

# Accepted Manuscript

Computational Identification of T-Cell Epitopes in Conserved Sequences  
Belonging to Genotype 6 of Hepatitis C Virus

Maneet Matharu, Kamini Bhardwaj

To appear in: *Journal of Neoteric Life Science*

Received Date: 03 February 2024

Accepted Date: 03 February 2024

Please cite this article as:

Matharu M., Bhardwaj K. Computational Identification of T-Cell Epitopes in Conserved Sequences Belonging to Genotype 6 of Hepatitis C Virus. *JNL*. 2024;2(1):49-67.

This PDF file contains an unedited article that has been accepted for publication. We are offering this draft of the article to our readers as a service.



# Computational Identification of T-Cell Epitopes in Conserved Sequences Belonging to Genotype 6 of Hepatitis C Virus

Author name: Matharu M.

Author affiliation: ADI Intellect, 400 Tradecenter Dr, Woburn, MA 01801, United States.

## Corresponding author details

Name: Bhardwaj K.

Address: ADI Intellect, 400 Tradecenter Dr, Woburn, MA 01801, United States.

Phone: +91-8278704376

Email: kaminibhardwaj199923@gmail.com

## Abstract

**Purpose:** Hepatitis C Virus (HCV) is the leading cause of hepatocellular carcinoma and liver infections, which can be acute or chronic. For vaccine development, computational methods give us the best approaches for epitope-based vaccines against different types of viruses. In the early phase of acute hepatitis infection, only CD8+ and CD4+ T cells play a major role in fighting against HCV infection. In the current study, computational identification of multiple T-cell epitopes was done for the human genome with genotype 6.

**Methods:** Various bioinformatics tools were used for the prediction of unique sequences of epitopes containing HCV glycoproteins, mainly E1 and E2 transmembrane proteins that show interaction with unique HLA allele sequences. The objective of the T-cell prediction is to find out the shortest sequence of antigen peptides that elicits the T cell immune response.

**Results:** Class I HLA alleles bound to the 8–11 amino acids, while Class II HLA bound to the 15–24 amino acids. The various tools used for epitope prediction of MHC class I

CD8+ T cells were SYFPEITHI, NetCTL1.2, and IEDB. Similarly, the tools used for MHC class II CD4+ T cells were ProPred, MHC2 Pred, and IEDB for epitope prediction. Various unique HLA allele sequences were enlisted, for which various unique epitopes were predicted. A possible treatment for HCV is not known yet.

**Conclusion:** This study gives an approach to the future identification of vaccines in the case of HCV via in silico selection of epitopes, which can be further validated by in vivo and in vitro methods.

**Keywords:** Hepatitis C virus, Epitopes, HLA alleles, T-cells.

## Introduction

The genetic structure and characteristics of HCV have been described well in the history.<sup>1</sup> The viral HCV could be a major blood-borne human infectious agent. There are currently 170 and 200 million cases of the virus worldwide, and 40,000 new cases are added each year. The HCV virus is an RNA-positive, positively polarized, 9500 nucleotide long virus.<sup>2</sup> It is considered as a significant public health issue since the virus is the etiological issue of chronic liver disease that often results in cirrhosis of the liver and

malignant hepatoma (HCC). In developed countries, the foremost vital route of HCV transmission is blood vessel misuse, whereas in resource-poor countries, invasive procedures or injection-based therapies with contaminated instruments are the predominant sources of the latest infections.

Interferon-alpha and drugs like telaprevir and sofosbuvir can be used for the treatment of the HCV infection, but these therapies are costly.<sup>3-4</sup> Since people in developing countries can't afford such expensive treatments, there is a need to develop an affordable antiviral treatment. Thus, we need to gain knowledge about the epitopes of the virus for the advancement of the vaccine. At present, two vaccines are clinically used: one is developed by Chiron, which is recombinant protein E1 and E2, and the other, which makes CD8+ active, is made from the non-structural protein of the recombinant virus.<sup>5</sup>

Cytotoxic T cells (CTLs) are effective against the short sequences of peptide epitopes that are bonded with the human leucocyte antigen (HLA) and are expressed by the antigen-presenting cells. Through this, the evasion of HCV infection by immune responses could be avoided.<sup>6</sup>

Epitopes, also called antigen determinants, are the specific short peptide sequences in an antigen that stimulate the immune response when bound to the specific antibody. T-cell receptors (TCRs) are present on the surface of T cells and recognize the APC when bonded to the major histocompatibility complex (MHC). T-cell epitopes consist of two classes: MHC Class I and MHC Class II. CD4 cells bond to the MHC Class II through the receptors present on their surface, while CD8 cells bond with the MHC Class I. The MHC is present on chromosome 6, which codes for 21 genes and shows polymorphism. The three loci present on MHC Class I (A, B, and C) are only able to detect peptide

sequences of eight to twelve amino acids in length. While Class II has HLA (DP, DQ, and DR) and detects amino acids of length eighteen to twenty-four, if we want to predict the T-cell epitopes, we must first know about the various HLA alleles.<sup>7</sup>

Various bioinformatics tools are available for the prediction of epitopes. Among this Class I bound with 8 to 11 AA, the C-terminal region and the 2nd position are known as anchor sites because these are important positions for the binding. A very limited number of proteins are acceptable for binding to the anchor site and binding to an HLA allele. Matrix-based methods are not able to correlate the relationship between the amino acids and how the binding affinity at one position of AA depends on the other amino acid position. In this study, we use three types of tools for class I prediction: SYFPEITHI, NetCTL1.2, and IEDB.

For Class II, prediction became difficult because both ends are open and allow more peptides to be bound, which shows higher polymorphism. Also, their binding motifs have weaker relativity. For their prediction, it was assumed that the 9 AA core motifs determined the affinity of the protein, MHC. This idea was used by TEPITOPE.<sup>8</sup>

In the early days, this was only for the use of PCs, but after a long time, this was provided as a part of the web server for the class II predictor, namely ProPred.<sup>9</sup> In this study, we have used IEDB, ProPred, and MHC2pred.

Epitope-based vaccines are becoming increasingly popular for the development of vaccines against viruses with conserved sequences or pathogens that have mutated rapidly. Using these various tools, appropriate in silico selection of epitopes could be possible.

## **Methods**

### ***Sequence retrieval***

## ***ViPR***

ViPR stores information for virus families that are classified as either human priority pathogens or potential public health threats. ViPR works to combine data from three different sorts of sources: (i) data transmitted to ViPR from public archives; (ii) unique data created by ViPR utilizing a variety of computational algorithms and bioinformatics methodologies; and (iii) data directly submitted to ViPR by independent researchers. In a relational database, these data are kept so that user-specified queries can quickly retrieve them.<sup>10</sup>

Whole genome sequences for E2 unique protein sequences were selected for downloading. Then, 121 protein sequences in FASTA format were retrieved for genotype VI.

### ***Multiple sequence alignment***

Multiple sequence alignment is a method for aligning more than two homologous nucleotide or amino acid sequences so that homologous residues from the respective sequences line up in columns as far as possible.

### ***Clustal Omega***

Clustal Omega is one of the bioinformatics tools used for aligning various multiple amino acid and nucleotide sequences precisely and rapidly. For decades, this has become one of the most used bioinformatics tools since it is a requirement for most phylogenetic or comparative studies of homologous genes or proteins.<sup>11</sup> In this study, downloaded E2 protein sequences of Hepatitis C virus genotype 6 were used as input for Clustal Omega for multiple sequence alignment.

### ***Conservancy analysis***

#### **AVANA**

AVANA was further used to analyze the conserved region in the E2 protein after multiple sequence alignments through Clustal Omega. AVANA stands for Antigen Variability Analyzer Tool. This was used for the alignment of various subsets of the collected sequence that were entirely based on the viral subtype of the host. The tool provides useful information about the conserved region that is entirely based on information entropy analysis.<sup>12</sup> Information entropy is measured by the uncertainty of a random variable, or in the case of proteins, by protein mutation.<sup>11</sup> Highly conserved sites are shown due to lower entropy, which displays the degree of randomness.

To retrieve the conserved region of E2 protein, multiple sequence alignment of E2 protein in FASTA format was used by setting the parameter in AVANA as conservation region 70%, with a minimum length of 9 bp and a maximum length of 25 bp.

### ***T-cell epitope prediction***

#### ***Tools used for CLASS I MHC Epitope and HLA allele prediction***

##### ***SYFPEITHI***

The MHC databank is called SYFPEITHI. It regularly updates its database of MHC class I and class II ligands and peptide patterns from humans and other species, including apes, cattle, chicken, and mice, among others. There are distinct entries available for each motif that is currently available. It is possible to conduct searches for references, source proteins/organisms, T-cell epitopes, natural ligands, MHC alleles, and MHC motifs. Links are provided to the EMBL and PubMed databases.<sup>13</sup>

In this study, MHC-I prediction was performed by SYFPEITHI. To perform this task, all HLA were selected from HLA-A \* 01 through HLA-DRB1 \* 1501 (DR2b). Other parameters include the selection of

nanomers, and 70% of the conserved sequence obtained from the above step was used as input. As a result, the HLA, epitope, and score appeared. The score must be less than or equal to 20.

### ***NetCTL1.2***

The NetCTL1.2 service is used to predict CTL epitopes in protein sequences. Version 1.2 of the MHC class I binding prediction now includes 12 MHC supertypes, including supertypes A26 and B39. In the technique, the efficiency of TAP transport, proteasomal C-terminal cleavage, and peptide MHC class I binding are all predicted. The service can be used to anticipate CTL epitopes that are restricted to 12 MHC class I supertypes. In order to achieve MHC class I binding and proteasomal cleavage, artificial neural networks are used. To forecast TAP transit effectiveness, use the weight matrix.<sup>14</sup> Subtype A1 was used to choose input from 70% of conserved sequences, starting with C1. The ordering by score was then altered to "Combination Score."

The remaining C1-C7 sequences were processed using the same procedures. The threshold for epitope recognition, the weight of C-terminal cleavage, and the weight of TAP transport effectiveness are all unchanged.

### ***IEDB***

Researchers from the La Jolla Institute for Allergy and Immunology (LIAI) have developed the Immune Epitope Database and Analysis Resource (IEDB) in collaboration with the Department of Health and Human Services (HHS) and the National Institute of Allergy and Infectious Diseases (NIAID). In order to generate new research tools, diagnostic techniques, vaccines, and therapies, immunological epitope data must be disseminated. Data curation for the ongoing NIAID Emerging and Re-emerging

Infectious Diseases and Category A, B, and C priority pathogens (which includes influenza) is given top priority. Epitopes from diverse allergens, autoantigens, and infectious pathogens are also being catalogued. The database also contains information on immune epitopes from the TopBank, FIMM, HLA Ligand, and MHC binding databases, as well as information on MHC binding from a number of antigenic sources.<sup>15</sup>

C1-C7 conserved sequences were used as input, much as the prior technique. The ANN.4.0.0 prediction format was chosen. When you type HLA-A \* 01:01, the outcome is shown. Afterward, choose HLA and epitope based on a score of 500 or fewer.

### ***Tools used for CLASS II MHC epitope and HLA allele prediction***

#### ***PROPPRED***

MHC class II binding sites in antigenic protein sequences were predicted using ProPred, a graphical online application. The server uses an amino-acid/position coefficient database derived from literature to create a matrix-based prediction method. In the graphical interface, the predicted binders may be seen as peaks, whereas in the HTML interface, they can be seen as colored residues. This website might help researchers find promiscuous binding areas that can connect to several HLA-DR alleles.<sup>9</sup>

Conserved sequences obtained from AVANA were used as input sequences. Select HLA alleles and epitopes on the basis of red and blue color. At a time, only nine sequences can be selected from ProPred.

#### ***MHC2 Pred***

An SVM-based method is employed to predict the peptides that bind to MHC class II. SVM-based techniques are 80% accurate for 42 alleles.

Because the dataset was smaller, the method's performance was worse for a few alleles. The method's performance was evaluated using 5-fold cross-validation. MHC2Pred input sequences were 70% AVANA conserved sequences for the human allele HLA-DR1-HLA-DRB5 \* 0101, with a score of 0.5. The display mode would be tabular, and the tabular results would be 50. The results were then presented by HLA, epitope, and score. The score must be 0.5 or greater.

### ***IEDB Analysis resource***

MHC-2 prediction using IEDB was similar to one that we used for MHC-1 epitope prediction. By selecting human, HLA-DR, HLA-DP, and HLA-DQ species or loci one at a time. Then the results appear, with HLA,

core sequence, peptide sequence, and score. Their selection was based on the score, which must be less than or equal to 500.

## ***Results***

### ***Sequence retrieval***

#### ***ViPR***

In sequence retrieval, the unique sequence of 121 E2 proteins of the human genome had genotype 6, was retrieved in the FASTA format, and all had a unique sequence ID for each protein. The results for the sequence retrieval are shown in the **Table 1 A unique sequence of E2 protein in FASTA format**.

**Table 1 A unique sequence of E2 protein in FASTA format**

| <b>S. No.</b> | <b>GenBank ID</b> | <b>Host</b> | <b>Country</b> | <b>Sequence ID</b>                 |
|---------------|-------------------|-------------|----------------|------------------------------------|
| 1             | KC191671          | Human       | Malaysia       | >gb:VIPR_ALG4_440234843_1468_2565  |
| 2             | KJ567649          | Human       | Vietnam        | >gb:VIPR_ALG4_685430775_1488_2594  |
| 3             | KJ567648          | Human       | Vietnam        | >gb:VIPR_ALG4_685430773_1489_2583  |
| 4             | KJ567646          | Human       | Vietnam        | >gb:VIPR_ALG4_685430769_1488_2588  |
| 5             | KJ567651          | Human       | Vietnam        | >gb:VIPR_ALG4_685430779_1488_2597  |
| 6             | KJ567650          | Human       | Vietnam        | >gb:VIPR_ALG4_685430777_1488_2594  |
| 7             | KJ567652          | Human       | Vietnam        | >gb:VIPR_ALG4_685430781_1488_2591  |
| 8             | KJ567645          | Human       | Vietnam        | >gb:VIPR_ALG4_685430767_1488_2582  |
| 9             | KJ567644          | Human       | Vietnam        | >gb:VIPR_ALG4_685430765_1473_2567  |
| 10            | KJ567647          | Human       | Vietnam        | >gb:VIPR_ALG4_685430771_1489_2580  |
| 11            | KM504109          | Human       | -N/A           | >gb:VIPR_ALG4_758739587_1489_2580  |
| 12            | KJ678744          | Human       | China          | >gb:VIPR_ALG4_758739124_1492_2595  |
| 13            | KJ678745          | Human       | China          | >gb:VIPR_ALG4_758739126_1492_2595  |
| 14            | KM504110          | Human       | Thailand       | >gb:VIPR_ALG4_758739589_1488_2576  |
| 15            | KY014622          | Human       | China          | >gb:VIPR_ALG4_AQT18930_1_1489_2583 |
| 16            | JX183557          | Human       | China          | >gb:VIPR_ALG4_407751085_1488_2585  |
| 17            | KJ678746          | Human       | China          | >gb:VIPR_ALG4_758739128_1449_2552  |
| 18            | KJ678747          | Human       | China          | >gb:VIPR_ALG4_758739130_1449_2552  |
| 19            | KJ678748          | Human       | China          | >gb:VIPR_ALG4_758739132_1429_2532  |
| 20            | KJ678749          | Human       | China          | >gb:VIPR_ALG4_758739134_1448_2551  |
| 21            | KJ678750          | Human       | China          | >gb:VIPR_ALG4_758739136_1449_2552  |
| 22            | KJ678751          | Human       | China          | >gb:VIPR_ALG4_758739138_1449_2552  |
| 23            | KJ678752          | Human       | China          | >gb:VIPR_ALG4_758739140_1425_2528  |
| 24            | KJ678753          | Human       | China          | >gb:VIPR_ALG4_758739142_1449_2549  |
| 25            | KJ678754          | Human       | China          | >gb:VIPR_ALG4_758739144_1492_2595  |
| 26            | KJ678755          | Human       | China          | >gb:VIPR_ALG4_758739146_1449_2552  |
| 27            | KJ678756          | Human       | China          | >gb:VIPR_ALG4_758739148_1491_2594  |
| 28            | KJ678757          | Human       | China          | >gb:VIPR_ALG4_758739150_1491_2594  |

| S. No. | GenBank ID | Host  | Country | Sequence ID                       |
|--------|------------|-------|---------|-----------------------------------|
| 29     | KJ678758   | Human | China   | >gb:VIPR_ALG4_758739152_1491_2594 |
| 30     | KJ678759   | Human | China   | >gb:VIPR_ALG4_758739154_1491_2594 |
| 31     | KJ678760   | Human | China   | >gb:VIPR_ALG4_758739156_1491_2594 |
| 32     | KJ678761   | Human | China   | >gb:VIPR_ALG4_758739158_1491_2594 |
| 33     | KJ678762   | Human | China   | >gb:VIPR_ALG4_758739160_1491_2594 |
| 35     | KJ678764   | Human | China   | >gb:VIPR_ALG4_758739164_1491_2594 |
| 36     | KJ678765   | Human | China   | >gb:VIPR_ALG4_758739166_1491_2594 |
| 37     | KJ678766   | Human | China   | >gb:VIPR_ALG4_758739168_1449_2552 |
| 38     | KJ678767   | Human | China   | >gb:VIPR_ALG4_758739170_1449_2552 |
| 39     | KJ678768   | Human | China   | >gb:VIPR_ALG4_758739172_1446_2549 |
| 40     | KJ678769   | Human | China   | >gb:VIPR_ALG4_758739174_1449_2552 |
| 41     | KJ678770   | Human | China   | >gb:VIPR_ALG4_758739176_1449_2552 |
| 42     | JX183549   | Human | China   | >gb:VIPR_ALG4_407751069_1488_2579 |
| 43     | KM252792   | Human | China   | >gb:VIPR_ALG4_749396734_1488_2585 |
| 44     | KM252793   | Human | Laos    | >gb:VIPR_ALG4_749396736_1488_2579 |
| 45     | KM252794   | Human | Laos    | >gb:VIPR_ALG4_749396738_1488_2585 |
| 46     | KM252789   | Human | Laos    | >gb:VIPR_ALG4_749396728_1491_2594 |
| 47     | KM252795   | Human | Laos    | >gb:VIPR_ALG4_749396740_1488_2582 |
| 48     | KM252796   | Human | Laos    | >gb:VIPR_ALG4_749396742_1488_2582 |
| 49     | KM252797   | Human | Laos    | >gb:VIPR_ALG4_749396744_1474_2568 |
| 50     | KM252790   | Human | Laos    | >gb:VIPR_ALG4_749396730_1491_2597 |
| 51     | JX183554   | Human | Laos    | >gb:VIPR_ALG4_407751079_1488_2582 |
| 52     | KM252798   | Human | Laos    | >gb:VIPR_ALG4_749396746_1491_2582 |
| 53     | KM252799   | Human | Laos    | >gb:VIPR_ALG4_749396748_1491_2591 |
| 54     | KM252800   | Human | Laos    | >gb:VIPR_ALG4_749396750_1488_2579 |
| 55     | KM252791   | Human | Laos    | >gb:VIPR_ALG4_749396732_1491_2597 |
| 56     | KM587629   | Human | USA     | >gb:VIPR_ALG4_751660982_1488_2585 |
| 57     | KM587630   | Human | USA     | >gb:VIPR_ALG4_751660984_1488_2582 |
| 58     | KM504112   | Human | Canada  | >gb:VIPR_ALG4_758739593_1488_2588 |
| 59     | KM504113   | Human | Canada  | >gb:VIPR_ALG4_758739595_1489_2580 |
| 60     | KM504114   | Human | Canada  | >gb:VIPR_ALG4_758739597_1488_2579 |
| 61     | KJ678771   | Human | Canada  | >gb:VIPR_ALG4_758739178_1449_2552 |
| 62     | KJ678772   | Human | Canada  | >gb:VIPR_ALG4_758739180_1449_2552 |
| 63     | KJ678773   | Human | Canada  | >gb:VIPR_ALG4_758739182_1449_2552 |
| 64     | KM504115   | Human | Canada  | >gb:VIPR_ALG4_758739599_1489_2580 |
| 65     | KM504116   | Human | Canada  | >gb:VIPR_ALG4_758739601_1489_2580 |
| 66     | KM504118   | Human | Canada  | >gb:VIPR_ALG4_758739603_1488_2579 |
| 67     | KJ678774   | Human | Canada  | >gb:VIPR_ALG4_758739184_1452_2555 |
| 68     | KM504118   | Human | Canada  | >gb:VIPR_ALG4_758739605_1488_2579 |
| 69     | KM504119   | Human | Canada  | >gb:VIPR_ALG4_758739607_1488_2570 |
| 70     | KJ678775   | Human | Canada  | >gb:VIPR_ALG4_758739186_1449_2552 |
| 71     | KM504120   | Human | Canada  | >gb:VIPR_ALG4_758739609_1488_2576 |
| 72     | KM504121   | Human | Canada  | >gb:VIPR_ALG4_758739611_1489_2580 |
| 73     | JX183558   | Human | Canada  | >gb:VIPR_ALG4_407751087_1488_2582 |
| 74     | JX183550   | Human | Canada  | >gb:VIPR_ALG4_407751071_1488_2579 |
| 75     | KM504122   | Human | Canada  | >gb:VIPR_ALG4_758739613_1488_2579 |
| 76     | KJ678776   | Human | Canada  | >gb:VIPR_ALG4_758739188_1449_2552 |
| 77     | KJ678777   | Human | Canada  | >gb:VIPR_ALG4_758739190_1449_2549 |
| 78     | KM504124   | Human | Canada  | >gb:VIPR_ALG4_758739617_1489_2577 |
| 79     | KJ678778   | Human | Canada  | >gb:VIPR_ALG4_758739192_1449_2552 |
| 80     | KJ678779   | Human | Canada  | >gb:VIPR_ALG4_758739194_1449_2552 |

| S. No. | GenBank ID | Host  | Country | Sequence ID                       |
|--------|------------|-------|---------|-----------------------------------|
| 81     | KJ678780   | Human | Canada  | >gb:VIPR_ALG4_758739196_1449_2549 |
| 82     | KM504111   | Human | Canada  | >gb:VIPR_ALG4_758739591_1488_2579 |
| 83     | KJ678781   | Human | China   | >gb:VIPR_ALG4_758739198_1492_2595 |
| 84     | KJ678782   | Human | China   | >gb:VIPR_ALG4_758739200_1492_2595 |
| 85     | KJ678784   | Human | China   | >gb:VIPR_ALG4_758739202_1492_2595 |
| 86     | KJ678784   | Human | China   | >gb:VIPR_ALG4_758739204_1424_2527 |
| 87     | KJ678785   | Human | China   | >gb:VIPR_ALG4_758739206_1449_2552 |
| 88     | JX183551   | Human | Vietnam | >gb:VIPR_ALG4_407751073_1488_2582 |
| 89     | KM252779   | Human | Vietnam | >gb:VIPR_ALG4_749396708_1488_2588 |
| 90     | JX183555   | Human | Vietnam | >gb:VIPR_ALG4_407751081_1488_2582 |
| 91     | KM252780   | Human | Vietnam | >gb:VIPR_ALG4_749396710_1488_2588 |
| 92     | KM252787   | Human | Vietnam | >gb:VIPR_ALG4_749396724_1488_2582 |
| 93     | KM252786   | Human | Vietnam | >gb:VIPR_ALG4_749396722_1488_2579 |
| 94     | KM252782   | Human | Vietnam | >gb:VIPR_ALG4_749396714_1489_2583 |
| 95     | KM252783   | Human | Vietnam | >gb:VIPR_ALG4_749396716_1474_2562 |
| 96     | KM252785   | Human | Vietnam | >gb:VIPR_ALG4_749396720_1489_2580 |
| 97     | KJ678786   | Human | Vietnam | >gb:VIPR_ALG4_758739208_1492_2595 |
| 98     | KJ678787   | Human | Vietnam | >gb:VIPR_ALG4_758739210_1449_2552 |
| 99     | KJ678788   | Human | Vietnam | >gb:VIPR_ALG4_758739212_1491_2594 |
| 100    | JX183552   | Human | Vietnam | >gb:VIPR_ALG4_407751075_1488_2582 |
| 101    | KJ678789   | Human | Vietnam | >gb:VIPR_ALG4_758739214_1491_2594 |
| 102    | KJ678790   | Human | Vietnam | >gb:VIPR_ALG4_758739216_1449_2552 |
| 103    | KJ678791   | Human | -N/A-   | >gb:VIPR_ALG4_758739218_1492_2595 |
| 104    | KJ678792   | Human | Vietnam | >gb:VIPR_ALG4_758739220_1492_2595 |
| 105    | JX183556   | Human | Vietnam | >gb:VIPR_ALG4_407751083_1488_2582 |
| 106    | KM252782   | Human | -N/A-   | >gb:VIPR_ALG4_749396712_1488_2588 |
| 107    | KJ678793   | Human | Vietnam | >gb:VIPR_ALG4_758739222_1449_2552 |
| 108    | KJ678794   | Human | Vietnam | >gb:VIPR_ALG4_758739224_1492_2595 |
| 109    | KJ678795   | Human | Vietnam | >gb:VIPR_ALG4_758739226_1492_2595 |
| 110    | KJ678796   | Human | Vietnam | >gb:VIPR_ALG4_758739228_1449_2552 |
| 111    | JX183553   | Human | Vietnam | >gb:VIPR_ALG4_407751077_1488_2582 |
| 112    | KM252788   | Human | Vietnam | >gb:VIPR_ALG4_749396726_1489_2586 |
| 113    | KJ678797   | Human | Vietnam | >gb:VIPR_ALG4_758739230_1449_2552 |
| 114    | KJ678798   | Human | -N/A-   | >gb:VIPR_ALG4_758739232_1449_2552 |
| 115    | KM252784   | Human | Vietnam | >gb:VIPR_ALG4_749396718_1487_2581 |
| 116    | KC844040   | Human | China   | >gb:VIPR_ALG4_576294938_1488_2585 |
| 117    | KC844039   | Human | China   | >gb:VIPR_ALG4_576294935_1488_2600 |
| 118    | HQ912955   | Human | China   | >gb:VIPR_ALG4_324330491_1150_2253 |
| 119    | HQ912954   | Human | China   | >gb:VIPR_ALG4_324330489_1150_2253 |
| 120    | KC844037   | Human | China   | >gb:VIPR_ALG4_576294931_1448_2551 |
| 121    | KC844038   | Human | China   | >gb:VIPR_ALG4_576294933_1448_2551 |

### ***Multiple Sequence Alignment***

#### ***Clustal Omega***

The multiple sequence alignment shows the Consensus sequence of each unique 121 E2 protein sequence. The results for E2 protein

sequences of HCV genotype 6 is shown below:

#### ***GENOTYPE VI***

```
xxTTTVGxAVxRtTxGLxxLFSxGxKQNL
QLINTNGSWHINRTALNCNDSLQTFIA
SLFYxHKFNSSGCPERMAACKPLxDFR
```

QGWGxITYKxNISGPSDDRPYCWHYAP  
 RPCxVVPARTVCGPVYCFTPSPVVVGT  
 TDRRGNPTYTWGENETDVMLES LRPP  
 xGGWFGCTWMNSTGFTKTCGAPPCQI  
 VPGDYNsxxxSxNELLCPTDCFRKHPEA  
 TYQRCGSGPWLTPRCLVDYPYRLWHY  
 PCTVNFTLHKVRMFVGGIEHRFDAACN  
 WTRGERCELxDRDRIEMSPLLFSTTQLA  
 ILPCSFTTMPALSTGLIHLHQNIVDVQY  
 LYGVSSSVTSWVVKWEYVVLMLFLVA  
 DARICTCLWLMLLIxVEA

**Conservancy analysis**

**AVANA**

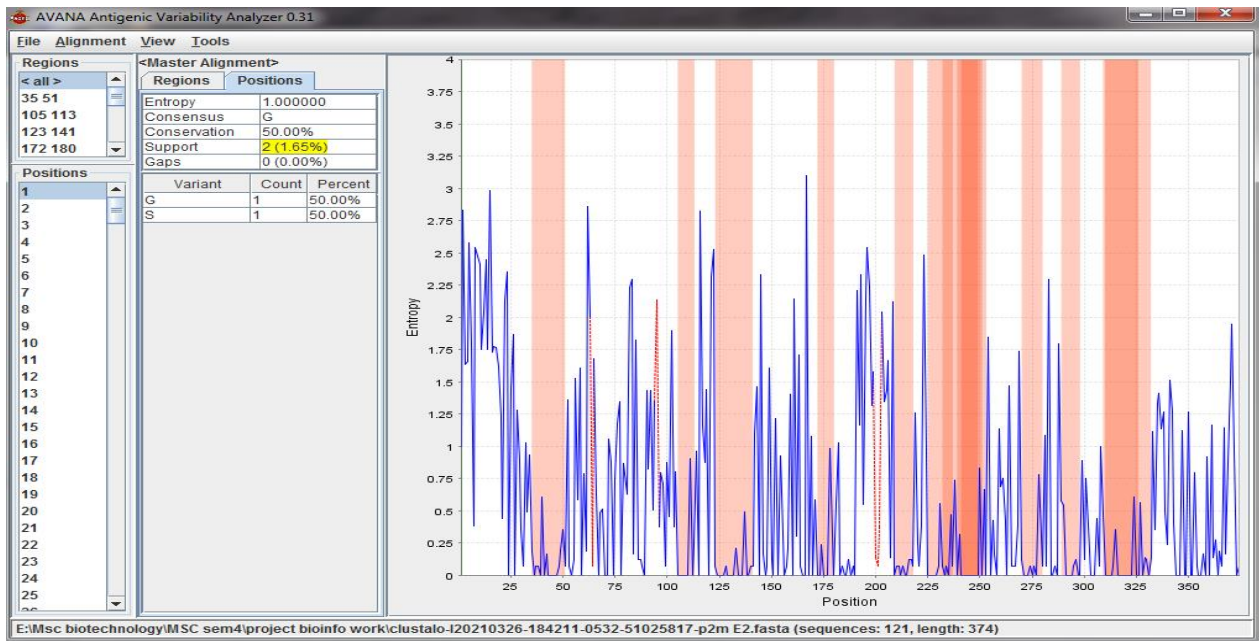
The tool AVANA was used for the conservancy analysis. In this analysis, we took 70% of the conserved region as it shows the more unique sequence had a sequence length greater than 9. C1, C2, C3, C4, C5, C6, and C7 show the unique and non-repeated sequences with lengths greater than 9 (

Table 2 and Figure 1). This also represents the location of each sequence in the consensus sequence.

**Table 2 70% conserved region had unique sequence more than 9 length**

| Sr. No. | Sequence                      | Location |
|---------|-------------------------------|----------|
| C1      | NGSWHINRTALNCNDSL             | 35-51    |
| C2      | VCGPVYCFTPSPVVVGT             | 122-140  |
| C3      | CPTDCFRKHP                    | 208-217  |
| C4      | CGSGPWLTPRCLVDYPYRLWHYPCTVNFT | 224-252  |
| C5      | AACNWTRGERCE                  | 269-280  |
| C6      | EMSPLLSTTQ                    | 288-298  |
| C7      | TMPALSTGLIHLHQNIVDVQYLYG      | 308-331  |

**Figure 1 Orange color shows the 70% conserved region by using AVANA**



**Epitope prediction**

It was done for both MHC class I and II with different tools. The results of these tools are

mentioned below based on their category and tools.

**MHC Class I**

In class I epitope prediction, epitopes were predicted for conserved sequences C1, C2, C4, and C7 by using the tools SYFEITHI and NetCTL1.2 (**Table 3**). In the case of the IEDB epitope, it was predicted for conserved sequences C2, C4, C6, and C7. The SYFEITHI tool had predicted 1 epitope

for C1, 2 epitopes for C2, 10 for C4, and 7 for C7. Similarly, NetCTL1.2 predicted 1 epitope for C1, 2 epitopes for C2, 10 for C4, and 6 for C7, while 1 epitope for C2, 7 epitopes for C4, 1 epitope for C5, 1 epitope for C6, and 6 epitopes for C7 were predicted by IEDB (**Table 3**).

**Table 3 Epitopes predicted by SYFEITHI, NetCTL1.2, and IEDB tools for MHC I**

| Conserved Sequence | SYFEITHI   | NetCTL1.2   | IEDB  |
|--------------------|--|---|---|
| C1                 | TALNCNDSL  | SWHINRTAL   | Not applicable  |
| C2                 | TPSPVVVGT, CFTSPVVV  | VYCFTSPV,GPVYCFTPS  | VYCFTSPV  |
| C4                 | LTPRCLVDY,RLWHYPCTV,C<br>LVDYPYRL,WLTPRCLVD,GS<br>GPWLTPR,<br>DYPYRLWHY,GPWLTPRCL,<br>WHYPCTVNF,PRCLVDYPY<br>LVDYPYRLW | LTPRCLVDY,RLWHYPCTV,C<br>LVDYPYRL,DYPYRLWHY,H<br>YPCTVNFT,<br>DYPYURLWHY,GPWLTPRCL<br>,YRLWHYPCT,PRCLVDYPY<br>WHYPCTVNF | RLWHYPCTV,GSGPWL<br>TPR,CLVDYPYRL,WHY<br>PCTVNF,<br>RCLVDYPYR,GPWLTPR<br>CL,LVDYPYRLW |
| C5                 | Not applicable   | Not applicable  | ACNWTRGER   |
| C6                 | Not applicable   | Not applicable  | EMSPLLFST   |
| C7                 | ALSTGLIHL,HLHQNIVDV,G<br>LIHLHQNI<br>LIHLHQNIV,TMPALSTGL,H<br>QNIVDVQY<br>MPALSTGLI                                    | NIVDVQYLY,HQNIVDVQY,<br>ALSTGLTHL,TMPALSTGL<br>,HLHQNIVDV,MPALSTGLI   | NIVDVQYLY,ALSTGLI<br>HL,MPALSTGLI,TMPAL<br>STGL,<br>HQNIVDVQY,GLIHLHQ<br>NI           |

**MHC Class II**

In class II epitope prediction, epitopes were predicted for conserved sequences C1, C2, C4, and C7 by using ProPred, while MHC2Pred predicted epitopes for all conserved sequences C1, C2, C3, C4, C5, C6, and C7. Whereas in the case of IEDB, it was predicted for conserved sequences C1, C2, C4, C5, C6, and C7. ProPred predicted 1

epitope for C1, 3 for C2, 3 for C4, and 3 for C7 as well. MHC2Pred predicted 9 epitopes for C1, 11 for C2, 2 for C3, 21 for C4, 5 for C5, 3 for C6, and 14 for C7. IEDB predicted 4 epitopes for C1, 12 for C2, 16 for C4, 2 each for C5, and C6, and 16 for C7 (**Table 4**).

As compared to MHC class I tools, MHC class II tools have predicted epitopes for more conserved sequences.

**Table 4 Epitopes predicted by ProPred, MHC2Pred, and IEDB tools for MHCII**

| Conserved Sequence | ProPred   | MHC2Pred   | IEDB                                       |
|--------------------|-----------|--|--|
| C1                 | WHINRTALN | NGSWHINRT,INRTALNCN,N<br>RTALNCND,WHINRTALN,<br>GSHWIRTA,TALNCNDSL,H<br>INRTALNC , | WHINRTALN,GSHWIRTA,I<br>NRTALNCN,SWHINRTAL |

| Conserved Sequence | ProPred                               | MHC2Pred  | IEDB  |
|--------------------|---------------------------------------|---|---|
|                    |                                       | SWHINRTAL, RTALNCNDS  |   |
| C2                 | VYCFTPSPV<br>,FTPSPVVVG,<br>YCFTPSPVV | PSPVVVGGT,CGPVYCFTP,SP<br>VVVGGTT,FTPSPVVVG,<br>CFTPSPVVV,YCFTPSPVV,TP<br>SPVVVGT,VCGPVYCFT,GPV<br>YCFTPS<br>PVYCFTPSP,VYCFTPSPV  | YCFTPSPVV,VYCFTPSPV,CF<br>TPSPVVV,FTPSPVVVG,VCG<br>PVYCFT,PVYCFTPSP<br>TPSPVVVGT,YCFTPSPVX,GP<br>VYCFTPS,SPVVVGGTT,CGP<br>VYCFTP,PSPVVVGGT  |
| C3                 | Not applicable                        | CPTDCFRKH,PTDCFRKHP<br>PWLTPRCLV,CGSGPWLT,H<br>YPCTVNFT,LWHYPCTVN,W<br>LTPRCLVD,<br>LVDYPYRLW,TPRCLVDYP,R<br>LWHYPCTV,LTPRCLVDY,GP<br>WLTPRCL,<br>SGPWLT,PRC,CLVDYPYRL,D<br>YPYRLWHY,RCLVDYPYR,W<br>HYPCTVNF<br>GSGPWLT,PRCLVDYPY,Y<br>PYRLWHYP,PYRLWHYPC,V<br>DYPYRLWH,<br>YRLWHYPCT | Not applicable  |
| C4                 | WHYPCTVNF<br>,YRLWHYPCT,<br>YRLWHYPC  | NWTRGERCE,CNWTRGERC,<br>NWTRGERC,EACNWTRGER,<br>AACNWTRGE   | YPYRLWHYP,CLVDYPYRL,P<br>YRLWHYPC,LVDYPYRLW,V<br>DYPYRLWH,<br>LWHYPCTVN,RLWHYPCTV,<br>YRLWHYPCT,WHYPCTVNF,<br>PWLTPRCLV,<br>RCLVDYPYR,CGSGPWLT,L<br>TPRCLVDY,WLTPRCLVD,PR<br>CLVDYPY<br>HYPCTVNFT |
| C5                 | Not applicable                        | MSPLLSTT,EMSPLLST,SP<br>LLSTTQ  | ACNWTRGER,CNWTRGERC   |
| C6                 | Not applicable                        | STGLIHLHQ,LSTGLIHLH,HQ<br>NIVDVQY,QNIVDVQYL,LHQ<br>NIVDVQ,MPALSTGLI<br>HLHQNIVDV,IVDVQYLYG,T<br>MPALSTGL,NIVDVQYLY,TG<br>LIHLHQN,PALSTGLIH<br>LIHLHQNIV,IHLHQNIVD,AL<br>STGLIHL,GLIHLHQNI   | EMSPLLST,MSPLLSTT   |
| C7                 | LIHLHQNIV,<br>LHQNIVDVQ,<br>IHLHQNIVD | LIHLHQNIV,IHLHQNIVD,AL<br>STGLIHL,GLIHLHQNI   | LIHLHQNIV,LSTGLIHLH,IHL<br>HQNIVD,ALSTGLIHL,HLHQ<br>NIVDV,GLIHLHQNI<br>MPALSTGLI,LHQNIVDVQ,P<br>ALSTGLIH,NIVDVQYLY,TG<br>LIHLHQN,TMPALSTGL<br>IVDVQYLYG,HQNIVDVQY,<br>QNIVDVQYL,STGLIHLHQ         |

### HLA analysis

#### MHC Class I

In class I HLA allele prediction, HLA alleles were predicted for conserved sequences C1, C2, C4, and C7 by tools SYFPEITHI and NetCTL 1.2. But in the case of IEDB, HLA

alleles are predicted for C2, C4, C5, C6, and C7. SYFPEITHI predicted 1 HLA allele for C1, 2 for C2, 15 for C4, 4 for C7, whereas NetCTL1.2 predicted 3 HLA alleles for C1, 2 for C2, 6 for C4, 5 alleles for C7. IEDB predicted 1 HLA allele for C2, 12 for C4, 1 for C5, 2 for C6, and 12 for C7 (**Table 5**).

**Table 5 HLA alleles predicted by SYFPEITHI, NetCTL1.2, and IEDB tools for MHC class I**

| Conserved Sequence | SYFPEITHI   | NetCTL1.2             | IEDB  |
|--------------------|---|-----------------------|---|
| C1                 | HLA-B*51:01   | A24,B8,B39            | Not applicable  |
| C2                 | HLA-B*07:02 ,HLA-B*15:16<br>HLA-A*01,HLA-A*02:01,HLA-A*03 ,HLA-A*11:01,HLA-A*26 ,<br>HLA-A*68:01 ,HLA-B*07:02,HLA-B*13 ,HLA-B*15:10<br>HLA-B*15:16 ,HLA-B*27:05 ,HLA-B*35:01<br>HLA-B*38:01 ,HLA-B*51:01 ,HLA-B*58:02 | A24,B7                | HLA-A*23:01<br><br>HLA-A*02:01,HLA-A*02:02,HLA-A*02:03<br>HLA-A*02:04,HLA-A*02:05,HLA-A*02:06<br>HLA-A*02:07,HLA-A*02:08,HLA-A*02:09<br>HLA-A*02:10,HLA-A*02:11,HLA-A*02:12         |
| C4                 | Not applicable  | A1,A2,A24,A26,B27,B39 | HLA-A*31:01   |
| C5                 | Not applicable  | Not applicable        | HLA-A*02:06,HLA-A*68:02<br>HLA-A*29:02,HLA-A*02:01,HLA-A*02:06<br>HLA-A*68:01,HLA-B*07:02,HLA-B*53:01<br>HLA-A*26:01,HLA-A*30:02,HLA-B*51:01<br>HLA-B*15:01,HLA-B*35:01,HLA-A*68:02 |
| C6                 | Not applicable  | Not applicable        |   |
| C7                 | HLA-A*02:01 ,HLA-B*13 ,HLA-B*15:01 (B62) ,HLA-B*51:01   | A1,A2,A26,B7,B62      |   |

**MHC Class II**

In the case of Class II HLA prediction, HLA alleles were predicted for conserved sequences C1, C2, C4, and C7 by using the tool ProPred. In MHC2Pred, it was predicted for conserved sequences C1, C2, C3, C4, C5, C6, and C7 and in the case of IEDB, the HLA allele was predicted for conserved sequences C1, C2, C4, C5, C6, and C7. ProPred

predicted 2 HLA alleles for C1, 8 for C2, 6 for C4, and 14 for C7. MHC2Pred predicted HLA allele 17 for C1, 16 for C2, 5 for C3, 22 for C4, 5 for C5, 9 for C6, and 23 for C7. The IEDB tool predicted 17 HLA alleles for C1, 32 for C2, 38 for C4, 3 for C5, 4 for C6, and 35 for C7. The maximum HLA alleles were predicted by the IEDB tool for MHCII as compared to MHCI (**Table 6**).

**Table 6 HLA alleles predicted by ProPred, MHC2Pred, and IEDB tools for MHC II**

| Conserved Sequence | ProPred             | MHC2Pred  | IEDB   |
|--------------------|---------------------|---|--|
| C1                 | DRB1_0305,DRB1_0309 | HLA-DR9,HLA-DR2,HLA-DR3,HLA-DR51<br>HLA-DQ4,HLA-DQ6,HLA-DQ7,HLA-DQ8 | HLA-DRB3*02:02,HLA-DQA1*02:01/DQB1*04:02<br>HLA-DQA1*03:03/DQB1*04:02<br>HLA-DQA1*06:01/DQB1*04:02 |

| Conserved Sequence | ProPred   | MHC2Pred   | IEDB   |
|--------------------|---|--|--|
|                    |   | HLA-DQB1*03,HLA-DQB1*0301,HLA-DRB1*0101<br>HLA-DRB4*0101,HLA-DRB1*0301,HLA-DRB1*0405<br>HLA-DRB1*0901,HLA-DRB1*1101,HLA-DRB1*1501  | HLA-DRB1*04:05,HLA-DRB1*11:01<br>HLA-DRB1*08:01,HLA-DQA1*05:01/DQB1*04:02<br>HLA-DRB3*03:01,HLA-DQA1*01:02/DQB1*06:02<br>HLA-DQA1*01:02/DQB1*05:01<br>HLA-DRB1*13:02,HLA-DRB1*07:01<br>HLA-DRB1*01:01,HLA-DRB1*04:04<br>HLA-DRB1*13:01,HLA-DRB4*01:03<br>HLA-DRB1*07:01,HLA-DRB1*09:01<br>HLA-DQA1*05:01/DQB1*03:03<br>HLA-DQA1*06:01/DQB1*04:02<br>HLA-DQA1*02:01/DQB1*04:02<br>HLA-DQA1*05:01/DQB1*03:02,HLA-DRB3*02:02<br>HLA-DQA1*05:01/DQB1*04:02<br>HLA-DQA1*02:01/DQB1*03:03<br>HLA-DQA1*02:01/DQB1*03:01,HLA-DRB3*03:01<br>HLA-DQA1*01:03/DQB1*06:03,HLA-DRB1*04:01<br>HLA-DRB1*04:04,HLA-DPA1*01:03/DPB1*02:01<br>HLA-DRB1*13:02,HLA-DRB1*04:05<br>HLA-DQA1*05:01/DQB1*03:01<br>HLA-DRB1*16:02,HLA-DRB1*01:01<br>HLA-DRB1*08:02,HLA-DPA1*01:03/DPB1*06:01<br>HLA-DRB1*10:01,HLA-DQA1*03:01/DQB1*03:01<br>HLA-DRB1*13:01,HLA-DQA1*01:02/DQB1*05:01<br>HLA-DRB4*01:03,HLA-DRB1*08:01<br>HLA-DRB1*15:01,HLA-DRB1*11:01<br>HLA-DRB3*01:01,HLA-DPA1*02:01/DPB1*14:01 |
| C2                 | DRB1_0309,DRB1_0309,DRB1_0421<br>DRB1_0423,DRB1_0701,DRB1_0703<br>DRB1_1501,DRB1_1506 | HLA-DR9,HLA-DR2,HLA-DR3,HLA-DR13<br>HLA-DR15,HLA-DR52,HLA-DQ4,HLA-DQ6<br>HLA-DQ7,HLA-DQ8,HLA-DQB1*0301<br>HLA-DRB1*0101,HLA-DRB5*0101,HLA-DRB1*0401<br>HLA-DRB1*0405,HLA-DRB1*0901 |  |
| C3                 | Not applicable  | HLA-DR9,HLA-DQ4,HLA-DQ7,HLA-DRB1*0405  | Not applicable   |

| Conserved Sequence | ProPred  | MHC2Pred  | IEDB  |
|--------------------|--|---|---|
|                    |  | HLA-DRB1*0901   | HLA-DQA1*05:01/DQB1*04:02,HLA-DRB3*01:01<br>HLA-DPA1*01:03/DPB1*06:01<br>HLA-DPA1*01:03/DPB1*02:01,HLA-DRB3*03:01<br>HLA-DRB1*15:01,HLA-DPA1*03:01/DPB1*04:02<br>HLA-DQA1*01:01/DQB1*05:01,HLA-DRB1*03:01<br>HLA-DRB4*01:03,HLA-DRB1*09:01<br>HLA-DRB1*04:04,HLA-DRB1*16:02<br>HLA-DPA1*01:03/DPB1*03:01,HLA-DRB1*13:01<br>HLA-DRB1*04:05,HLA-DQA1*02:01/DQB1*04:02<br>HLA-DRB1*10:01,HLA-DPA1*02:01/DPB1*01:01<br>HLA-DQA1*01:02/DQB1*05:01,HLA-DRB1*01:01<br>HLA-DPA1*01:03/DPB1*04:01<br>HLA-DQA1*02:01/DQB1*03:01,HLA-DRB1*04:02<br>HLA-DRB1*07:01,HLA-DRB1*13:02,HLA-DRB1*12:01,HLA-DPA1*02:01/DPB1*05:01<br>HLA-DQA1*03:01/DQB1*03:01<br>HLA-DQA1*03:03/DQB1*04:02,HLA-DRB5*01:01<br>HLA-DRB1*08:02,HLA-DQA1*01:02/DQB1*05:02<br>HLA-DQA1*06:01/DQB1*04:02,HLA-DRB1*08:01<br>HLA-DRB3*02:02,HLA-DQA1*05:01/DQB1*03:01<br>HLA-DQA1*05:01/DQB1*03:03<br>HLA-DRB1*13:01,HLA-DRB1*07:01<br>HLA-DRB4*01:03 |
| C4                 | DRB1_0309,DRB1_0405,DRB1_0421<br>DRB1_0703,DRB1_1502,DRB1_1506 | HLA-DR4,HLA-DR9,HLA-DR2,HLA-DR3<br>HLA-DR8,HLA-DR13,HLA-DR15<br>HLA-DR52,HLA-DQ4,HLA-DQ6<br>HLA-DQ7,HLA-DQ8,HLA-DQB1*0301<br>HLA-DRB4*0101,HLA-DRB5*0101<br>HLA-DRB1*0301,HLA-DRB1*0401<br>HLA-DRB1*0405,HLA-DRB1*0802<br>HLA-DRB1*0901,HLA-DRB1*1501,<br>HLA-DRB5*0101 |   |
| C5                 | Not applicable   | HLA-DR4,HLA-DR9,HLA-DRB1*0301   |   |

| Conserved Sequence | ProPred   | MHC2Pred   | IEDB   |
|--------------------|---|--|--|
| C6                 | Not applicable  | HLA-DRB1*0405,HLA-DRB1*0901<br>HLA-DR9,HLA-DQ4,HLA-DQ7,HLA-DQ8,<br>HLA-DQB1*0301,HLA-DRB1*0401<br>HLA-DRB1*0405,HLA-DRB1*0901<br>HLA-DRB1*1501   | HLA-DRB4*01:03,HLA-DQA1*01:02/DQB1*05:01<br>HLA-DQA1*05:01/DQB1*03:02<br>HLA-DPA1*01:03/DPB1*06:01<br><br>HLA-DRB3*03:01,HLA-DRB1*13:01<br>HLA-DRB1*01:01,HLA-DPA1*03:01/DPB1*04:02<br>HLA-DQA1*01:02/DQB1*05:01,HLA-DRB1*04:04<br>HLA-DRB4*01:01,HLA-DRB1*13:02<br>HLA-DQA1*05:01/DQB1*03:03<br>HLA-DQA1*02:01/DQB1*03:01<br>HLA-DRB4*01:03,HLA-DQA1*01:02/DQB1*05:02<br>HLA-DPA1*01:03/DPB1*06:01,HLA-DRB1*07:01<br>HLA-DQA1*03:01/DQB1*03:01<br>HLA-DRB1*10:01,HLA-DRB3*02:02<br>HLA-DQA1*01:02/DQB1*06:02<br>HLA-DPA1*02:01/DPB1*01:01,HLA-DRB1*04:05<br>HLA-DQA1*02:01/DQB1*03:03<br>HLA-DQA1*01:01/DQB1*05:01,HLA-DRB1*16:02<br>HLA-DRB1*12:01,HLA-DQA1*05:01/DQB1*03:01<br>HLA-DRB1*15:01,HLA-DQA1*05:01/DQB1*03:02<br>HLA-DRB1*09:01,HLA-DQA1*02:01/DQB1*04:02<br>HLA-DPA1*01:03/DPB1*02:01,HLA-DRB1*04:02<br>HLA-DRB1*04:01,HLA-DQA1*05:01/DQB1*04:02<br>HLA-DRB1*03:01,HLA-DQA1*05:01/DQB1*02:01 |
| C7                 | DRB1_0102,DRB1_0306,DRB1_0307<br>DRB1_0308,DRB1_0311,DRB1_0401<br>DRB1_0402,DRB1_0405,DRB1_0410<br>DRB1_0421,DRB1_0426,DRB1_1107<br>DRB1_1304,DRB1_1506 | HLA-DR4,HLA-DR9,HLA-DR2<br>HLA-DR7,HLA-DR13,HLA-DR15<br>HLA-DR52,HLA-DR53,HLA-DQ4<br>HLA-DQ6,HLA-DQ7,HLA-DQ8<br>HLA-DQB1*03,HLA-DQB1*0301<br>HLA-DRB1*0101,HLA-DRB4*0101<br>HLA-DRB1*0301,HLA-DRB1*0401<br>HLA-DRB1*0405,HLA-DRB1*0802<br>HLA-DRB1*0901,HLA-DRB1*1101<br>HLA-DRB1*1501 | HLA-DRB3*03:01,HLA-DRB1*13:01<br>HLA-DRB1*01:01,HLA-DPA1*03:01/DPB1*04:02<br>HLA-DQA1*01:02/DQB1*05:01,HLA-DRB1*04:04<br>HLA-DRB4*01:01,HLA-DRB1*13:02<br>HLA-DQA1*05:01/DQB1*03:03<br>HLA-DQA1*02:01/DQB1*03:01<br>HLA-DRB4*01:03,HLA-DQA1*01:02/DQB1*05:02<br>HLA-DPA1*01:03/DPB1*06:01,HLA-DRB1*07:01<br>HLA-DQA1*03:01/DQB1*03:01<br>HLA-DRB1*10:01,HLA-DRB3*02:02<br>HLA-DQA1*01:02/DQB1*06:02<br>HLA-DPA1*02:01/DPB1*01:01,HLA-DRB1*04:05<br>HLA-DQA1*02:01/DQB1*03:03<br>HLA-DQA1*01:01/DQB1*05:01,HLA-DRB1*16:02<br>HLA-DRB1*12:01,HLA-DQA1*05:01/DQB1*03:01<br>HLA-DRB1*15:01,HLA-DQA1*05:01/DQB1*03:02<br>HLA-DRB1*09:01,HLA-DQA1*02:01/DQB1*04:02<br>HLA-DPA1*01:03/DPB1*02:01,HLA-DRB1*04:02<br>HLA-DRB1*04:01,HLA-DQA1*05:01/DQB1*04:02<br>HLA-DRB1*03:01,HLA-DQA1*05:01/DQB1*02:01   |

## Discussion

There is identification of the requirement to stimulate each arm of the adaptive reaction for an efficient preventative HCV immunogen and proof to support the inclusion of each structural and non-structural protein.<sup>16</sup> Previous studies have shown proof of cross-reactive neutralizing antibodies (NAbs), significantly within the great ape model.<sup>17-19</sup> However, they have restricted knowledge on immunogen candidates that elicit neutralizing and HCV-specific T-lymphocyte responses.

Some positive results have been shown in a research on African monkeys and immunized mice by mixing the glycoproteins E1, E2, non-structural protein 3 (NS3), and the core protein in the alum (consisting of conserved T-cell epitopes) for the induction of sensitive neutralizing antibodies and wide cellular responses (CD4+ and CD8+).<sup>20</sup> Few vectors and DNA-based vaccines were also present, but results from the phase 2 trial showed that they increased productivity because they were likely to be in an "at-risk" population.<sup>14</sup>

In the relatively large and currently thirty-year field of HIV immunizing agent development, the dearth of an efficient immunizing agent points to the various remaining barriers to inducement, generally neutralizing antibodies or effective CD8+ T cells capable of acting and continuous at the positioning of tissue layer HIV entry. Several current immunizing agent methods going into clinical studies look to handle a number of the evasion mechanisms mentioned here. For instance, there has been testing of assorted diversity-combating antigen-style approaches as well as mosaic vaccines, during which the inclusion of variant epitopes is optimized, as well as ways to preserve immunogen sequences. There are varied adjuvants, vectors, and delivery vehicles designed to boost the potency of the presentation of immunizing agent antigens so

as to stimulate effective antiviral responses. A unique "tolerogenic" immunizing agent consisting of inactivated simian immunological disorder virus (SIV) mac239 particles with explicit microorganism adjuvants has been shown to elicit CD8+ T-regulatory cells in immunized macaques.

T lymphocytes were not lysed but were ready to suppress the activation of SIV-positive CD4+ T cells, making them less prone to SIV infection once challenged. Additionally, these CD8+ T cells were found to have unusual resistance provided by non-classical MHCIIb/E molecules<sup>20</sup>, such as HLA-E in humans. Curiously, recent knowledge shows HLA-E expression in liver biopsies correlates with HCV microorganism load in chronic HCV-infected subjects, and NK cells lacking the repressive receptor for HLA-E (NKG2A) are related to protection from HCV infection in bad-exposure subjects.<sup>21</sup> Thus far, there has been no examination of non-classical HLA-restricted CD8+ T cells in HCV infection. In distinction to the CD8+ "T-regulatory" sort cells represented above, an immunizing agent supported by a rhesus herpes virus vector has made sturdy protection or clearance of SIV challenge infections in unsusceptible macaques related to the induction of effector memory CD8+ T lymphocyte responses. However, these CD8+ T cells are found to focus on a variety of promiscuous or dominant epitopes restricted by HLA category II alleles instead of HLA category I.<sup>22-23</sup>

## Conclusion

In this study, computational identification of multiple T cell epitopes was done for genotype 6 of the human genome with the help of various epitope prediction tools. The tools used for MHC class I epitope prediction were SYFPEITHI, NetCTL1.2, and IEDB for CD8+ T cells. Similarly, for MHC class II epitope prediction, the tools were ProPred, MHC2Pred, and IEDB for CD4+ T cells.

These tools give unique epitope sequences. However, various unique HLA allele sequences were enlisted, for which the epitopes were predicted. This study gives the platform to identify future Hepatitis C vaccine candidates for genotype 6 for further in silico and experimental validation in vivo and in vitro cultures.

### **Conflict of interest**

The authors whose names are listed in the paper have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria, educational grants, participation in speakers' bureaus, membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements) or non-financial interest (such as personal or professional relationships, affiliations, knowledge, or beliefs) in the paper's subject matter or accoutrements.

### **Data availability statement**

The data can be made available upon request from the author.

### **Ethics statement**

Not applicable.

### **Acknowledgement**

Not applicable.

### **References**

1. Choo QL, Richman KH, Han JH, et al. Genetic organization and diversity of the hepatitis C virus. *Proc Natl Acad Sci U S A*. 1991;88(6):2451-2455. doi:10.1073/pnas.88.6.2451
2. Salari N, Kazemini M, Hemati N, Ammari-Allahyari M, Mohammadi M, Shohaimi S. Global prevalence of hepatitis C in general population: A systematic review and meta-analysis. *Travel Med Infect Dis*. 2022;46:102255. doi:10.1016/j.tmaid.2022.102255
3. Nakamura M, Kanda T, Haga Y, et al. Sofosbuvir treatment and hepatitis C virus infection. *World J Hepatol*. 2016;8(3):183-190. doi:10.4254/wjh.v8.i3.183
4. Rong L, Perelson AS. Treatment of hepatitis C virus infection with interferon and small molecule direct antivirals: viral kinetics and modeling. *Crit Rev Immunol*. 2010;30(2):131-148. doi:10.1615/critrevimmunol.v30.i2.30
5. Fournillier A, Wychowski C, Boucreux D, et al. Induction of hepatitis C virus E1 envelope protein-specific immune response can be enhanced by mutation of N-glycosylation sites. *J Virol*. 2001;75(24):12088-12097. doi:10.1128/JVI.75.24.12088-12097.2001
6. Lafuente EM, Reche PA. Prediction of MHC-peptide binding: a systematic and comprehensive overview. *Curr Pharm Des*. 2009;15(28):3209-3220. doi:10.2174/138161209789105162
7. Demetz S, Grey HM, Sette A. The minimal number of class II MHC-antigen complexes needed for T cell activation. *Science*. 1990;249(4972):1028-30. doi:10.1126/science.2118680
8. Hammer J, Bono E, Gallazzi F, Belunis C, Nagy Z, Sinigaglia F. Precise prediction of major histocompatibility complex class II-peptide interaction based on peptide side chain scanning. *J Exp Med*. 1994;180(6):2353-2358. doi:10.1084/jem.180.6.2353
9. Singh H, Raghava GP. ProPred: prediction of HLA-DR binding sites.

- Bioinformatics*. 2001;17(12):1236-7.  
doi:  
10.1093/bioinformatics/17.12.1236
10. Pickett BE, Greer DS, Zhang Y, Stewart L, Zhou L, Sun G, Gu Z, Kumar S, Zaremba S, Larsen CN, et al. Virus Pathogen Database and Analysis Resource (ViPR): A Comprehensive Bioinformatics Database and Analysis Resource for the Coronavirus Research Community. *Viruses*. 2012; 4(11):3209-3226.  
<https://doi.org/10.3390/v4113209>
  11. Sievers F, Higgins DG. Clustal omega. *Curr Protoc Bioinformatics*. 2014;48:3.13.1-3.13.16. Published 2014 Dec 12.  
doi:10.1002/0471250953.bi0313s48
  12. Khan AM, Miotto O, Heiny AT, et al. A systematic bioinformatics approach for selection of epitope-based vaccine targets. *Cell Immunol*. 2006;244(2):141-147. doi: 10.1016/j.cellimm.2007.02.005
  13. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanović S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics*. 1999;50(3-4):213-219.  
doi:10.1007/s002510050595
  14. Larsen MV, Lundegaard C, Lamberth K, et al. An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur J Immunol*. 2005;35(8):2295-2303.  
doi:10.1002/eji.200425811
  15. Fleri W, Paul S, Dhanda SK, et al. The Immune Epitope Database and Analysis Resource in Epitope Discovery and Synthetic Vaccine Design. *Front Immunol*. 2017;8:278. Published 2017 Mar 14.  
doi:10.3389/fimmu.2017.00278
  16. Swadling L, Klenerman P, Barnes E. Ever closer to a prophylactic vaccine for HCV. *Expert Opin Biol Ther*. 2013;13(8):1109-1124.  
doi:10.1517/14712598.2013.791277
  17. Torresi J, Johnson D, Wedemeyer H. Progress in the development of preventive and therapeutic vaccines for hepatitis C virus. *J Hepatol*. 2011;54(6):1273-85.
  18. Fornis X, Payette PJ, Ma X, Satterfield W, Eder G, Mushahwar IK, Govindarajan S, Davis HL, Emerson SU, Purcell RH, Bukh J. Vaccination of chimpanzees with plasmid DNA encoding the hepatitis C virus (HCV) envelope E2 protein modified the infection after challenge with homologous monoclonal HCV. *Hepatology*. 2000;32(3):618-25.
  19. Rollier CS, Paranhos-Baccala G, Verschoor EJ, Verstrepen BE, Drexhage JA, Fagrouch Z, Berland JL, Komurian-Pradel F, Duverger B, Himoudi N, Staib C. Vaccine-induced early control of hepatitis C virus infection in chimpanzees fails to impact on hepatic PD-1 and chronicity. *Hepatology*. 2007;45(3):602-13.
  20. Martinez-Donato G, Amador-Cañizares Y, Alvarez-Lajonchere L, et al. Neutralizing antibodies and broad, functional T cell immune response following immunization with hepatitis C virus proteins-based vaccine formulation. *Vaccine*. 2014;32(15):1720-1726.  
doi:10.1016/j.vaccine.2014.01.046
  21. Andrieu JM, Chen S, Lai C, Guo W, Lu W. Mucosal SIV Vaccines Comprising Inactivated Virus Particles and Bacterial Adjuvants

- Induce CD8(+) T-Regulatory Cells that Suppress SIV-Positive CD4(+) T-Cell Activation and Prevent SIV Infection in the Macaque Model. *Front Immunol.* 2014;5:297. Published 2014 Jun 30. doi:10.3389/fimmu.2014.00297
22. Thoens C, Berger C, Trippler M, et al. KIR2DL3<sup>+</sup>NKG2A<sup>-</sup> natural killer cells are associated with protection from productive hepatitis C virus infection in people who inject drugs. *J Hepatol.* 2014;61(3):475-481. doi:10.1016/j.jhep.2014.04.020
23. Hansen SG, Piatak M Jr, Ventura AB, et al. Immune clearance of highly pathogenic SIV infection [published correction appears in *Nature.* 2014 Oct 30;514(7524):654] [published correction appears in *Nature.* 2017 Jul 6;547(7661):123-124]. *Nature.* 2013;502(7469):100-104. doi:10.1038/nature12519