Prediction of multi-epitopic domains of a putative oral vaccine against hepatitis C virus

*Olatunde Micheal Adeoti*\(^{1a,b,c}\), Aanuoluwa Victoria Akinsooto\(^2a\), Elisabeth Olajumoke Adedokun \(^3a\) and Opeyemi Joy Olaoye\(^4a\)

\(^a\)The Oke- Ogun Polytechnic, Saki Oyo State Nigeria, Department of Science Laboratory Technology, Microbiology Option

\(^b\)Ladoke Akintola University of Technology, Department of Pure & Applied Biology, Ogbomoso

\(^c\)Cellular Parasitology Unit, Department of Zoology, University of Ibadan, Nigeria

**ARTICLE INFO**

**ABSTRACT**

For vaccine development, triggering an immune response is desired. Designing and assessing vaccine candidates for the appropriate immune response is critical for their success. Hepatitis C virus is the major cause of liver disease. Anti HCV vaccines if designed is rational decision to reinforce specific T-cell as a crucial aspect of effective antiviral treatment. This study explored the use of bioinformatics tools by retrieval of twenty (20) HCV proteins which were selected for vaccine design. These were retrieved from UniProt server based on their antigenicity, virulence, subcellular localization, essentiality non-homology and other physical parameters, including, TM helices, and relative molecular mass. BLASTp revealed 80% non-identity with Homo sapiens genes. The Epitopes obtained include: Q3S781_9HEPC\(^{52-71}\), POLG_HCVBK\(^{442-461}\), POLG_HCVJA\(^{2-21}\), POLG_HCVJ1\(^{77-95}\), POLG_HCVR6\(^{1107-1126}\), POLG_HCVJP\(^{47-66}\), POLG_HCVTW\(^{64-68}\), POLG_HCVTR\(^{46-465}\), LTOR5_HUMAN\(^{23-42}\), POLG_HCVT5\(^{100-119}\), POLG_HCVJT\(^{77-96}\), HOIL1_HUMAN\(^{169-188}\), POLG_HCJV4\(^{644-663}\), POLG_HCJV8\(^{47-66}\), TFB2M_HUMAN\(^{49-68}\), RSF1_HUMAN\(^{138-157}\), A8DGK3_9HEPC\(^{77-96}\), A8DFL0_9HEPC\(^{2-21}\), and A8DFL0_9HEPC\(^{2-21}\). An antigenicity score of 0.6004 was obtained with the use of VaxiJen server. The allergenicity prediction showed that the vaccine is not allergenic with the use of AllerTOP v.2.0 and AlgPred servers. The molecular weights and theoretical pI of protein were 45.1 kDa and 10.24 kDa respectively. A potentially suitable vaccine candidate with multivariant regions and immunogenic which could be antagonistic to HCV was designed.

2021 Sciforce Publications. All rights reserved.

**Introduction**

Hepatitis C is a disease caused by hepatitis C virus (HCV); the virus can cause both acute and chronic hepatitis, ranging in severity from mild illness lasting a couple of weeks to a significant live long illness. Hepatitis C is that the major explanation for cancer of the liver. Hepatitis C virus (HCV) may be a RNA virus known to infect human and chimpanzees, causing similar disease in these 2 species. HCV is most frequently transmitted parenterally but is additionally transmitted vertically and sexually (MMWR Recomm Rep. 1998). It also requires less exposure that HIV to cause infection (Te HS, Jensen DM. 2010). HCV may be a leading explanation for chronic disease within the world [Williams et al., 2006]. The World Health Organization (WHO) estimates that 170 million people are infected with HCV globally and 3 - 4 million new infections occur annually (Madhava *et al.* 2002). Despite decades of research, there's still no effective vaccine available for HCV thanks to high genetic heterogeneity for the HCV RNA (Manns *et al.* 2017). Despite its high prevalence and highly infectious nature, HCV remains under-diagnosed and underreported in Africa (with the exception of Egypt). Most of the available data on HCV in Africa are old and outdated. We therefore began to review available medical literature on HCV in Africa with a view to determining the prevalence, disease burden
and customary transmission modes. Additionally, we draw attention to diagnosis, treatment and prevention of HCV.

Recently many studies are done introducing effective vaccines but so far no approved vaccine for HCV infection has been introduced. Generally, there are several approaches to style vaccines against microbial infection including living, inactivated, subunit, toxoids, conjugate, DNA, recombinant vector and peptide vaccines. Fundamental information about the microbe like life cycle, virulence factor(s) and host cell receptor(s) also as practical considerations, like the world where vaccine is employed and prevalence play key roles in select the sort of vaccine (Kallerup, Rie S., and Camilla Foged, 2015) The technological advances within the fields of genomics, proteomics, human immunology, and structural biology have provided the molecular information for the invention and prediction by bioinformatics tools of novel antigens, epitopes, and style of vaccines against pathogenic bacteria, like meningococcus B (Serruto et al.,2012; O’Ryan et al.,2013; Rappuoli et al.,2016).

Conventional vaccines are often characterized by an isolate, inactivate/attenuated, inject archetype of development (Ranjibar et al.,2015). The main target on vaccine design and development has changed to the assembly of peptides composed of multiple epitopes (multi epitope vaccines), supported linear arrangements, as a completely unique alternative. Additionally, epitope-based vaccines have demonstrated various advantages, including safety, the chance to rationally engineer the epitopes for increased potency, breadth, and antigenicity, and therefore the possibility to focus large repertoires of immune responses on conserved epitope sequences (Livingston et al., 2002; Oyarzu’ and Kobe, 2015). In the present study, an in silico attempt towards prediction of hypervariable vaccine against Hepatitis C. Epitopes are enriched with charged with water-loving domains, its loops are depleted in helix to \( \beta \)-helix. Predicting epitopes is vital to understanding of the basis of immunological differentiation between self and non-self as well as mechanisms of bio-recognition. Much of these multi-epitopic regions are desired for its multi functionality and hyper variability in developing good vaccine candidates.

Abbreviations

CELLO: CELLular Localization; HCV: Hepatitis C- Virus; BLAST: Basic Local Alignment Search Tools

Material and methods

Selection and Retrieving sequence and antigenic evaluation of Protein

Hepatitis C proteins were selected based on reported allergenicity, antigenicity, Virulence and their relatedness to the mechanisms of adhesion. The entire twenty protein sequences were retrieved from UNIPROT reference sequence database in FASTA format. Virulence factors were predicted, gene essentiality as related to essential genes of HCV were downloaded according to Ning et al., 2014, Sayers et al., 2018). For the virus proteins; subcellular localization prediction, CELLO v2.5 and CELLO2GO web server were used (Yu et al., 2006, 2014). The database of GepTop was used to evaluate gene essentiality (Chen et al., 2017).

PATRIC3.5.16 databases was used for study of the virulence role of proteins (Watam et al., 2017, Garg and Gupta, 2008) and VirulentPred database for SVM based prediction method for virulent proteins in bacterial pathogens was adapted for HCV. Proteins were screened by Vaxign and BLASTp alignment were used to detect sequence homology to Homo sapiens. Blasting against the protein sequences was to detect and eliminate proteins that can self-react to human leukocytes or otherwise. Auto-immunity of the selected sequences should be guarded against; a good vaccine candidate must not have similarity/homology with human genes. Also, transmembrane (TM) helices prediction was done by using TMHMM v2.0 server (Krogh et al., 2001). Compute pi/Mw tool was used to calculate the estimated isoelectric point and molecular weight of all amino acid sequences (Wilkins et al., 1999).

Phylogenetic Evolution

Construction of Phylogenetic tree and sequence similarity/dissimilarity matrices of the retrieved sequence of the capsid protein of Hepatitis C was created using Mega X software. 20 of the protein tree was constructed using maximum likelihood parameter in the software (fig 2). Bootstrap values was detected based on 100 replications.

T cell and B Cell Epitope Prediction

There are several antibodies which are of animal origin especially mouse, monkey or synthetic repertoires that are vital in the diagnosis and treatment of diseases. Designing and engineering of such is of vital importance to ward off infections by non- or infectious agents. This will go a long way to overcome laboratory and facility limitations as well as its economic viability. (Baran et al., 2017). To identify MHC-I binding epitopes, NetMHC 4.0 server was used (Andreatta and Nielsden, 2016). In the alternative, NetCTL 1.2 server could be used to predict CTL epitopes by integrating predictions of proteasomal cleavage, TAP transport efficiency, and MHC class I binding (Larsen et al., 2007). Fifty-one human leukocyte antigen (HLA) alleles (HLA-A, -B, -C, and -E) and six murine alleles (H-2) were evaluated. Predictions were calculated for nine-mers epitopes with a threshold for strong binders of 0.5% and a threshold for weak binders of 2%. For MHC-II binding epitopes, NetMHCII 2.3 server; predictions were obtained for 20HLA- DR alleles, 20 HLA-DQ, 9 HLA-DP, and 7 mouse H2 class II alleles using a threshold of -99.9, threshold for the strong binder of 5%, and threshold for the weak binder of 20%. Linear B cell epitopes of 20-mers were predicted utilizing ABCpred with a threshold of 0.7. The second was BCPred server which was applied with a specificity threshold of 75%. For BepiPred server, only amino acids with score >1.0 were considered for the downstream analysis (Jespersen et al., 2017).
Epitopes were selected based on the following criteria: (1) 20-mer epitopes, (2) epitopes matching on all algorithms, if possible, and (3) potential to bind with the maximum number of MHC-I and MHC-II alleles. For selection, sequences were aligned and overlapped using Clustal Omega server. Also, predicted epitopes were searched in the IEDB database (www.immuneepitope.org) to find out the already discovered experimental epitopes.

**Vaccine Design**

Vaxign Server (Xian and He, 2013) was used for the vaccine design which shows the protein accession, gene symbol, localization probability, adhesion probability, transmembrane helixes and the protein length.

**Protein Prediction and validation of secondary and tertiary structures**

The secondary structure of the multi-epitope antigen was predicted using PSIPREDv3.3 (McGuffin et al., 2000). The three-dimensional (3D) structure modeling was performed using Swiss-Model server (Yang and Zhang, 2015). Jmol was used for visualizing 3D structures of proteins. For refinement of 3D model structure, Galaxy Refine and Galaxy Loop were applied (Park et al., 2011).

The best model was validated by the ProSA web (Wiederstein and Sippl, 2007) and ERRAT (Colovos and Yeates, 1993). The residue-by-residue stereochemical qualities of models were validated by Ramachandran plot obtained from PROCHECK server (Laskowski et al., 2012). The best-refined model was selected.

**Antigenicity, Allergenicity, Solubility, and physicochemical predictions of Vaccine**

For antigenicity prediction, VaxiJen server was used. For allergenicity evaluation, AllerTOP v.2.0 and AlgPred servers were used. For solubility prediction, SOLpro server was used. Finally, ProtParam allowed the computation of various physical and chemical parameters (Wilkins et al., 1999).

**Results and Discussion**

**Protein Selection and Evaluation**

Both 9HEPC and HCVJP possessed an inner membrane location, HCVBK, HCVJA, HCVJ1, HCVCO, HCVR6, HCVTV, HCVTR, HCVT5, HCVJT, HOIL1, HCVJ4, A8DGK3, A8DHNL1 and A8DFL0 were located in periplasmic region while LTOR5, TFB2M and RSF1 were of cytoplasmic location. Only HCVJ8 was predicted as outer membrane protein as shown in Table1. By convention and traditions, surface and extracellularly located proteins are good to developing a vaccine that is aiming toward prevention of viral infections and diseases (Dwivedi et al., 2016). For genes essentiality designed by GepTop, almost half of the twenty selected proteins were qualified to possess essential genes: HCVJ1, LTOR5, HCVT5, HOIL1, HCVJ8, TFB2M, RSF1 and A8DGK3. These essential genes are needed for survivability of all organisms and are critical for its existence. These essential genes are of particular importance as result of their theoretical and practical relevance for studying the viability performance of a biological system and identifying effective therapeutic targets in pathogens (Chen et al., 2017). Homology analysis of the 20 prioritized proteins using Vaxign showed about 80% non-relatedness. The prediction of the topology of proteins by TMHMM showed that LTOR5_HUMAN, POLG_HCVJ4, TFB2M_HUMAN and RSF1_HUMAN had 0, POLG_HCVT5 had 9, POLG_HCVJA, POLG_HCVJ1, POLG_HCVCO, POLG_HCVJT, POLG_HCVJ4 POLG_HCVJ8 and A8DFL0_9HEPC had 14, Q3S781_9HEPC, POLG_HCVR6, POLG_HCVJP, POLG_HCVTV, POLG_HCVTR and A8DHN1_9HEPC had 12 TM helix, POLG_HCVBK and A8DGK3_9HEPC had 11 TMH located at 2-3 amino acid position.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Accession No</th>
<th>Subcellular Localization</th>
<th>Gene Essentiality</th>
<th>Human homology</th>
<th>TM helix</th>
<th>pI/MW(kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q3S781_9HEPC</td>
<td>tr</td>
<td>Q3S781</td>
<td>I-1.467</td>
<td>N-ES</td>
<td>NH</td>
<td>12</td>
</tr>
<tr>
<td>POLG_HCVBK</td>
<td>sp</td>
<td>P26663</td>
<td>P-1.395</td>
<td>N-ES</td>
<td>NH</td>
<td>11</td>
</tr>
<tr>
<td>POLG_HCVJA</td>
<td>sp</td>
<td>P26662</td>
<td>P-1.466</td>
<td>N-ES</td>
<td>NH</td>
<td>14</td>
</tr>
<tr>
<td>POLG_HCVJ1</td>
<td>sp</td>
<td>Q03463</td>
<td>P-1.248</td>
<td>ES</td>
<td>NH</td>
<td>14</td>
</tr>
<tr>
<td>POLG_HCVCO</td>
<td>sp</td>
<td>Q9WXM2</td>
<td>P-1.283</td>
<td>N-ES</td>
<td>NH</td>
<td>14</td>
</tr>
<tr>
<td>POLG_HCVR6</td>
<td>sp</td>
<td>Q9133V3</td>
<td>P-1.303</td>
<td>N-ES</td>
<td>NH</td>
<td>12</td>
</tr>
<tr>
<td>POLG_HCVJP</td>
<td>sp</td>
<td>Q9DHLD6</td>
<td>I-1.399</td>
<td>N-ES</td>
<td>NH</td>
<td>12</td>
</tr>
<tr>
<td>POLG_HCVTV</td>
<td>sp</td>
<td>P29846</td>
<td>P-1.476</td>
<td>N-ES</td>
<td>NH</td>
<td>12</td>
</tr>
<tr>
<td>POLG_HCVTR</td>
<td>sp</td>
<td>Q81487</td>
<td>P-1.344</td>
<td>N-ES</td>
<td>NH</td>
<td>12</td>
</tr>
<tr>
<td>LTOR5_HUMAN</td>
<td>sp</td>
<td>O43504</td>
<td>C-3.254</td>
<td>ES</td>
<td>H</td>
<td>0</td>
</tr>
</tbody>
</table>
## Table2: Potential antigenic epitopes predicted by different servers

<table>
<thead>
<tr>
<th>Order</th>
<th>Protein</th>
<th>Position</th>
<th>Sequence</th>
<th>B cell Epitope</th>
<th>T Cell Epitope</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Q3S781_9HEPC</td>
<td>52-71</td>
<td>TSERSQPRGRQRRQPIPDKRRT</td>
<td>0.998</td>
<td>0.9</td>
</tr>
<tr>
<td>2</td>
<td>POLG_HCVBK</td>
<td>442-461</td>
<td>FYANSFNSGCPERMACHRS</td>
<td>0.966</td>
<td>0.81</td>
</tr>
<tr>
<td>3</td>
<td>POLG_HCVJA</td>
<td>21-Feb</td>
<td>STNPKPQRKTKRNTNRRQD</td>
<td>0.984</td>
<td>0.83</td>
</tr>
<tr>
<td>4</td>
<td>POLG_HCVJ1</td>
<td>77-95</td>
<td>AQPGYPWPLYGNECGWAG</td>
<td>0.978</td>
<td>0.92</td>
</tr>
<tr>
<td>5</td>
<td>POLG_HCVCO</td>
<td>445-464</td>
<td>HKFNSSGCPERMASCSPIDA</td>
<td>0.951</td>
<td>0.88</td>
</tr>
<tr>
<td>6</td>
<td>POLG_HCVR6</td>
<td>1107-1126</td>
<td>DLVGWQAPPGSRSLTPCTCG</td>
<td>0.995</td>
<td>0.92</td>
</tr>
<tr>
<td>7</td>
<td>POLG_HCVJP</td>
<td>47-66</td>
<td>RATTRKTSERSQPRGRQPIP</td>
<td>0.995</td>
<td>0.9</td>
</tr>
<tr>
<td>8</td>
<td>POLG_HCVTW</td>
<td>664-683</td>
<td>CNWTRGERCDELDRDRESLS</td>
<td>0.997</td>
<td>0.86</td>
</tr>
<tr>
<td>9</td>
<td>POLG_HCVTR</td>
<td>446-465</td>
<td>HKFNSSGCPERMSSCKPITY</td>
<td>0.497</td>
<td>0.88</td>
</tr>
<tr>
<td>10</td>
<td>LTOR5_HUMAN</td>
<td>23-42</td>
<td>CTDSQGLNLGCRGTLSEHA</td>
<td>0.845</td>
<td>0.59</td>
</tr>
<tr>
<td>11</td>
<td>POLG_HCVT5</td>
<td>100-119</td>
<td>PRGSRPSWGPNDPRRSRNL</td>
<td>0.997</td>
<td>0.88</td>
</tr>
<tr>
<td>12</td>
<td>POLG_HCVJT</td>
<td>77-96</td>
<td>AQPQYPWPLYGNEGLGWAGW</td>
<td>0.967</td>
<td>0.87</td>
</tr>
<tr>
<td>13</td>
<td>HOIL1_HUMAN</td>
<td>169-188</td>
<td>EPGPPKPQGPQEPGRGQPDA</td>
<td>1</td>
<td>0.94</td>
</tr>
<tr>
<td>14</td>
<td>POLG_HCVJ4</td>
<td>644-663</td>
<td>CNWTRGERCNLEDREDRESLS</td>
<td>0.994</td>
<td>0.82</td>
</tr>
<tr>
<td>15</td>
<td>POLG_HCVJ8</td>
<td>47-66</td>
<td>RATTRKTSERSQPRGRQPIP</td>
<td>0.995</td>
<td>0.9</td>
</tr>
<tr>
<td>16</td>
<td>TFB2M_HUMAN</td>
<td>49-68</td>
<td>QLWPEPDPFRNPRPKASKAS</td>
<td>0.999</td>
<td>0.97</td>
</tr>
<tr>
<td>17</td>
<td>RSF1_HUMAN</td>
<td>138-157</td>
<td>KNIINEEDADTMRLQPIGRD</td>
<td>0.989</td>
<td>0.82</td>
</tr>
<tr>
<td>18</td>
<td>A8DGK3_9HEPC</td>
<td>77-96</td>
<td>AQPQYPWPLYGNEGLGWAGW</td>
<td>0.985</td>
<td>0.86</td>
</tr>
</tbody>
</table>

aa, amino acid; C, cytoplasmic; E, essential; ES, extracellular; N-ES, nonessential; N-H, non-homology; N-V, nonvirulent; OM, Outer membrane; P, periplasmic; TM, transmembrane; V, virulent.
The prediction of T and B-cell epitopes by different bioinformatics servers for T and B cells (NetCTL 1.2 server and using MHC-I/-II alleles for human and mouse BALB/c allowed the selection of 20 epitopes based on their score, number of alleles, and agreement between the servers. Epitopes obtained using the above mentioned servers include: Q3S781_9HEPC52-71, POLG_HCVB442-461, POLG_HCVJ210, POLG_HCVJ177, POLG_HCVCO445-464, POLG_HCVR5107-1126, POLG_HCVJ277-96, HOIL1_HUMAN169-188, POLG_HCV444-663, POLG_HCVR1107-1126, POLG_HCVJP47-66, POLG_HCVTW664-683, POLG_HCVTR446-465, LTOR5_HUMAN23-42, POLG_HCVT5100-119, POLG_HCVJT77-96, HOIL1_HUMAN169-188, POLG_HCV444-663, POLG_HCVJ464-663, HOIL1_HUMAN169-188, LTOR5_HUMAN23-42, POLG_HCVT5100-119, POLG_HCVJT77-96, RSF1_HUMAN134-157, A8DGK3_9HEPC77-96, A8DHN1_9HEPC54-73, and A8DFL0_9HEPC2-21.

Protein Structure Prediction

The vaccine is composed of 600 amino acids, and prediction of secondary structure showed that it composed of 7.75% α-helices, 2.75% β-sheets, and 89.5% others (random coil and β-turn), as shown in Figure 1.

Figure 1: Phylogenetic tree for the selected capsid protein of Hepatitis C created by Mega X

Figure 3: The z-score plot of unrefined and refined 3D structure of vaccine by ProSA-web. (a) The z-score of the starting model is -5.07, (b) The z-score of model after refinement steps is -4.77. The z-scores indicates overall model quality and is depicted as a black spot. The z-scores of all experimentally determined protein chains in current protein data bank (PDB) from NMR spectroscopy (Charcoal) and X-ray crystallography (silver). 3D, three dimensional.

Five 3D models of protein vaccine were generated among which the model with the highest c-score = 2 was selected for further refinement; the c-score range is typically from 2-3, the higher the value, the higher the confidence.
Prediction of vaccine by PSIPRED. The protein vaccine consists of 7.75% a helix (H, cylinder), 2.75% b strand (E, arrow), and 89.5% coil (C, line) secondary structural elements. The bar chart represents the percentage of confidence.

The quality and potential errors in the best model were analyzed. The initial input model z-score was -5.07, which falls within those commonly observed in similar size-native proteins (Fig.4A).

ProSA-web indicated that the preliminary model requires refinement processes. Hence, the raw model was subjected to loop refinement and energy minimization using galaxy refine. After all refinement procedures, ERRAT factor was improved from 91.56 to 98.58. The z-score of the final model reached a value of -4.77 (Fig. 4B). The starting models was given (Fig.5).

Fig. 5: Validation of vaccine 3D model using Ramachandran plots of (a) the unrefined model and (b) the refined model. The most favored (A, B, and L) and additional allowed (a, b, l and p) regions were demonstrated with charcoal and silver gray colors respectively. The generously allowed regions (-a, -b, -l and -p) are indicated in silver, and the disallowed regions are in white color. Glycine residues are shown in black triangles and other residues of protein are shown in black squares.

To validate the 3D models, Ramachandran plot analysis was performed before and after refinement processes. The Ramachandran plots of the unrefined model indicated that 92.8% of residues were located in most-favored regions, 7.2% in the additional allowed region, 0.1% in generously allowed regions, and 0.3% in disallowed regions of the plot (Fig. 6A). The refined model showed that 96.7% of residues were located in most-favored regions, 3.3% in additional allowed regions, 0.0% in generously allowed regions, and only 0.0% in disallowed regions (Fig.6B). Our results indicated that the quality and stability of the final refined model were slightly improved based on Ramachandran plot predictions.

Antigenicity, Allergenicity, Solubility and physicochemical parameters of the vaccine

An antigenicity score of 0.6004 was obtained. The allergenicity prediction showed that the vaccine is not allergenic. The molecular weight and theoretical pI of protein were 45.1 kDa and 10.24, respectively. The recombinant protein vaccine solubility upon over expression in Hepatitis C was 0.821589. Half-life was estimated to be 7.2 hours in mammalian reticulocytes, >20 hour in yeast and >10 hour in Hepatitis C. The vaccine was found as unstable within instability index of 72.40 GRAVY and aliphatic index were assessed as -1.469 and 30.73, respectively.

Conclusion

In this study, we designed presumptive multi epitopes oral vaccine against Hepatitis C Virus based on bioinformatics approach to predict structure that could be capable of provoking cellular and humoral immune response against Hepatitis C Virus could be a good vaccine candidate against Hepatitis C. However, to make the therapeutic and prophylactic effect of our oral vaccine design valid, in vitro and in vivo immunological studies are required.

Conflicts of interest

The authors hereby declare no conflict of interests

References


