ</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>233</td>
<td>247</td>
<td>LFEVNLTYVQLESR</td>
<td>15</td>
</tr>
<tr>
<td>13</td>
<td>249</td>
<td>259</td>
<td>TPQFLLQLNET</td>
<td>11</td>
</tr>
<tr>
<td>14</td>
<td>274</td>
<td>280</td>
<td>IWKVNPE</td>
<td>7</td>
</tr>
</tbody>
</table>

**Table 5** Prediction of the Emini surface accessibility of the protein under study

<table>
<thead>
<tr>
<th>Number</th>
<th>Start</th>
<th>End</th>
<th>Peptide</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>17</td>
<td>LPDRFKRT</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>56</td>
<td>63</td>
<td>KLSTNQL</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>81</td>
<td>87</td>
<td>SATKRWG</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>111</td>
<td>117</td>
<td>LEIKKPD</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>188</td>
<td>193</td>
<td>QAKKDF</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>197</td>
<td>221</td>
<td>HPQPLPVNATEDPSSGYSSATTIRYQ</td>
<td>25</td>
</tr>
<tr>
<td>7</td>
<td>227</td>
<td>232</td>
<td>TNMFS</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>244</td>
<td>249</td>
<td>LERSFT</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>263</td>
<td>270</td>
<td>SGKRSNTT</td>
<td>8</td>
</tr>
</tbody>
</table>

**Table 6** Identification of the Bepipred linear epitope from the protein sequence

<table>
<thead>
<tr>
<th>Number</th>
<th>Start</th>
<th>End</th>
<th>Peptide</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14</td>
<td>14</td>
<td>F</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>57</td>
<td>59</td>
<td>LSS</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>73</td>
<td>106</td>
<td>NGVATDVPATKRWFRSGVPPKVYNTEAGWAE</td>
<td>34</td>
</tr>
<tr>
<td>4</td>
<td>114</td>
<td>131</td>
<td>KKPDGSECLPAAPDGIRG</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>141</td>
<td>148</td>
<td>VSGTGPAC</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>175</td>
<td>176</td>
<td>TF</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>191</td>
<td>193</td>
<td>KDF</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>198</td>
<td>215</td>
<td>PLREPYNATEDPSSYYS</td>
<td>18</td>
</tr>
<tr>
<td>9</td>
<td>223</td>
<td>229</td>
<td>TGFGTNE</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>261</td>
<td>270</td>
<td>YTSGKRSNTT</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>279</td>
<td>285</td>
<td>PEIDTTI</td>
<td>7</td>
</tr>
</tbody>
</table>
similarity and support the previous analysis. From this set of GPs, the most antigenic protein sequence was determined by Vaxijen server. Based on auto cross covariance (ACC), the Vaxijen server transform the protein sequence into uniform vectors of physicochemical properties of proteins. With 91% sensitive, 82% accuracy and 72 specificity, the leave one – out cross validation was used to identify antigenicity of protein for viral species. The resultant antigenic protein (VaxiJen score ≥0.5) was then subjected for various immunoinformatics analysis, followed by IEDB web server.

At the beginning five potent 9-mer epitopes have been predicted from NetCTL 1.2 server and selected for further study. Using the threshold of 0.5, the NetCTL 1.2 server predicts maximum number of epitopes without compromising the specificity or sensitivity levels, covering all 12 MHC class I supertypes. The five most potent epitopes are represented in Table 1, and the scores are the predicted MHC class I affinities in the form of –logIC50 and IC50 value.

For MHC-I binding prediction, peptides with IC50 values <50 nM are considered high affinity, <500 nM for intermediate affinity, and <5000 nM for low affinity. Therefore, we selected maximum alleles having binding affinity <200 nM. It is advocated that T-cell epitope binding to specific multiple HLA supertypes are termed as promiscuous in vaccine design, since they effectively increase the coverage of higher proportions of human populations. According to the results, both HKEGAFFLY and LFEVDNLTY bind to the highest number of alleles. However, HKEGAFFLY represents highest conservancy and was hence considered as epitope of choice.

We also validated each epitope by molecular docking simulation and MM-GBSA/MM-PBSA studies with HLA-A*32:15 protein, as it was found common in the results from MHC-I binding interaction analysis. Prior of docking simulation, the three dimensional structure of HLA molecule was prepared by using the Phyre2, followed by intensive mood. As a result, 179 residues (99%) of HLA-A*32:15 modelled at >90% accuracy. In these study, 2BCK_Chain A, crystal structure of HLA-A*2402 showed the highest similarity of 90% (Figure S1). The selected model has been chosen from the twenty models generated by Phyre2, on the basis of similarity and confidence level. Phyre2 is one of the best protein prediction servers that allows remote fold reorganization and homology detection. Using hidden Markov model (HMM), this server predicts the structure of given protein sequence by constructing backbone, loop modelling and adding side chains. However in intensive mode, additional ab initio approach is used for reconstruction of missing region, backbone and side chain. In docking simulation, among the other epitopes, HKEGAFFLY obtained highest binding affinity (Table S1, supplementary material). In addition, MM-PBSA and MM-GBSA techniques are frequently to re-rank docking poses from molecular docking study, as they achieve a much better performance than docking scoring functions. Nevertheless, the success rate of the absolute binding free energy prediction strongly depends on the systems. Thence, we used both of these solvation models to predict binding energy more accurately, where the results from MM-PBSA examine the accuracy and reliability of the results from MM-GBSA. Results of binding energy calculation are shown in Table S1. The relative magnitude of binding free energy obtained from GB methods is found to be consistent with those calculated using PB method, despite of the differences in the absolute value of salvation energy. As a corollary, these results also demonstrated the consistence with relative stabilities of HLA-epitope complexes. In previous published reports regarding in silico epitope identification of EBOV, the studies are limited to sequence-based scoring function techniques and some extend to docking simulation. These techniques have certain limitations, though these are very useful. In docking simulation of peptide and protein, it faces problem like peptide’s flexibility. Whereas, energy based approach like molecular mechanics and interaction energy scoring can add valuable information to sequence based results.

Therefore, we have performed MD simulation study of 100 ns long to enhance the predictive power of the peptide affinity calculations to MHC molecule. In molecular dynamics simulation, both epitope and HLA protein were seen to achieve equilibration, while different fluctuations of RMSD were seen by the time evolution. Higher RMSD values of epitope indicate the flexibility in binding with HLA molecule, during the simulation. These results were further confirmed by the analysis of per residue contributions in dynamics simulation by means of RMSF and RMSD. Low values of RMSF indicate the core region of the HLA protein was stable, while high values of RMSD demonstrated the motion of the protein during the simulation. In like manner, the RMSD and RMSF profiles of epitope confirm the synergic conformation changes to accommodate the binding pocket of HLA. The hydrogen bond occupancy analysis between the HLA-epitope further confirmed the stability of the complex during the simulation. Overall, these results evidently demonstrate that both HLA and epitope have remarkable conformation changes to facilitate the binding and formed stable complex in thermodynamic environment (Figure 7).
In recent years, most vaccines have been developed based on B-cell immunity; however, the current strategy relies mostly on T-cell epitope owing to long-lasting immunity. Both B-cell and T-cell epitopes are offered in this study for stimulating immunity in several ways. The resulting peptides showed B-cell and T-cell selectivity, better conservancy, population coverage, and significant interaction with MHC-1 allele with good affinity. Above all, the predicted epitopes are anticipated to offer long-term and high protective immunity against EBOV.

**Disclosure**

The authors report no conflicts of interest in this work.

**References**

In silico-based vaccine design against EBOV GP


**Supplementary materials**

**Figure S1** Multiple sequence alignment and secondary structure predictions of query sequence (HLA*A15:02) and template sequence (PDB: 2BCK).

*Notes:* Green color represents alpha helix, whereas beta sheet is represented by blue color.

*Abbreviations:* HLA, human leukocyte antigen; PDB, Protein Data Bank.

**Figure S2** Visualization of the best docked HLA–epitope complexes obtained from molecular docking analysis. The complexes are representing the interaction of HLA (HLA*A15:02) with (A) ATEDPSSGY, (B) TEDPSSGYY, (C) LFEVDNLTY, and (D) LLIQNETIY epitope.

*Abbreviation:* HLA, human leukocyte antigen.
Figure S3 Kolaskar and Tongaonkar antigenicity prediction of the most antigenic protein.
Notes: Here, the x-axis and y-axis represent sequence position and antigenic propensity, respectively. The threshold value is 1.000. The regions above the threshold are antigenic, shown in yellow while, green color reflects the polypeptide regions that could not satisfy the threshold margin.

Figure S4 Emini surface accessibility prediction of the most antigenic protein.
Notes: Here, the x-axis and y-axis represent sequence position and antigenic propensity, respectively. The threshold value is 1.000. The regions above the threshold are antigenic, shown in yellow while, green color reflects the polypeptide regions that could not satisfy the threshold margin.
Figure S5 Chou and Fasman β-turn prediction of the most antigenic protein.

Notes: Here, the x-axis and y-axis represent sequence position and antigenic propensity, respectively. The threshold value is 1.000. The regions above the threshold are antigenic, shown in yellow while, green color reflects the polypeptide regions that could not satisfy the threshold margin.

Figure S6 Karplus and Schulz flexibility prediction.

Notes: Here, x-axis and y-axis represent position and score, respectively. The threshold is 1.00. The flexible regions are shown in yellow while, green color reflects the polypeptide regions that could not satisfy the threshold margin.
Figure S7 Bepipred linear epitope prediction of the most antigenic protein.

Notes: Here, x-axis and y-axis represent position and score, respectively. The threshold is 0.35. The flexible regions are shown in yellow while, green color reflects the polypeptide regions that could not satisfy the threshold margin.

Table S1 Results of binding affinity and binding energy calculations of the selected epitopes

<table>
<thead>
<tr>
<th>Epitope</th>
<th>Binding affinity (kcal/mol)</th>
<th>(ΔG) MM–PBSA</th>
<th>(ΔG) MM–GBSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATEDPSSGY</td>
<td>−6.8</td>
<td>−21.89</td>
<td>−44.86</td>
</tr>
<tr>
<td>TEDPSSGYY</td>
<td>−6.5</td>
<td>−20.98</td>
<td>−57.63</td>
</tr>
<tr>
<td>LFEVDNLTY</td>
<td>−6.3</td>
<td>−33.58</td>
<td>−55.96</td>
</tr>
<tr>
<td>HKEGAFFLY</td>
<td>−7.6</td>
<td>−38.48</td>
<td>−63.89</td>
</tr>
<tr>
<td>LLQLNETIY</td>
<td>−7.0</td>
<td>−37.25</td>
<td>−48.63</td>
</tr>
</tbody>
</table>