Journal of Intellectuals (JoI)

Original Article

Peptide vaccine prediction against Hepatitis-C virus causing hepatocellular cancer using immunoinformatics approach

Afrin Yasmin[*1]

[1] Department of Biotechnology, Assam University, Silchar 788011, Assam, India
[*]Corresponding author email: yasafrinl@gmail.com

Abstract: The prime objective of this study is to predict peptide vaccine against all genotypes of HCV that might be a part of combination therapy. We used bioinformatic approaches to identify the T cell and B cell epitopes having high antigenicity from the envelope glycoproteins E1 and E2 of HCV. For the analysis, HLA-A1, HLA-A*01:01, HLA-A3, HLA-B*2702, HLA-B*2705, HLA-B7, HLA-B8, HLA-Cw*0702, HLA-DRB1*1101, HLA-DRB1*03:01, DRB1*0401 and DRB1*0311 alleles were selected for the prediction of T cell epitopes. Consequently, potential B cell epitopes for each surface protein were predicted *i.e.,* ALYVGMCGA, VNYRNVSGIY, GAAFCSALYV and CGVVSAKTVC. Likewise, T cell peptide epitopes having potential to bind with MHC class I molecule were also identified *i.e.,* AWAKVVVIL, VRYVGATTA, VAPTLAVRY, WEYVVLAFL, WEYIMLVFL, AVKWEYVVL and WEYVVLLFL. Further, T cell epitope capable of binding to MHC class II molecule was identified *i.e.,* VAIIMVMFS, LVLAQVMRI, VVIDIIAGG, LVGSATLCS, VVASATLCS, LLADARVCA, IQLINTNGS, LQLINTNGS and VVLLFLLLA. These peptides are potential candidates for design of vaccine.

Keywords: Hepatitis C virus; Peptide Epitopes; Vaccine prediction; Hepatocellular cancer

Citation: Yasmin, A. (2021). Peptide vaccine prediction against Hepatitis-C virus causing hepatocellular cancer using immunoinformatics approach. Journal of Intellectuals, 1(1), 82–97. Retrieved from https://journals.bahonacollege.edu.in/index.php/joi/article/view/joi2021-1-1-8

Received: 15 July, 2021
Revised: 12 September, 2021
Accepted: 12 December, 2021
Published: 25 December, 2021

Publisher's Note: JOI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

HCV is a global blood-borne virus that caused infection to an estimated of 1.3 million people worldwide [1]. The genus was first identified in 1988 [2-3]. It is a small circular, enveloped, positive sense and single-stranded RNA virus and belongs to the hepacivirus genus within the Flaviviridae family with a diameter of approximately 50-60 nm [4]. Individuals suffering from chronic HCV infection usually remains asymptomatic for a long period and certainly undiagnosed and this leads to severe fibrosis, cirrhosis [5], hepatocellular carcinoma [6], requiring liver transplantation in many patients [7] **Figure 1**. Several studies revealed that the developing countries are at more risk for viral diseases like hepatitis C virus [8]. The entry of HCV into the host cell is a complex mechanism. HCV has 1-7 distinct genotypes with 30% genetic variability [9]. The polyprotein translated from viral RNA genome is cleaved into three structural and seven

nonstructural proteins [10]. The viral nucleocapsid is surrounded by the envelope glycoproteins *i.e.* E1 and E2 which are covalently linked complexes of transmembrane proteins [11], and the surface of these proteins is highly glycosylated [10]. Additionally, one of the major contents of E1 protein is a C-terminal transmembrane domain (TMD) and it has 4-5 glycosylation sites [12], whereas E2 protein contains four O-linked glycosylation sites [13]. Apart from their glycosylation dissimilarities, both function as a heterodimer to mediate the viral passage [14]. Further, the viral entry into the hepatic cell is controlled by several intracellular signaling steps [15]. It requires four essential receptors and co-receptors *i.e.* CD81, scavenger receptor class B type I (SR-BI), and junction proteins named occludin and claudin-1 [16]. The virus particles are linked with certain lipoproteins such as apolipoprotein E (ApoE) to form lipoviral particles (15). SR-BI receptor first interacts with ApoE on the lipoviral molecule of the virus [17], the virus is then mediated into the cell by the action of clathrin-mediated endocytosis process that finally leads to the discharge of viral genome into the host cell [18].

Infection by HCV can be detected within a period of 1–3 weeks, but HCV-specific T-cells or antibodies are recognizable only after 1–2 months [19], however, this leads to the chronicity [20]. In that case, chronic infection and re-infection of the virus can be controlled in the acute phase by the role of neutralizing antibodies (nAbs) [20], but the delayed response of these nAbs may be risky [21]. An HCV vaccine will be successful if it can stimulate both cellular as well as humoral immune responses. It has been observed that a strong cellular response is essential for viral clearance and defense [22]. However, for inducing the humoral (B-cell) immune response against viral infections, prophylactic vaccines are preferred, while for activating both humoral and cellular (T-cell) immune responses the therapeutic vaccines are favored [23].

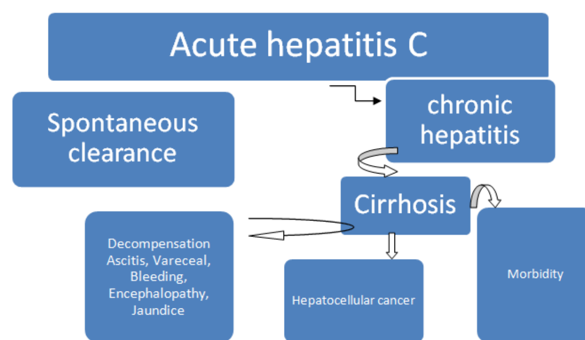

Figure 1. Occurrence of HCV infection

The struggle for the development of vaccines against hepatitis C virus continues since the traditional way of vaccine design and testing the efficacy of the vaccine is difficult and remains a matter of great concern. Notably, HCV is characterized by very high genetic variability [24]. Reverse vaccinology has opened up a new dimension in the field of vaccine prediction using bioinformatic tools [25]. The basic idea behind this technique is to use the complete genomic sequence of an organism to predict potential antigens of encoded proteins using algorithms. For any synthetic vaccine immune dominant epitopes must be there for triggering immune response against the pathogen [26] so that B cells and T cells can recognize the epitopes of an antigen. Using reverse vaccinology approach, the first pathogen against which synthetic vaccine developed was for serogroup B *Neisseria meningitides* [27]. With the advent of reverse vaccinology approach, it is now possible to predict epitopes of high efficacy without culturing any extremely infectious pathogen in the laboratory. The prime requirement of this technique is the availability of whole genome sequence of an infectious

organism. Amino acid residues form the building blocks of a peptide vaccine. Multiple epitope loaded peptide vaccine is now effectively used for the treatment of various diseases, since several clinical trials have been carried out. The first peptide-based vaccine clinical trial on melanoma antigen was initiated in 1995 [28]. For instance, if there is no vaccine yet available against a pathogen or difficult to develop on the basis of its macromolecule, the first attempt made in this regard is to develop a peptide-based multiple-epitope loaded vaccine using the modern tools of immunoinformatics. Traditional way of vaccination using the live-attenuated organism can be risky and sometimes unsafe as the virus may turn into an active form by mutation and might produce several side effects to the host cell [29]. But, vaccines containing purified or recombinant macromolecules of the pathogen seem to have lesser side effects (28) and hence less risky in contrast to the traditional ones [30].

The advantages of peptide-based vaccine include easy production at low cost and does not need any cold preservation, both T cell and B cell responses can directly be stimulated and personalized immune response can be triggered. In this study, we identified the T cell and B cell epitopes having high antigenicity for Major Histocompatibility Complex (MHC) I and II molecules from the evolutionarily conserved regions of envelope glycoproteins E1 and E2. The peptide-based vaccines formulated on the basis of reverse vaccinology do not run the risk of back-mutation into an active form of virus [31]. Several studies reported the importance of immunoinformatics for identifying T cell epitopes, B cell epitopes, and estimating the antigenicity and other biochemical properties of epitopes [32], but this work would provide insights into peptide vaccine formulation against HCV.

2. Methods

Data retrieval and amino acid sequence alignment

From the protein database of NCBI, amino acid sequences of the envelope glycoproteins E1 and E2 of different genotypes of HCV with their accession IDs were retrieved (http://www.ncbi.nlm.nih.gov/). Conserved regions in E1 and E2 protein sequences of HCV were identified using the CLUSTALW program [33]. The capsid protein E1 and E2 were selected, as they are conserved sequences and exists on the outer membrane of virion. Therefore, the epitopes derived from this region could potentially trigger the immune responses against HCV infection specifically.

Prediction of B cell and T cell epitopes of E1 and E2 proteins of HCV

B cell epitope

Predictions of B cell epitopes (linear octapeptides) of E1 and E2 protein sequences of HCV were done using ABCpred analytical tool. All potential B cell continuous linear epitopes were identified using artificial neural network method [34].

T cell epitope

Potential T cell epitopes ((linear hexapeptides) of E1 and E2 glycoprotein sequences of HCV were determined by ProPred-1 [35] and ProPred prediction tool [36]. The peptides having potential to bind with MHC class I molecule were

identified using matrix based method, and the epitopes having affinity for binding to MHC class II molecule were determined using quantitative matrices obtained from published work [37].

Immunogenicity and Antigenicity Score

Immunogenicity of potential T cell epitope capable of binding with MHC Class I molecule was predicted on the basis of immunogenicity score using IEDB Immunogenicity prediction tool as per Calis method [38] similarly, immunogenicity score of T cell epitopes capable of binding with MHC Class II molecule was predicted according to Dikhit et al approach [39]. The immunogenicity is calculated on the basis of amino acid properties and their positions within the peptide. In addition to this, antigenicity of predicted epitope was calculated. Antigenicity is a property that measures the antigenic propensity of a peptide sequence [40]. The antigenicity score of each predicted peptide epitope was determined using the online tool VaxiJen v2.0 [41-43]. However, **Beta-turn prediction on the basis of Chou and Fasman scale** [44] of secondary protein structure was identified to determine the immunogenicity of the epitopes as the beta-turn region of proteins are directly responsible for providing immunogenicity (Rini1992 SC)**.**

In addition to this, the 3D structures of E1 and E2 protein of HCV were retrieved from the Protein Data Bank (PDB) available at http:www.pdb.org.

Physiological properties of predicted B-cell and T-cell Epitopes

The identified epitopes were used for further *in-silico* analysis for its physiochemical characterizations. Parameters such as hydrophilicity, GRAVY score, polarity etc were taken into the account. Hydrophilicity was determined using Parker hydrophilicity scale, which is based on peptide retention time in high-performance liquid chromatography (HPLC) on a reversed-phase column [45]. GRAVY score, atom counts for each peptide and instability index (the level of protein stability) were determined using ProtParam software [46]. Further, aliphatic index, polar as well as non-polar amino acids and aromaticity of those predicted epitopes were determined using EMBOSS Pepstats web tool [47] while the n*et charge* at neutral pH (7.0) was estimated using Protein calculator v3.4 [48]. **Further, t**he surface accessibility of each predicted epitope was carried out using the web tool Emini surface accessibility scale [49]. Additionally, flexibility for each epitope was then estimated using Karplus and Schulze scale [50].

For detailed analysis of isoelectric point and molecular weight of each predicted epitope, the ProtParam ExPasy web tool was used, which allows computation of the theoretical pI and Mw [46, 51-52].

3. Results

Alignment of amino acid sequences of capsid protein E1and E2 of Hepatitis C-Virus

Generally, antigenic epitopes on the surface of capsid protein have a special chemical structure that triggers the immune system. This peptide based epitope prediction is based on the alignment of conserved sequences. These conserved regions can be considered to be the major targets for the development of neutralizing antibodies. Further, the conserved epitopic regions are inversely proportional to the antigen escape and mutation, so these regions were

identified using bioinformatic tools [53]. Hepatitis C-virus genome has high mutation rate which results in increased genotypic diversity. The E1 and E2 capsid proteins comprise of a consistent pattern of conserved and variable regions, so these two proteins were used for prediction and analysis of the strongly antigenic epitopes contained in them. Further, E1 and E2 proteins harbour the target binding region of neutralizing antibodies. We had aligned seven E1 and six E2 proteins from different HCV genotypes obtained from the National Center for Biotechnology Information database (http://ncbi.nlm.nih.gov) using CLUSTALW program to find out the conserved regions within the protein sequences **Figure 2.** It is believed that the highly conserved epitopic regions are less prone to antigen escape and viral mutation [54]. From the alignment figures, it was observed that the alignment score for E1 was 17568 and for E2 the score was 27917. It revealed that the intertype variability of E2 proteins was higher than that of E1 proteins. From these results, we could conclude that E2 proteins showed less consistency in the positions of variable regions as compared to the E1 proteins.

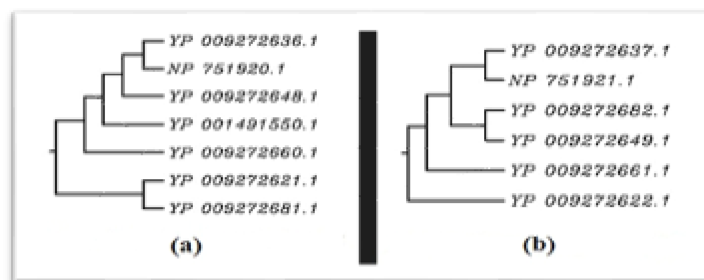

Figure 2a & 2b. Phylogenetic analysis of E1 and E2 proteins from different genotypes of HCV, respectively.

Identification of B-cell and T-cell Epitopes and their Characterization

After a detailed analysis newly, efficient B-cell peptide epitopes were predicted using immunoinformatics analytical tool **Table 1.** From a comprehensive analysis of antigenicity, accessibility, flexibility, hydrophilicity, gravy score and Chou-Fasman conformation in the present study, the selection of promising continuous linear B-cell epitopes of each surface protein were determined based on maximum hydrophilicity score of the epitopes using ABCpred software that relies on artificial neural network. Potential B cell epitopes for E1 and E2 protein sequences of different HCV genotypes were found within the conserved locales and the start position of each epitope in the protein sequence, along with their pI, polarity, atom count, aliphatic and aromatic index were determined as shown in **Table 2**. The T cell epitopes capable of binding to MHC class I molecules were identified using the ProPred-1 software. Constant dysregulation of HLA expression was found to be associated with increased HCV chronic infection [55], which could influence cell-mediated vaccine therapies. Therefore, HLAA*01:01 allele along with HLA-A1, HLA-A3, HLA-B*2702, HLA-B*2705, HLA-B7, HLA-B8, HLA-Cw*0702 [56] were preferred for predicting T cell MHC-I epitopes of E1 and E2 proteins of HCV using ProPred I analysis tool at 4% default threshold value **Table 3**. In addition, the predictions for T cell MHC-II epitopes of E1 and E2 proteins were done using ProPred at a 3% threshold value. The DRB1*1101, DRB1*0201, DRB1*0311 and DRB1*0401 alleles were found to be associated with chronic HCV infection [55-56] **Table 4**.

Table 1: B cell peptide epitopes against E1 and E2 proteins with their accession numbers and antigenicity scores of HCV

Envelope protein	HCV genotypes	Accession numbers	Peptide Sequence	Score	Start Position	Antigenicity Score
E1	1	YP_009272621	NDCTNDSITW	0.81	14	0.02
	2	YP_001491550	ALYVGMCGA	0.79	83	0.17
	3	YP_009272636	VNYRNVSGIY	0.79	1	1.21
	4	YP_009272648	GAAFCSALYV	0.71	77	0.17
	5	YP_009272660	CVPCVRAGNI	0.77	35	2.54
	6	NP_751920	VFLVGQLFTF	0.79	93	-0.21
	7	YP_009272681	ILHEPGCVPC	0.83	29	-0.34
E2	1	YP_009272622	CGVVSAKTVC	0.83	113	1.38
	2	YP_009272637	FLASLFYTHK	0.86	54	0.24
	3	YP_009272649	CNDSLQTGFI	0.81	46	0.33
	4	YP_009272661	GWFGCTWMNS	0.83	166	-0.01
	5	YP_009272682	PPQGSWFGCS	0.87	162	-0.33
	6	NP_751921	FTPSPVVVGT	0.86	126	1.46

Table 2: Chemical properties of predicted B cell epitopes of HCV

Peptide epitopes of B cell	Chou-Fasman conformat-ion	Hydrophilicity	Accessibility	Flexibility	Instability	pI	pH	Mw	Gravy score	Aliphatic index	Aroma-ticity index	Polar %	Non-polar %
ALYVGMCGA	0.8	-1.4	1.5	0.9	-9.98 (stable)	5.56	-0.1	88.08	1.54	44.4	11.1	0	100
VNYRNVSGIY	1.0	2.2	1.0	0.9	-4.06 (stable)	8.56	0.9	1184.32	-0.24	30.0	20.0	40	60
GAAFCSALYV	0.9	1.5	0.9	0.9	0.51 (stable)	5.52	0.1	1001.72	1.62	50.0	20.0	10	90
CVPCVRAGNI	0.9	0.5	0.4	0.9	41.09 (unstable)	8.07	0.8	1031.26	0.97	40.0	0	20	80
CGVVSAKTVC	0.9	2.0	0.1	0.9	9.0 (stable)	8.03	0.8	966.18	1.36	40.0	0	30	70
FLASLFYTHK	0.8	-4.3	0.3	0.9	56.9 (unstable)	8.88	1.1	1226.44	0.51	30.0	40.0	40	60
CNDSLQTGFI	1.1	3.8	0.7	1.0	52.37 (unstable)	3.81	-1.1	1097.21	0.12	20.0	10.0	50	50
FTPSPVVVGT	1.0	-0.1	1.7	1.0	74.48 (unstable)	5.50	-0.1	1003.16	0.96	30.0	10.0	30	70

Table 3: T cell peptide epitopes E1 and E2 proteins that can bind with MHC class I molecule with their accession numbers and antigenicity scores of HCV

Envelope protein	HCV genotypes	Accession numbers	MHC class	Allele	Peptides	Length	Residue no.	Percentage score	Antigenicity Score
E1	1	YP_009272621	I	HLA-Cw*0702	AWAKVVVIL	9	176	29.56	0.32
	2	YP_001491550	I	HLA-B*2705	VRYVGATTA	9	57	67.01	0.46
	3	YP_009272636	I	HLA-B*2705	NQSRCWVAL	9	43	51.40	0.56
	4	YP_009272648	I	HLA-B*7	APGCVPCVL	9	32	50.30	-0.02
	5	YP_009272660	I	HLA-Cw*0702	VPEIVLEVF	9	149	46.83	-0.48
	6	NP_751920	I	HLA-B*2705	MNWSPTAAL	9	133	46.67	1.68
	7	YP_009272681	I	HLA-Cw*0702	VAPTLAVRY	9	51	54.05	1.50
E2	1	YP_009272622	I	HLA-B*2705	VRWEWVILL	9	335	89.34	2.12
	2	YP_009272637	I	HLA-B*2705	WEYVVLAFL	9	333	48.60	1.16
	3	YP_009272649	I	HLA-B*2705	WEYIMLVFL	9	334	48.60	1.14
	4	YP_009272661	I	HLA-B*7	AVKWEYVVL	9	335	37.57	1.59
	5	YP_009272682	I	HLA-B*2705	IRWEWVVLV	9	335	77.66	.2.38
	6	NP_751921	I	HLA-B*2705	WEYVVLLFL	9	333	48.60	0.84

Table 4: T cell peptide epitopes E1 and E2 proteins that can bind with MHC class II molecule with their accession numbers and antigenicity scores of HCV

Capsid protein	HCV genotypes	Accession numbers	MHC Class	Allele	Peptide sequence	Length	Position	Allele Score	% of highest score	Threshold for 3% with score	Antigenicity Score
E1	1	YP_009272621	II	HLA-DRB1*0401	VVMSATLCS	9	73	6.6	76.74	1.48	0.28
	2	YP_001491550	II	HLA-DRB1*0301	VAIIMVMFS	9	179	4.6	55.42	1.1	0.74
	3	YP_009272636	II	HLA-DRB1*0301	LVLAQVMRI	9	140	6.0	63.16	2.96	0.37
	4	YP_009272648	II	HLA-DRB1*0301	VVIDIIAGG	9	151	5.6	63.64	2.08	0.58
	5	YP_009272660	II	HLA-DRB1*0301	LVLSSILRV	9	140	5.0	52.63	2.96	0.10
	6	NP_751920	II	HLA-DRB1*0401	LVGSATLCS	9	73	5.4	62.79	1.48	0.13
	7	YP_009272681	II	HLA-DRB1*0401	VVASATLCS	9	73	5.2	60.47	1.48	0.24
E2	1	YP_009272622	II	HLA-DRB1*0311	LLADARVCA	9	345	4.3	49.77	2.08	0.14
	2	YP_009272637	II	HLA-DRB1*0401	LQLINSNGS	9	27	4.3	50.00	1.48	-0.06
	3	YP_009272649	II	HLA-DRB1*0301	LMYAMKFNS	9	57	4.8	50.53	2.96	0.56
	4	YP_009272661	II	HLA-DRB1*0401	IQLINTNGS	9	26	5.2	60.47	1.48	0.33
	5	YP_009272682	II	HLA-DRB1*0401	LQLINTNGS	9	27	5.2	60.46	1.48	0.28
	6	NP_751921	II	HLA-DRB1*1101	VVLLFLLLA	9	335	4.9	59.04	1.1	0.52

3D-Structure of Capsid ProteinE1 and E2

The three-dimensional structures of E1 and E2 proteins of HCV were retrieved from PDB (Protein Databank) having ID 2KNU [57] and 4WEB [16], respectively. E1 glycoprotein showed the molecular weight of 3328.97D, atom count: 229, residue count: 29, unique protein chains: 1, with unique A chain **(Figure 3a)**. Additionally, E2 glycoprotein had molecular weight 102877.65D, atom count: 4201, residue count: 924, unique protein chains: 3, with unique E chain **(Figure 3b)**.

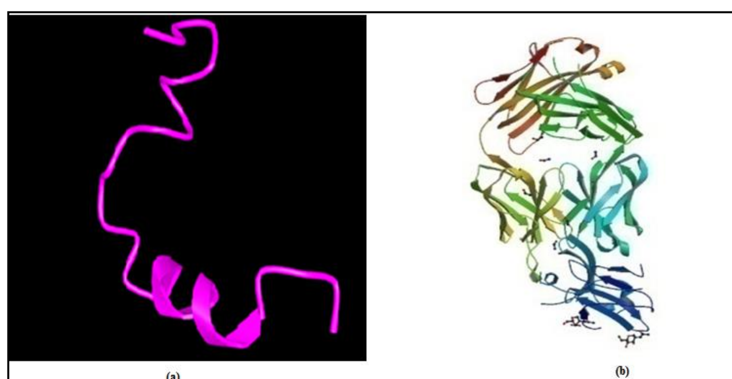

Figure 3. (a). 3D structure of E1 (2KNU). **(b).** 3D structure of E2 (4WEB)

4. Discussion

Since more than one B-cell epitope could be recognized in a protein, so the antigenic propensity of each B-cell epitope was further determined. Sometimes not all epitopes in a protein induce high humoral immunity. In the present study, seven B-cell epitopes *i.e.* NDCTNDSITW, ALYVGMCGA, VNYRNVSGIY, GAAFCSALYV, CVPCVRAGNI, VFLVGQLFTF and ILHEPGCVPC were predicted for E1 protein. Likewise, six B cell epitopes namely CGVVSAKTVC, FLASLFYTHK, CNDSLQTGFI, GWFGCTWMNS, PPQGSWFGCS and FTPSPVVVGT were predicted for E2 protein. The antigenicity of these epitopes was then predicted and we identified the epitopes having antigenic score > 0.5 as highly antigenic epitope **Table 1**. Following analysis, the filtered B cell epitopes of both E1 and E2 proteins were found to be ALYVGMCGA, VNYRNVSGIY, GAAFCSALYV, CVPCVRAGNI, CGVVSAKTVC, FLASLFYTHK, CNDSLQTGFI

and FTPSPVVVGT **Table 2**. These selected epitopes were analyzed for other biochemical properties and consequently, three epitopes namely ALYVGMCGA (antigenic score 0.17), VNYRNVSGIY (antigenic score=1.21), GAAFCSALYV (antigenic score=0.17) for E1 and the sole epitope CGVVSAKTVC (antigenic score=1.38) for E2 were recognized as suitable B cell epitopes based on instability index (stable peptides were selected as B cell epitope). These B-cell epitopes might be associated with the production of neutralizing antibody response against HCV. Based on those characterization B-cell epitopes were represented graphically **Figure 4.**

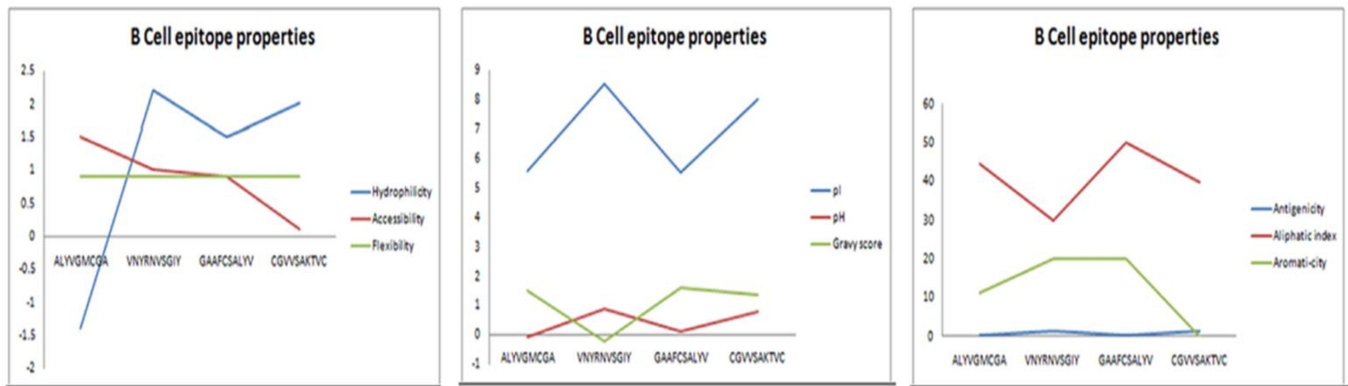

Figure 4a **Figure 4**b **Figure** 4c

Figure 4(a). Graphical representation for B cell epitopes of E1 and E2 protein and their predicted hydrophilicity, accessibilityand flexibility and (b) Graphical representation for B cell epitopes of E1 and E2 protein and their predicted net charge at pH 7 along with Gravy score and isoelectric point. (c) Graphical representation for B cell epitopes of E1 and E2 protein and their predicted antigenicity, aliphatic index and aromaticity.

In addition, potential T cell epitopes that could bind with MHC Class I molecule for E1 and for E2 were predicted within the conserved regions **Table 3**. The predicted peptides from E1 were AWAKVVVIL, VRYVGATTA, NQSRCWVAL, APGCVPCVL, VPEIVLEVF, MNWSPTAAL, VAPTLAVRY and from E2 were VRWEWVILL, WEYVVLAFL, WEYIMLVFL, AVKWEYVVL, IRWEWVVLV and WEYVVLLFL. Of these, five peptides from E1 *i.e.* AWAKVVVIL, VRYVGATTA, NQSRCWVAL, MNWSPTAAL and VAPTLAVRY, and six peptides from E2 *i.e.* VRWEWVILL, WEYVVLAFL, WEYIMLVFL, AVKWEYVVL, IRWEWVVLV and WEYVVLLFL were found to be in the antigenic range. Moreover, peptides were further analyzed for instability index and polarity of amino acids in the peptide **Table 5**.

Depending on these parameters, we could reduce the number of predicted epitopes to seven for E1 and E2 proteins *i.e.* AWAKVVVIL, VRYVGATTA, VAPTLAVRY, WEYVVLAFL, WEYIMLVFL, AVKWEYVVL and WEYVVLLFL. Further, the atom count of each epitope was also determined **Table 6**.

Table 5: Chemical properties of T-cell epitopes binding with both MHC Class I and Class II complex of HCV

Peptide epitopes of T cell binding with MHC Class I molecule ⇩	Instability	pI	pH	Mw	Gravy score	Aliphatic index	Aromaticity index	Polar%	Non-polar %
AWAKVVVIL	**4.16 (stable)**	**8.80**	**0.9**	**998.28**	**2.18**	**77.7**	**11.1**	**11.1**	**88.8**
VRYVGATTA	**-18.36 (stable)**	**8.72**	**0.9**	**937.06**	**0.48**	**44.4**	**11.1**	**33.3**	**66.6**
NQSRCWVAL	**87.61 (unstable)**	**8.25**	**0.9**	**1076.24**	**-0.10**	**33.3**	**11.1**	**44.4**	**55.5**
MNWSPTAAL	**46.19 (unstable)**	**5.28**	**-0.1**	**990.14**	**0.20**	**33.3**	**11.1**	**33.3**	**66.6**
VAPTLAVRY	**21.91 (stable)**	**8.72**	**0.9**	**989.18**	**0.85**	**55.5**	**11.1**	**22.2**	**77.7**
VRWEWVILL	**67.80 (unstable)**	**5.97**	**-0.1**	**1213.49**	**1.18**	**55.5**	**22.2**	**22.2**	**77.7**
WEYVVLAFL	**8.89 (stable)**	**4.0**	**-1.1**	**1139.36**	**1.65**	**55.5**	**33.3**	**11.1**	**88.8**
WEYIMLVFL	**8.89 (stable)**	**4.0**	**-1.1**	**1213.50**	**1.70**	**44.4**	**33.3**	**11.1**	**88.8**
AVKWEYVVL	**5.69 (stable)**	**6.05**	**-0.1**	**1106.33**	**0.95**	**55.5**	**22.2**	**22.2**	**77.7**
IRWEWVVLV	**46.40 (unstable)**	**6.0**	**-0.1**	**1199.46**	**1.23**	**55.5**	**22.2**	**22.2**	**77.7**
WEYVVLLFL	**8.89 (stable)**	**4.0**	**-0.1**	**1181.44**	**1.87**	**55.5**	**33.3**	**11.1**	**88.8**
Peptide epitopes of T cell binding with MHC Class II molecule ⇩									
VVMSATLCS	**57.71 (unstable)**	**5.49**	**-0.1**	**910.11**	**1.78**	**44.4**	**0.0**	**33.3**	**66.6**
VAIIMVMFS	**8.89 (stable)**	**5.49**	**-0.1**	**1010.32**	**2.77**	**55.5**	**11.1**	**11.1**	**88.8**
LVLAQVMRI	**-7.87 (stable)**	**9.75**	**0.9**	**1042.35**	**1.80**	**66.6**	**0.0**	**22.2**	**77.7**
VVIDIIAGG	**22.60 (stable)**	**3.80**	**-1.1**	**856.03**	**2.15**	**66.6**	**0.0**	**11.1**	**88.8**
LVLSSILRV	**73.09 (unstable)**	**9.75**	**0.9**	**999.26**	**2.02**	**66.6**	**0.0**	**33.3**	**66.6**
LVGSATLCS	**-0.54 (stable)**	**5.52**	**-0.1**	**850.0**	**1.48**	**44.4**	**0.0**	**33.3**	**66.6**
VVASATLCS	**8.89 (stable)**	**5.49**	**-0.1**	**850.0**	**1.77**	**55.5**	**0.0**	**33.3**	**66.6**
LLADARVCA	**-0.54 (stable)**	**5.83**	**-0.1**	**931.12**	**1.30**	**66.6**	**0.0**	**22.2**	**77.7**
LMYAMKFNS	**61.51 (unstable)**	**8.59**	**0.9**	**1104.35**	**0.30**	**22.2**	**22.2**	**33.3**	**66.6**
IQLINTNGS	**-33.94 (stable)**	**5.52**	**-0.1**	**959.07**	**0.04**	**33.3**	**0.0**	**55.5**	**44.4**
LQLINTNGS	**2.82 (stable)**	**5.52**	**-0.1**	**959.07**	**-0.03**	**33.3**	**0.0**	**55.5**	**44.4**

VVLLFLLLA	**8.89 (stable)**	**5.49**	**-0.1**	**1000.33**	**8.89**	**88.8**	**11.1**	**0.0**	**100**

Table 6: Atom counts for each predicted B-cell and T-cell epitope of HCV

Atoms ⇨	Carbon	Hydrogen	Nitrogen	Oxygen	Sulphur
B cell peptide epitopes ⇩					
ALYVGMCGA	38	61	9	11	12
VNYRNVSGIY	53	81	15	16	0
GAAFCSALYV	46	68	10	13	1
CGVVSAKTVC	39	71	11	12	2
T cell of peptide epitopes that binds to MHC Class I molecule ⇩					
AWAKVVVIL	50	83	11	10	0
VRYVGATTA	41	68	12	13	0
VAPTLAVRY	46	76	12	2	0
WEYVVLAFL	59	82	10	13	0
WEYIMLVFL	62	88	10	13	1
AVKWEYVVL	55	83	11	13	0
WEYVVLLFL	62	88	10	13	0
T cell of peptide epitopes that binds to MHC Class II molecule ⇩					
VAIIMVMFS	47	79	9	11	2
LVLAQVMRI	47	87	13	11	1
VVIDIIAGG	39	69	9	12	0
LVGSATLCS	35	63	9	13	1
VVASATLCS	35	63	9	13	1
LLADARVCA	39	70	12	12	1
IQLINTNGS	40	70	12	15	0
LQLINTNGS	40	70	12	15	0
VVLLFLLLA	52	89	9	10	0

Moreover, the predicted potential epitopes that would bind with MHC class II complex from E1 protein were identified as VVMSATLCS, AIIMVMFS, LVLAQVMRI, VVIDIIAGG, LVLSSILRV, LVGSATLCS and VVASATLCS and from E2 protein were LLADARVCA, LQLINSNGS, LMYAMKFNS, IQLINTNGS, LQLINTNGS and VVLLFLLLA **Table 4**. Here, all the peptides from E1 *i.e.* VVMSATLCS, VAIIMVMFS, VVIDIIAGG, LVLAQVMRI, LVLSSILRV, LVGSATLCS and VVASATLCS and five peptides from E2 *i.e.* LLADARVCA, LMYAMKFNS, LQLINTNGS, IQLINTNGS and VVLLFLLLA were found to be in the antigenic range. Further biochemical characterizations of these predicted peptides having potential to bind with both MHC I and II complex were performed **Table 5**. From the findings, the epitopes having potential to bind with MHC II complex were identified in E1 and E2 proteins. These epitopes along with atom counts were VAIIMVMFS, LVLAQVMRI, VVIDIIAGG, LVGSATLCS, VVASATLCS, LLADARVCA, IQLINTNGS, LQLINTNGS and VVLLFLLLA **Table 6**.

Furthermore, biochemical characterization was done graphically for both T cell epitopes (HLA class I and HLA class II) for validating and is presented below **Figure 5.** Preferentially, B cell and T cell linear epitopes could be selected on the basis of length, conservancy, antigenicity and other biochemical properties for peptide vaccine formulation.

For determining the antigenicity of each predicted peptide, Vaxijen server was used, since it performs well in both validation and prediction at a threshold of 0.5 antigenic scores for viruses [41-43]. The tool is based on the physiochemical properties of peptides.

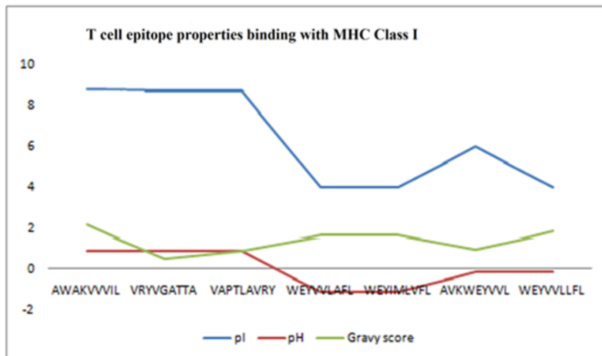

Figure 5(a)

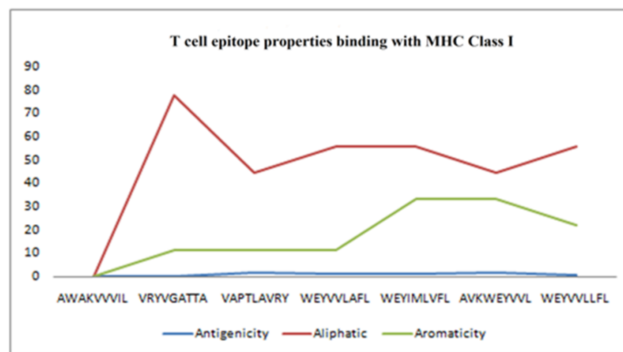

Figure 5(b)

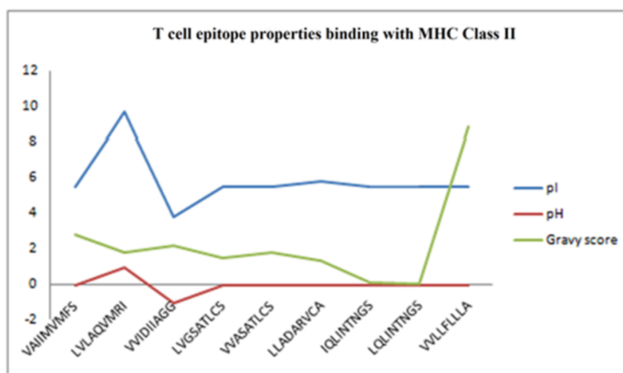

Figure 5(c)

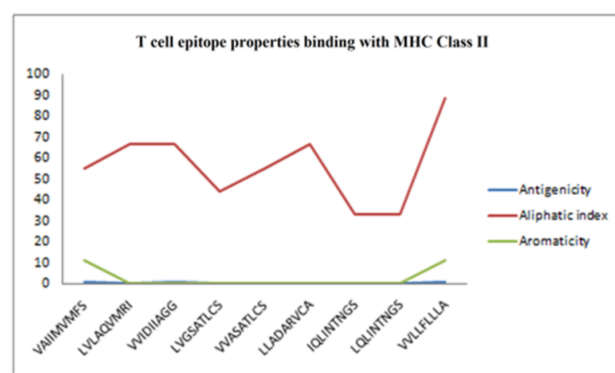

Figure 5(d)

Figure 5(a). Graphical representation for T cell epitopes of E1 and E2 protein binding with MHC Class I molecule with their predicted isoelectric point, net charge at pH 7 and gravy score and (b). Graphical representation for T cell epitopes of E1 and E2 protein binding with MHC Class I molecule and their predicted antigenicity, aliphatic index, and aromaticity.(c). Graphical representation for T cell epitopes of E1 and E2 protein binding with MHC Class II molecule with their predicted isoelectric point, net charge at pH 7 and gravy score and (d). Graphical representation for T cell epitopes of E1 and E2 protein binding with MHC Class I molecule and their predicted antigenicity, aliphatic index, and aromaticity.

5. Limitations

The development of the epitope-based study has advanced the vaccinology approach that leads to analysis of large data sets as well its complete properties. This research report has evaluated all possible efficient epitopes from E1 and E2 and there immunogeneic properties. However, the present work also comes with some caveats, as the epitope sequence is not the only factor dictating the effectiveness of an antibody whereas, the affinity of the antibody for its target is also important. Furthermore, the report also not tells us about the linkers to bind to its predicted peptides sequences.

6. Conclusion

In this work, an attempt was made to predict the epitopes of the E1 and E2 proteins of high-risk Hepatitis C virus for designing a peptide vaccine. The potential T cell epitopes (HLA class I and HLA class II) and B cell epitopes of E1 and E2 proteins were identified using different bioinformatic tools. From the present study, the epitopes predicted as potential B cell epitopes were ALYVGMCGA, VNYRNVSGIY, GAAFCSALYV and CGVVSAKTVC. Similarly, the T cell epitopes that could bind with MHC class I molecule were predicted as AWAKVVVIL, VRYVGATTA, VAPTLAVRY, WEYVVLAFL, WEYIMLVFL, AVKWEYVVL and WEYVVLLFL and likewise, the T cell epitopes binding to MHC class II molecule were predicted as VAIIMVMFS, LVLAQVMRI, VVIDIIAGG, LVGSATLCS, VVASATLCS, LLADARVCA, IQLINTNGS, LQLINTNGS and VVLLFLLLA. The study provides insights for designing an efficient peptide vaccine against HCV.

Acknowledgement

The authors are grateful to Assam University, Silchar, Assam, India for providing the necessary facilities for carrying out this work.

Funding: Unfunded

Conflict of Interest: The authors declare that no conflict of interest exists in this work.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

List of Tables

Table 1: B cell peptide epitopes against E1 and E2 proteins with their accession numbers and antigenicity scores of HCV

Table 2: Chemical properties of B cell epitopes of HCV

Table 3: T cell peptide epitopes E1 and E2 proteins that can bind with MHC class I molecule with their accession numbers and antigenicity scores of HCV

Table 4: T cell peptide epitopes E1 and E2 proteins that can bind with MHC class II molecule with their accession numbers and antigenicity scores of HCV

Table 5: Chemical properties of T-cell epitopes binding with both MHC Class I and Class II complex of HCV

Table 6: Atom counts for each predicted B-cell and T-cell epitope of HCV

List of Figures

Figure 1: Occurrence of HCV infection

Figure 2: Phylogenetic analysis of E1 and E2 proteins from different genotypes of HCV

Figure 3(a): 3D structure of E1 (2KNU)

(b): 3D structure of E2 (4WEB)

Figure 4(a): B cell epitopes of E1 and E2 protein and their predicted hydrophilicity, accessibility and flexibility

(b): B cell epitopes of E1 and E2 protein and their predicted net charge at pH 7 along with Gravy score and isoelectric point.

(c): B cell epitopes of E1 and E2 protein and their predicted antigenicity, aliphatic index and aromaticity

Figure 5(a): T cell epitopes of E1 and E2 protein binding with MHC Class I molecule with their predicted isoelectric point, net charge at pH 7 and GRAVY score

(b): T cell epitopes of E1 and E2 protein binding with MHC Class I molecule and their ppredicted antigenicity, aliphatic index, and aromaticity.

(c): T cell epitopes of E1 and E2 protein binding with MHC Class II molecule with their ppredicted isoelectric point, net charge at pH 7 and GRAVY score

(d): T cell epitopes of E1 and E2 protein binding with MHC Class I molecule and their predicted antigenicity, aliphatic index, and aromaticity

References

1. Ward JW, Hinman AR. What is needed to eliminate hepatitis B virus and hepatitis C virus as global health threats. Gastroenterology. 2019;156(2):297-310.
2. Choo Q-L, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science. 1989;244(4902):359-62.
3. Kuo G, Choo Q-L, Alter H, Gitnick G, Redeker A, Purcell R, et al. An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. Science. 1989;244(4902):362-4.
4. Bostan N, Mahmood T. An overview about hepatitis C: a devastating virus. Critical reviews in microbiology. 2010;36(2):91-133.
5. Freeman AJ, Dore GJ, Law MG, Thorpe M, Von Overbeck J, Lloyd AR, et al. Estimating progression to cirrhosis in chronic hepatitis C virus infection. Hepatology. 2001;34(4):809-16.
6. Mühlberger N, Schwarzer R, Lettmeier B, Sroczynski G, Zeuzem S, Siebert U. HCV-related burden of disease in Europe: a systematic assessment of incidence, prevalence, morbidity, and mortality. BMC public health. 2009;9(1):34.
7. Perz JF, Armstrong GL, Farrington LA, Hutin YJ, Bell BP. The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. Journal of hepatology. 2006;45(4):529-38.
8. Hunziker IP, Zurbriggen R, Glueck R, Engler OB, Reichen J, Dai WJ, et al. Perspectives: towards a peptide-based vaccine against hepatitis C virus. Molecular immunology. 2001;38(6):475-84.
9. Messina JP, Humphreys I, Flaxman A, Brown A, Cooke GS, Pybus OG, et al. Global distribution and prevalence of hepatitis C virus genotypes. Hepatology. 2015;61(1):77-87.
10. Vieyres G, Thomas X, Descamps V, Duverlie G, Patel AH, Dubuisson J. Characterization of the envelope glycoproteins associated with infectious hepatitis C virus. Journal of virology. 2010;84(19):10159-68.
11. Hill A, Cooke G. Hepatitis C can be cured globally, but at what cost? Science. 2014;345(6193):141-2.
12. El Omari K, Iourin O, Kadlec J, Sutton G, Harlos K, Grimes JM, et al. Unexpected structure for the N-terminal domain of hepatitis C virus envelope glycoprotein E1. Nature communications. 2014;5:4874.
13. Drummer HE, Poumbourios P. Hepatitis C virus glycoprotein E2 contains a membrane-proximal heptad repeat sequence that is essential for E1E2 glycoprotein heterodimerization and viral entry. Journal of Biological Chemistry. 2004;279(29): 30066-72.
14. Falkowska E, Kajumo F, Garcia E, Reinus J, Dragic T. Hepatitis C virus envelope glycoprotein E2 glycans modulate entry, CD81 binding, and neutralization. Journal of virology. 2007;81(15):8072-9.
15. Duns G. Challenges and rewards: a career as a generalist. Australian family physician. 2013;42(7):439.

16. Khan AG, Whidby J, Miller MT, Scarborough H, Zatorski AV, Cygan A, et al. Structure of the core ectodomain of the hepatitis C virus envelope glycoprotein 2. Nature. 2014;509(7500):381.
17. Khan AG, Whidby J, Miller MT, Scarborough H, Zatorski AV, Cygan A, et al. Structure of the core ectodomain of the hepatitis C virus envelope glycoprotein 2. Nature. 2014;509(7500):381-4.
18. Sharma NR, Mateu G, Dreux M, Grakoui A, Cosset F-L, Melikyan GB. Hepatitis C virus is primed by CD81 protein for low pH-dependent fusion. Journal of Biological Chemistry. 2011;286(35):30361-76.
19. Vernelen K, Claeys H, Verhaert H, Volckaerts A, Vermylen C, Courouce A-M, et al. Significance of NS3 and NS5 antigens in screening for HCV antibody. The Lancet. 1994;343(8901):853-4.
20. Park S-H, Rehermann B. Immune responses to HCV and other hepatitis viruses. Immunity. 2014;40(1):13-24.
21. Pestka JM, Zeisel MB, Bläser E, Schürmann P, Bartosch B, Cosset F-L, et al. Rapid induction of virus-neutralizing antibodies and viral clearance in a single-source outbreak of hepatitis C. Proceedings of the National Academy of Sciences. 2007;104(14):6025-30.
22. Ashfaq UA, Javed T, Rehman S, Nawaz Z, Riazuddin S. An overview of HCV molecular biology, replication and immune responses. Virology journal. 2011;8(1):161.
23. Ip PP, Nijman HW, Wilschut J, Daemen T. Therapeutic vaccination against chronic hepatitis C virus infection. Antiviral research. 2012;96(1):36-50.
24. Dunlop J, Owsianka A, Cowton V, Patel A. Current and future prophylactic vaccines for hepatitis C virus. Vaccine: Development and Therapy. 2015;2015(5):31-44.
25. Pizza M, Scarlato V, Masignani V, Giuliani MM, Aricò B, Comanducci M, et al. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. Science. 2000;287(5459):1816-20.
26. Agadjanyan MG, Ghochikyan A, Petrushina I, Vasilevko V, Movsesyan N, Mkrtichyan M, et al. Prototype Alzheimer's disease vaccine using the immunodominant B cell epitope from β-amyloid and promiscuous T cell epitope pan HLA DR-binding peptide. The Journal of Immunology. 2005;174(3):1580-6.
27. Kaboré N, Poda G, Barro M, Cessouma R, Héma A, Ouedraogo A, et al. Impact de la vaccination sur les admissions pour méningites à Haemophilus influenzae b de 2004 à 2008, à Bobo Dioulasso (Burkina Faso). Medecine et sante tropicales. 2012;22(4):425-9.
28. Wang CY, Walfield AM. Site-specific peptide vaccines for immunotherapy and immunization against chronic diseases, cancer, infectious diseases, and for veterinary applications. Vaccine. 2005;23(17):2049-56.
29. Harrison G, Shakes T, Robinson C, Lawrence S, Heath D, Dempster R, et al. Duration of immunity, efficacy and safety in sheep of a recombinant Taenia ovis vaccine formulated with saponin or selected adjuvants. Veterinary immunology and immunopathology. 1999;70(3):161-72.
30. Chakraborty S. Ebola vaccine: multiple peptide-epitope loaded vaccine formulation from proteome using reverse vaccinology approach. Int J Pharm Pharm Sci. 2014;6:407-12.
31. Amer FA. Progress in developing hepatitis C virus prophylactic and therapeutic vaccines. Int J Curr Microbiol App Sci. 2014;3(7):891-906.
32. Matsueda S, Yamada A, Takao Y, Tamura M, Komatsu N, Yutani S, et al. A new epitope peptide derived from hepatitis C virus 1b possessing the capacity to induce cytotoxic T-lymphocytes in HCV1b-infected patients with HLA-A11,-A31, and-A33. Cancer Immunology, Immunotherapy. 2007;56(9):1359-66.

33. Murphy E, Yu D, Grimwood J, Schmutz J, Dickson M, Jarvis MA, et al. Coding potential of laboratory and clinical strains of human cytomegalovirus. Proceedings of the National Academy of Sciences. 2003;100(25):14976-81.
34. Saha S, Raghava G. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins: Structure, Function, and Bioinformatics. 2006;65(1):40-8.
35. Singh H, Raghava G. ProPred1: prediction of promiscuous MHC Class-I binding sites. Bioinformatics. 2003;19(8):1009-14.
36. Singh H, Raghava G. ProPred: prediction of HLA-DR binding sites. Bioinformatics. 2001;17(12):1236-7.
37. Sturniolo T, Bono E, Ding J, Raddrizzani L, Tuereci O, Sahin U, et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nature biotechnology. 1999;17(6):555.
38. Calis JJ, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, et al. Properties of MHC class I presented peptides that enhance immunogenicity. PLoS computational biology. 2013;9(10):e1003266.
39. Dikhit MR, Kumar A, Das S, Dehury B, Rout AK, Jamal F, et al. Identification of potential MHC Class II-restricted epitopes derived from Leishmania donovani antigens by reverse vaccinology and evaluation of their CD4+ T-Cell responsiveness against visceral leishmaniasis. Frontiers in immunology. 2017;8:1763.
40. Cruse JM, Lewis RE. Atlas of immunology: CRC Press; 2010.
41. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC bioinformatics. 2007;8(1):4.
42. Doytchinova IA, Flower DR. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. Vaccine. 2007;25(5):856-66.
43. Doytchinova IA, Flower DR. Bioinformatic approach for identifying parasite and fungal candidate subunit vaccines. Open Vaccine J. 2008;1(1):4.
44. Chou P, Fasman GD. Amino acid sequence. Adv Enzymol Relat Areas molec Biol. 2009;47:45.
45. Parker J, Guo D, Hodges R. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry. 1986;25(19):5425-32.
46. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. The proteomics protocols handbook: Springer; 2005. p. 571-607.
47. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. Elsevier Current Trends; 2000.
48. HAIBIN C. Design of protein linkers for the controlled assembly of nanoparticles 2009.
49. Emini EA, Hughes JV, Perlow D, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. Journal of virology. 1985;55(3):836-9.
50. Karplus P, Schulz G. Prediction of chain flexibility in proteins. Naturwissenschaften. 1985;72(4):212-3.
51. Bjellqvist B, Hughes GJ, Pasquali C, Paquet N, Ravier F, Sanchez JC, et al. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. Electrophoresis. 1993;14(1):1023-31.
52. Bjellqvist B, Basse B, Olsen E, Celis JE. Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. Electrophoresis. 1994;15(1):529-39.
53. Arias-Vásquez A, Altink ME, Rommelse N, Slaats–Willemse D, Buschgens C, Fliers E, et al. CDH13 is associated with working memory performance in attention deficit/hyperactivity disorder. Genes, Brain and Behavior. 2011;10(8):844-51.

54. Shehzadi A, ur Rehman S, Idrees M. Promiscuous prediction and conservancy analysis of CTL binding epitopes of HCV 3a viral proteome from Punjab Pakistan: an in silico approach. Virology journal. 2011;8(1):55.
55. McKiernan SM, Hagan R, Curry M, McDonald GS, Kelly A, Nolan N, et al. Distinct MHC class I and II alleles are associated with hepatitis C viral clearance, originating from a single source. Hepatology. 2004;40(1):108-14.
56. Gededzha MP, Mphahlele MJ, Selabe SG. Prediction of T-cell epitopes of hepatitis C virus genotype 5a. Virology journal. 2014;11(1):187.
57. Madej T, Lanczycki CJ, Zhang D, Thiessen PA, Geer RC, Marchler-Bauer A, et al. MMDB and VAST+: tracking structural similarities between macromolecular complexes. Nucleic acids research. 2013;42(D1):D297-D303.