Biologicals 53 (2018) 63–71

Contents lists available at ScienceDirect

Biologicals

journal homepage: www.elsevier.com/locate/biologicals

Identification of T cell and B cell epitopes against Indian HCV-genotype-3a for vaccine development- An *in silico* analysis

Varun Chauhan, Mini P. Singh*, Radha K. Ratho

Department of Virology, Post Graduate Institute of Medical Education and Research (PGIMER), Chandigarh, Punjab 160012, India

ARTICLE INFO

Keywords:
Docking
Epitopes
Hepatitis C virus
Human leucocyte antigen
In silico

ABSTRACT

Hepatitis C virus (HCV) infects almost 150 million people and is a leading cause of liver disease worldwide. It has been classified into seven genotypes; the most common genotype affecting Indian population is genotype 3 (60–70%). Currently there is no vaccine for any genotype of HCV. In order to develop peptide based vaccine against HCV, it is important to identify the conservancy in the circulating genotypes, along with the Human Leucocyte Antigen (HLA) alleles in the target population. The present study aims to identify conserved CD4 and CD8 T cells and B cell epitopes against Indian HCV-genotype-3a using an *in silico* analysis. In the present study, 28 promiscuous CD4 T cell epitopes and some CD8 epitopes were identified. The NS4 region was predicted to be the most antigenic with maximum number of conserved and promiscuous CD4 T cell epitopes and CD8 T cell epitopes having strong and intermediate affinity towards a number of HLA alleles prevalent in Indian population. Additionally, some linear B cell epitopes were also identified, which could generate neutralizing antibodies. In order to ascertain the binding pattern of the identified epitopes with HLA alleles, molecular docking analysis was carried out. The authors suggest further experimental validation to investigate the immunogenicity of the identified epitopes.

1. Introduction

The Hepatitis C virus (HCV) affects nearly 3% of the world's population and is second to Human Immunodeficiency virus (HIV) in terms of morbidity and mortality among the emerging infections. The previously used interferon therapy for hepatitis C treatment was expensive, toxic, and required prolonged duration of therapy [1]. Recently the Direct Acting Antivirals (DAAs), have come up which can be given orally, require shorter duration of treatment and are known to achieve better sustained virological response (SVR). Recently the new HCV NS5A replication complex inhibitors of HCV, Ledipasvir and Daclatasvir along with Sofosbuvir (NS5B inhibitor) are usually given to patients infected with HCV genotype 1 and 3 [2,3]. However, resistance to these drugs too has been recently reported by several research groups [4–9]. Thus reinfection remains possible, making control of this leading cause of chronic liver disease difficult. The major hallmark of persistent HCV infection is the lack of B cells capable of generating neutralizing antibodies and CD4 T cells that proliferate in response to its antigenic stimulation [10–14]. There are several studies suggesting that neutralizing antibodies are crucial for protection against HCV infection [15,16].

The recent advances in the field of bioinformatics have contributed to the development of rationally designed peptide based vaccines. Several vaccine targets have been proposed so far for HCV, which have not proved to be promising due to lack of either ethnicity or specificity factors [17]. Genetic heterogeneity of HCV is the major obstacle in vaccine development [18]. The CD4 (HLA Class II) T cell epitopes are the key players of immune defense mechanism known for activation of protective B cell and CD8 T cell responses in host. The HLA molecules too are polymorphic, varying in different population. Hence, it is relevant to identify conservancy in the circulating HCV genotypes along with the HLA alleles prevalent in the target population [19]. Two peptide based vaccines are already running under clinical trials, IC41 has completed a randomized double-blind phase II study in patients with chronic HCV infection who had either relapsed or failed to respond to previous PEG-IFN/ribavirin therapy [20]. Another vaccine, composed of peptides from HCV core region (C35–44) with an emulsified incomplete Freund adjuvant ISA51, has shown to be well tolerated in HCV-infected patients in a phase I clinical trial [21]. Recently in phase II study, the researchers reported the peptide-specific IgG responses in HCV positive hepatocellular carcinoma patients and proposes further clinical trials for this vaccine [22]. These studies indicate that the peptide based vaccines could play an important role against HCV infections.

* Corresponding author. Room No-609, Department of Virology, Research Block A, PGIMER, Chandigarh 160012, India.
E-mail address: minipsingh@gmail.com (M.P. Singh).

https://doi.org/10.1016/j.biologicals.2018.02.003
Received 6 February 2017; Received in revised form 21 February 2018; Accepted 23 February 2018
Available online 05 March 2018
1045-1056/ © 2018 International Alliance for Biological Standardization. Published by Elsevier Ltd. All rights reserved.

Table 1
Physico-chemical properties and antigenicity prediction results of all the target proteins.

Protein	Molecular weight	Theoretical pI	Amino acid composition	Antigenicity score (Threshold 0.4)
Core	12963.8	12.05	Ala(A):7.4%; Arg(R):12.2%; Asn(N):3.7%; Asp(D):2.6%; Cys(C):2.1%; Gln(Q):3.7%; Glu(E):2.1%; Gly(G):13.8%; His(H):1.1%; Ile(I):4.2%; Leu(L):9.5%; Lys(K):3.7%; Met(M):0.5%; Phe(F):2.6%; Pro(P):10.6%; Ser(S):5.8%; Thr(T): 4.2%; Trp(W): 2.6%; Tyr(Y):2.1%; Val(V):5.3.%	0.3062 (non-antigenic)
E1	20974.3	6.38	Ala(A):9.4%; Arg(R):4.2%; Asn(N):3.1%; Asp(D):4.2%; Cys(C):3.1%; Gln(Q):4.2%; Glu(E):1.6%; Gly(G):8.4%; His(H):3.7%; Ile(I):4.2%; Leu(L):7.9%; Lys (K):0.5%; Met(M):5.2%; Phe(F):3.1%; Pro (P): 5.2%; Ser(S):5.8%; Thr (T):6.8%; Trp (W):4.2%; Tyr(Y):3.1%; Val(V):12.0%	0.4433 (antigenic)
E2/NS1	38787.1	8.24	Ala(A): 6.9%; Arg(R):5.7%; Asn(N):5.1%; Asp(D):4.0%; Cys(C):5.4%; Gln(Q):2.3%; Glu(E):2.9%; Gly(G):9.4%; His(H):2.9%; Ile(I):3.1%; Leu(L):8.3%; Lys (K):2.3%; Met(M):0.9%; Phe(F):6.0%; Pro(P):7.7%; Ser(S):5.7%; Thr(T):8.0%; Trp(W):3.1%; Tyr(Y):4.0%; Val(V):6.3%	0.4736 (antigenic)
NS3	15423.7	8.70	Ala(A):11.4%; Arg(R):5.4%; Asn(N):0.7%; Asp(D):3.4%; Cys(C):3.4%; Gln(Q):4.0%; Glu(E):2.7%; Gly(G):11.4%; His(H):2.0%; Ile(I):2.0%; Leu(L):9.4%; Lys(K):2.7%; Met(M):2.0%; Phe(F):2.0%; Pro(P):7.4%; Ser(S):6.7%; Thr(T):10.1%; Trp(W):1.3%; Tyr(Y):2.0%; Val(V):10.1%	0.3903 (non-antigenic)
NS4a	5751.7	4.25	Ala(A):9.3%; Arg(R):0.0%; Asn(N):0.0%; Asp(D):3.7%; Cys(C):5.6%; Gln(Q):3.7%; Glu(E):9.3%; Gly(G):11.1%; His(H):1.9%; Ile(I):3.7%; Leu(L):14.8%; Lys(K):3.7%; Met(M):1.9%; Phe(F):0.0%; Pro(P):3.7%; Ser(S):3.7%; Thr(T):1.9%; Trp(W):1.9%; Tyr(Y):5.6%; Val(V):14.8%	0.7566 (antigenic)
NS4b	20221.6	8.94	Ala(A):12.4%; Arg(R):2.6%; Asn(N):3.6%; Asp(D):1.0%; Cys(C):0.5%; Gln(Q):4.1%; Glu(E):3.1%; Gly(G):12.4%; His(H):2.1%; Ile(I):6.2%; Leu(L):12.4%; Lys(K):2.6%; Met(M):3.1%; Phe(F):4.1%; Pro(P):4.6%; Ser(S):5.7%; Thr(T):6.7%; Trp(W):2.6%; Tyr(Y):1.0%; Val(V):9.3%	0.4380 (antigenic)
NS51a	6704.7	9.44	Ala(A)8.1%; Arg(R)6.5%; Asn(N)4.8%; Asp(D)1.6%; Cys(C)6.5%; Gln(Q)1.6%; Glu(E)1.6%; Gly(G)14.5%; His(H)3.2%; Ile(I)4.8%; Leu(L)3.2%; Lys(K) 4.8%; Met(M)4.8%; Phe(F)3.2%; Pro(P)6.5%; Ser(S)4.8%; Thr(T)9.7%; Trp(W)3.2%; Tyr(Y)3.2%; Val(V)3.2%	0.1424 (non-antigenic)

Keeping these points under consideration, the following study aims to identify B cell epitopes and HLA Class II (CD4) T cell epitopes using *in silico* tools which could be utilized for vaccine development against HCV genotype 3 which is the predominant genotype in Indian population.

2. Methodology

2.1. Retrieval of amino acid sequences, molecular and structural analysis

The amino acid sequences of the poly-protein of Indian HCV genotype-3a were retrieved from NCBI database. The names and accession numbers of all the sequences analyzed in the present study were: ADE10208, AGQ17412, AGQ17413, AGQ17414, AGQ17415, AGQ17416, AFH74066, AFH74067, AFH74069, AFH74070, AFH74071, AFH74072, AFA36246, and ADV04529. The amino acid sequences obtained were further aligned using multiple sequence alignment tool Clustal Omega in order to find the conservancy among the proteins of HCV genotype 3a. Different physicochemical properties of the target proteins were also analyzed using ExPASy ProtParam tool (http://web.expasy.org/protparam/). The antigenicity determination of the target proteins was also carried out using VaxiJen v2.0 server, which is used for the prediction of subunit vaccines and protective antigens [23]. Here, the default parameter of the server was used for antigenicity determination. The threshold for antigenecity prediction was kept 0.4.

The secondary structure of the proteins was predicted using the improved self-optimized prediction method (SOPMA) software [24]. The four conformational states, including sheets, coils, helices and turns were analyzed in the protein sequence.

2.2. Identification of T and B cell epitopes

For the identification of HLA class-II T cell epitopes, the servers NetMHCIIpan 3.1, ProPred, MultiPred2, IEDB-NN and IEDB-SMM were used [25–29]. The alleles predominant in Indian population were selected for the study [30–34].

The alleles selected were: HLA-DRB1*01:01, HLA-DRB1*04:01, HLA-DRB1*07:01, HLA-DRB1*08:03, HLA-DRB1*11:01, HLA-DRB1*12:02, HLA-DRB1*14:01 and HLA-DRB1*15:01. Based on the binding affinity for HLA alleles, the T cell epitopes were classified into the following 3 groups. The peptides having half-maximal inhibitory concentration less than 50 (IC_{50s} < 50 nM) were considered as high-affinity binding epitopes, whereas IC_{50s} of < 500 nM were considered as intermediate-affinity binding epitopes; and IC_{50s} of < 5000 nM were low-affinity binding epitopes. HLA Class I T cell epitopes were identified using Net MHC 4.0 server. The threshold for predicting the strong binding epitopes was kept 0.5% and for weak binding epitopes was kept 2.0% [35].

The identification of linear B cell epitopes was carried out using BCPREDS, ABCpred, BepiPred 1.0, and LBtope online web servers [36–39]. All the servers were run at the default parameters without any alteration in the prediction methods. BLASTP search was also performed against human proteome in order to find any peptides showing identity to human proteins.

2.3. Characterization of the predicted peptides

Different parameters like molecular weight, theoretical isoelectric point (pI), estimated half-life of the predicted T and B cell epitopes were calculated using ExPASy ProtParam tool (http://web.expasy.org/protparam/). The antigenicity of each predicted epitope was analyzed by VaxiJen v2.0 tool (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html). The threshold for antigenicity prediction was kept 0.4. The conservancy analysis of the predicted T and B cell epitopes was carried out by IEDB Conservancy Analysis tool [40].

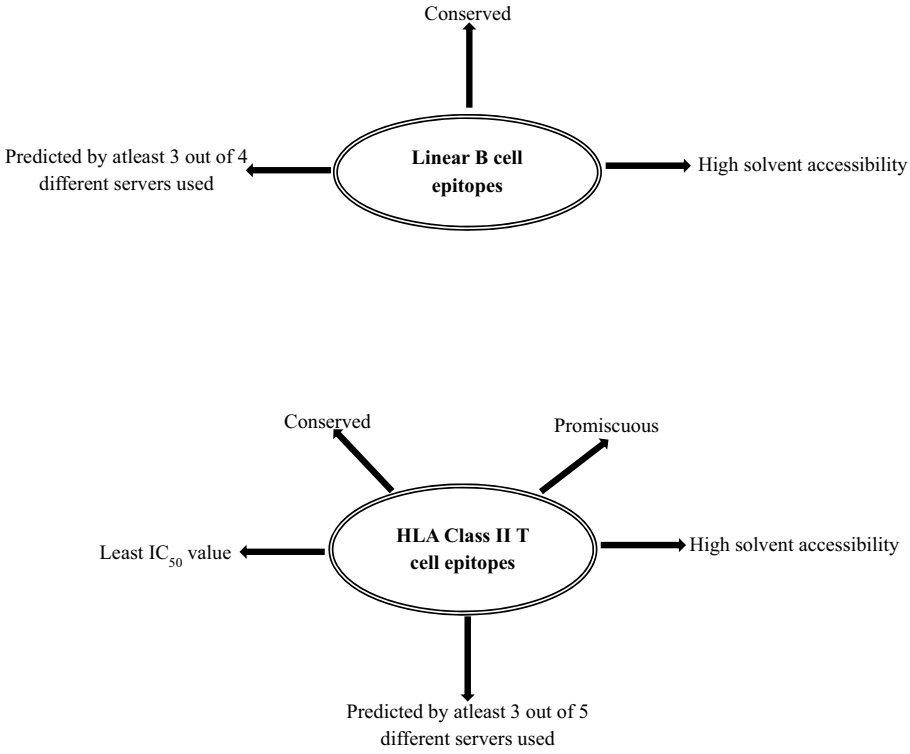

Fig. 1. Criteria for identification of T and B cell epitopes in the present study.

2.4. Modeling of peptides, HLA alleles and molecular docking analysis

The X-ray crystallographic structures of HLA-DRB1* 01:01 and HLA-DRB1* 15:01 alleles were retrieved from Protein Data Bank (PDB) bearing PDB ID- 2g9h and 1BX2 respectively. The energy minimization for the receptors (HLA alleles) was further carried out after successfully removing the previously bound ligand. The 3D structures of the rest of the HLA alleles were modeled using Modeller 9v7 Program [41]. The pdb files of the peptides were created using PEPFOLD server [42]. The receptor and the ligand files were prepared using AutoDock 4.2. The quality of the models of HLA alleles prepared were further checked by Ramachandran Plot analysis using Rampage server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php). Finally the most promising peptide predicted was further subjected to molecular docking analysis for analyzing its binding pattern towards different HLA alleles using AutoDock4.2 program [43]. This server provides 10 independent Lamarkian genetic algorithm (GA) runs for each docking analysis. The run showing lowest binding affinity was considered. Finally the interaction pattern of the peptide with HLA alleles was visualized using PYMOL version 1.7.4.4. (Schrodinger).

2.5. Validation of epitopes identification and molecular docking analysis

The T and B cell epitopes identified by different servers were also subjected to IEDB analysis database which already contains number of well-known validated epitopes. Finally the epitopes identified by IEDB were matched with the epitopes identified by other servers and those showing exact matches were chosen for further analysis. Molecular docking analysis was validated by carrying out the docking between the peptide Staphylococcal Enterotoxin I (SEI) with HLA-DRB1*01:01 (PDB id - 2g9h) using the same protocol as done for other predicted epitopes [44].

3. Results

3.1. Molecular and structural analysis

The multiple sequence alignment results revealed that the NS51b protein was found to be highly variable showing no conservancy and was thus excluded from further study. Rest of the proteins showed significant conservancy and were thus further investigated for *in silico* analysis. The amino acid composition, molecular weight and theoretical pI of each protein is shown in Table 1. The antigenicity prediction results showed NS4a region to be most antigenic with an antigenicity score of 0.7 which was much higher than the threshold value 0.4, the default parameter of the VaxiJen v2.0 server. The secondary structure prediction results (α helical region, β turns and random coils) of all the target proteins is shown in Supplementary Fig. 1.

3.2. Identification of T and B cell epitopes

Identification of T cell epitopes was carried out using 5 different servers. The peptides predicted by at least 3 servers were considered. The basis of using multiple number of servers was to assure the prediction accuracy of the results. The peptides having the least IC_{50} value towards the targeted HLA alleles were considered for further characterization. Narrowing down the prediction criteria (Fig. 1) 28 HLA Class II T cell epitopes (Core-9, E1-3, E2-NS1-3, NS3-5, NS4a-2, NS4b-5, NS51a-1) were predicted to be the most promising. In addition several conserved HLA Class I epitopes were also predicted. The epitopes predicted in the NS4a and NS4b regions were highly conserved (Supplementary Fig. 2). The T cell epitopes identified in the NS4b region lying at position 164, 165 and 166 were predicted to be highly promiscuous having high and intermediate affinity towards most of the HLA alleles included in the study and were also predicted to be antigenic in nature (Table 2). Similarly, the linear B cell epitopes predicted by multiple number of servers were considered. Overall 11linear B cell epitopes (Core-3, E1-1, E2-NS1-4, NS3-1, NS4a-1, NS4b-1) were identified. The B cell epitopes in the E2-NS1 region were predicted to be

Table 2
Predicted HLA Class I T cells epitopes.

Protein	Peptide position	Peptide seq	HLA Class I Allele	% Rank Bind Level
Core	170	CSFSIFLLA	A0101	0.80
	166	NLPGCSFSI	A0201	0.50
	75	AQPGYPWPL		0.50
	148	ALAHGVRAL		1.70
	0	STLPKPQRK	A0301	0.90
	13	TIRRPQDVK		1.0
	57	RGRRQPIPK		1.50
	8	KTKRNTIRR		1.70
	73	SWAQPGYPW	A2402	0.40
	83	LYGNEGCGW		1.60
	166	NLPGCSFSI		1.90
	55	QPRGRRQPI	B0702	0.01
	109	DPRRRSRNL		0.30
	148	ALAHGVRAL		0.70
	167	LPGCSFSIF		0.90
	140	APVGGVARA		1.20
	113	RSRNLGKVI		1.80
	86	NEGCGWAGW	B4403	0.07
	57	RGRRQPIPK	A3001	0.01
	8	KTKRNTIRR		0.07
	111	RRRSRNLGK		0.07
	49	KTSERSQPR		0.70
	0	STLPKPQRK		0.80
	13	TIRRPQDVK		0.90
	147	RALAHGVRA		0.90
	145	VARALAHGV		1.50
	167	LPGCSFSIF	B3501	0.12
	22	FPGGGQIVG		0.40
	76	QPGYPWPLY		0.60
	79	YPWPLYGNE		1.50
E1	76	ATMCSALYV	A0101	1
	75	AATMCSALY		1.2
	76	ATMCSALYV	A0201	0.30
	170	SMQGNWAKV		1.40
	167	AYYSMQGNW	A2402	0.60
	74	GAATMCSAL	B0702	2.00
	76	ATMCSALYV	A3001	0.17
	163	LAGLAYYSM	B3501	0.40
	75	AATMCSALY		0.50
	74	GAATMCSAL		1.70
E2/NS1	317	NIVDVQYLY	A0101	1.30
	121	VYCFTPSPV	A2402	1.30
	317	NIVDVQYLY	B3501	0.60
	315	HQNIVDVQY		1.00
	179	CGAPPCNIY		1.50
NS3	45	YTNVDQDLV	A0101	0.40
	13	FLGTTVGGV	A0201	0.50
	127	AVCTRGVAK	A0301	0.50
	44	MYTNVDQDL	A2402	1.90
	115	CPSGHVAGI	B0702	1.00
	129	CTRGVAKAL		1.20
	19	HVAGIFRAA		1.70
	125	RAAVCTRGV		1.70
	123	IFRAAVCTR	A3001	0.80
	127	AVCTRGVAK		1.20
	117	SGHVAGIFR		1.30
	126	AAVCTRGVA		1.70
NS4a	11	ALAAYCLSV	A0201	0.03
	3	VLLGGVLAA		0.40
	4	LLGGVLAAL		0.60
	10	AALAAYCLS	A3001	1.80
	7	GVLAALAAY	B3501	0.50
	9	LAALAAYCL		0.70
	12	LAAYCLSVG		1.70
NS4b	73	LTTNQTMFF	A0101	0.80
	40	FVSGIQYLA		1.70
	36	HMWNFVSGI	A0201	0.20
	158	ILSPGALVV		0.60
	40	FVSGIQYLA		0.60
	51	STLPGNPAV		1.60
	131	VSGALVAFK	A0301	0.50
	168	VICAAILRR		1.50
	32	FWHKHMWNF	A2402	0.08
	45	QYLAGLSTL		0.40
	75	TNQTMFFNI		0.80
	71	SPLTTNQTM	B0702	0.12
	56	NPAVASLMA		1.30
	12	YGAGVSGAL		1.50
	131	VSGALVAFK	A3001	1.20
	51	STLPGNPAV		1.30
	173	ILRRHVGPG		1.30
	71	SPLTTNQTM	B3501	0.25
	57	PAVASLMAF		0.80
	127	YGAGVSGAL		0.90
	155	LPAILSPGA		0.90
	151	MVNLLPAIL		1.20
	56	NPAVASLMA		1.30
	130	GVSGALVAF		1.40
	53	LPGNPAVAS		1.40
	32	FWHKHMWNF		1.40
	38	WNFVSGIQY		1.40
	73	LTTNQTMFF		1.70
NS5a	48	NMWHGTFPI	A0201	0.01
	2	FISCQKGYK	A0301	0.90
	42	GPRTCANMW	B0702	1.40
	32	HVKNGSMRL		1.90
	32	HVKNGSMRL	A3001	0.15

most antigenic (Table 3). Overall the region 161-SPGALVVGVICAAIL-RRHVGPGEGAVQWMNR-191 from NS4b protein was predicted most promising for vaccine development. The immunogenic properties of this region are detailed out in Fig. 2.

3.3. Modeling of peptides, HLA alleles and molecular docking analysis

The 3D models of all four epitopes were obtained using PEPFOLD server. The Ramachandran Plot analysis of the 3D structures of HLA alleles prepared through homology modeling revealed satisfactory results with more than 98% residues lying in the favored and allowed region (Supplementary Fig. 3, Supplementary Table 1). Further the docking analysis of the four most promising T cell epitopes of (2 from NS4b protein (VVGVICAAILRRHVG and SGIQYLAGLSTLPGN) at 165 and 42 position); 1 from NS4a protein (LLGGVLAALAAYCLS) at position 4; and 1 from Core protein (GGVARALAHGVRALE at position 143) was carried out with the HLA-DRB1*01:01, HLA-DRB1*03:01, HLA-DRB1*07:01, HLA-DRB1*10:01, HLA-DRB1*11:01, HLA-DRB1*13:01, HLA-DRB1*14:01 and HLA-DRB1*15:01 alleles in order to investigate the affinity and binding pattern of the epitopes. The binding pattern of all 4 peptides with HLA DRB1*0101 and HLA DRB1*1501 alleles as analyzed by docking analysis is shown in Fig. 3. Others are shown in Supplementary Fig. 4. The binding energies of peptides with respective HLA alleles are shown in Table 4. The peptide with sequence LLGGV-LAALAAYCLS located at position 4 from NS4a region showed the highest binding energy (−9.3 kcal/mol) with HLA DRB1*10:01 allele (Fig. 4) (Table 5). For validation of docking the Staphylococcal Enterotoxin I (SEI) peptide was docked with HLA DRB1*0101 allele using the same strategy followed for other predicted epitopes. The docked fig is shown in Supplementary Fig. 5. The overall strategy involved in HCV epitopes prediction in the present study is summarized in Fig. 5.

Table 3
Predicted HLA Class II T cells epitopes.

Proteins	Peptide sequence[a]	Peptide position	Antigenicity score	Mol. wt.	Minimum Identity[b]	HLA-DRB1 alleles[c]									Tools[d]
						*01:01	*03:01	*04:01	*07:01	*08:01	*10:01	*11:01	*12:01	*14:01	*15:01	
Core	IVggVyVLP**RR**gP**R**L	**28**	0.649	1652.0	100%											1,2,3,5
	ggVyVLP**RR**gP**R**LgV	**30**	0.859	1595.9	100%											1,2,3,5
	gVyVLP**RR**gP**R**LgV**R**	**31**	1.178	1695.0	93.3%											1,2,3,5
	LMgyIPLVgAPVggV	**131**	0.497	1442.7	93.3%											1,2,3,4,5
	MgyIPLVgAPVggVA	**132**	0.350	1400.7	93.3%											1,2,3,4,5
	VggVA**R**ALAHgV**R**AL	**142**	0.220	1446.7	100%											1,2,3,5
	ggVA**R**ALAHgV**R**AL**E**	**143**	0.299	1476.7	100%											1,2,3,5
	sIFLLALLsCLIHPA	**173**	0.031	1611.0	93.3%											1,2,3,5
	IFLLALLsCLIHPAA	**174**	0.009	1595.0	93.3%											1,2,3,5
E1	LVgqAFtF**RP****RR**Hqt	**94**	0.777	1814.0	93.3%											1,2,3,5
	VgqAFtF**RP****RR**HqtV	**95**	0.693	1800.0	93.3%											1,2,3,5
	MVVAHVL**R**LPqtLF**D**	**140**	-0.363	1739.1	93.3%											1,2,3,4,5
E2-NS1	CtVnFtLF**K**V**R**MFVg	**240**	0.305	1762.1	86.6%											1,3,5
	LPCsFtPMPALstgL	**295**	0.658	1534.8	93.3%											1,2,3
	PCsFtPMPALstgLI	**296**	0.640	1607.9	93.3%											1,2,3
NS3	HVAgIF**R**AAVCt**R**gV	**119**	-0.202	1556.8	100%											1,3,5
	gIF**R**AAVCt**R**gVA**K**A	**122**	-0.444	1519.8	100%											1,3,5
	IF**R**AAVCt**R**gVA**K**AL	**123**	-0.266	1575.9	100%											1,3,5
	VA**K**ALqFIPV**E**tLst	**133**	0.751	1616.9	93.3%											1,3,4,5
	A**K**ALqFIPV**E**tLstq	**134**	0.862	1645.9	93.3%											1,3,4,5
NS4a	LggVLAALAAyCLsV	**5**	0.627	1420.7	100%											1,2,3,5
	LLggVLAALAAyCLs	**4**	0.631	1434.7	100%											1,2,3,5
NS4b	LVVgVICAAIL**RR**HV	**164**	0.618	1619.0	100%											1,3,5
	VVgVICAAIL**RR**HVg	**165**	0.687	1562.9	100%											1,3,5
	VgVICAAIL**RR**HVgP	**166**	0.859	1560.9	100%											1,,3,5
	VsgIqyLAgLstLPg	**41**	0.077	1475.7	100%											1,,2,3,5
	sgIqyLAgLstLPgn	**42**	0.129	1490.6	100%											1,2,3,5
NS51a	**K**ngsM**R**LAgP**R**tCAn	**34**	-0.77	1575.8	93.3%											1,2,3,5

a- Non polar, polar and charged residues are shown in upper case, lower case and bold respectively.
b- Minimum identity of the identified epitopes among protein sequences included in the study predicted by IEDB Conservancy Analysis tool.
c- Heatmap showing the intermediate (IC_{50s} of <500 nM) and high affinity (IC_{50s} <50 nM) of the predicted epitopes towards the HLA alleles as predicted by NetMHCIIpan3.1 server. Dark grey color shows high affinity and light grey color shows intermediate affinity.
d- Tools that recognized the corresponding T cell epitope are numbered as follows: (1) IEDB-NN (2) IEDB-SMM, (3) NetMHCIIpan 3.1, (4) MultiPred, (5) ProPred.

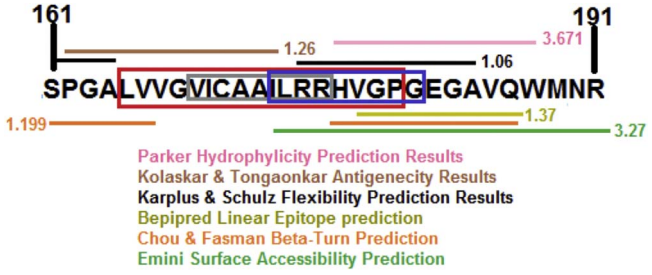

Fig. 2. One of the most promising region predicted for vaccine development. The region enclosed in red box predicted to be highly promiscuous having affinity towards maximum number of HLA DRB1 alleles (i.e DRB1*0301, DRB1*0801, DRB1*0301, DRB1*1201, DRB1*1401 and DRB1*1501). The region enclosed in grey and blue box were predicted to have affinity towards HLA*A03:01 and HLA*A30:01. Different colored lines indicate different properties with maximum scores for Linear B cell epitope prediction as predicted by IEDB Analysis Resource.

4. Discussion

Synthetic peptide based vaccines have a potential in future as these vaccines utilize only a small amount of microbial component which could elicit protective immune response in the host [45]. For developing a vaccine in the post genomic era, a major challenge is to identify the most immunogenic regions that can stimulate different arms of the immune responses [46]. A number of computational tools have been developed which could help in screening out the most immunogenic epitopes from a cluster of epitopes. Designing a peptide based vaccines using this technology has already provided promising results against human papilloma virus [47], influenza virus [48] and also vaccines against common allergens [49] have been prepared using this approach.

The major obstacle in vaccine development against HCV is its genetic heterogeneity. Hence, it will be relevant to identify conservancy in the circulating HCV genotypes in the target population in that particular area [18]. The most common genotype of HCV affecting the

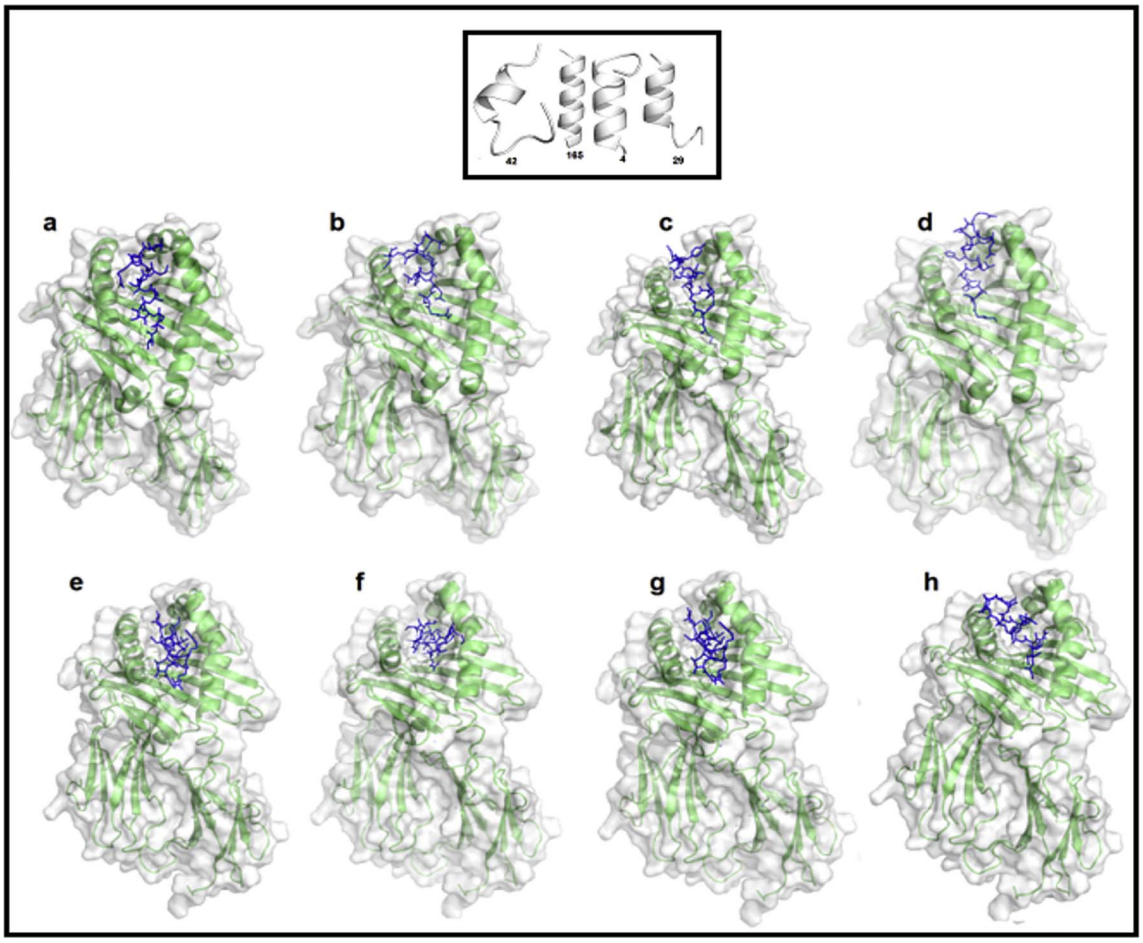

Fig. 3. Binding pattern of the peptides with HLA DRB1*01:01 and HLA DRB1*15:01 alleles. a, b, c and d are the binding patterns of peptides 165, 42, 4 and 29 with HLA DRB1*01:01 and e, f, g and h are the binding patterns of peptides 165, 42, 4 and 29 with HLA DRB1*15:01. The peptides deigned using PEPFOLD server are shown in top box.

Table 4
Linear B cell epitopes predicted by different servers used.

Protein	Peptide sequence[a]	Antigenicity	Location	Minimum Identity[b]	Tools[c]
Core	LP**K**Pq**RK**t**KR**ntl**RR**Pq**D**	−0.401	3–20	100%	1,2,3,4
	P**R**gs**R**PswgPn**D**P**RRR**	0.975	99–114	93.75%	1,2,3,4
	AtgnLPgCsFs	−1.17	164–174	100%	1,2,3
E1	tqFICwtPVtPPVAVsgVgA	1.140	43–62	60%	1,2,3
E2/NS1	t**D**AnItgPs**DER**PyCwHyAP	0.975	88	90%	1,2,3,4
	tCgAPPCnIygggE**nn**R	0.141	179–195	70.59%	1,2, 3,4
	tPsPVVVgtt**DAR**gqPtynw	1.218	126–145	80%	1,2,3
	RgERCDlEDRDRsEqH	1.260	269–284	87.5%	1,2,3
NS3	APPgA**K**s**L**EPCtCg	0.377	58–70	92.86%	2,3,4
NS4a	PALVP**DKE**VLyqqy**DEMEE**C	0.174	35–54	95%	2,3,4
NS4b	AIL**s**PgALVVgVICAAIL**RR**HVgPgEgAVqwMn**R**	0.602	161–191	100%	1,2,3

[a] Non polar, polar and charged residues are shown in upper case, lower case and bold respectively.
[b] Minimum identity of the identified epitopes among protein sequences included in the study predicted by IEDB Conservancy Analysis tool.
[c] Tools that recognized the corresponding B cell epitope are numbered as follows: (1) BCPREDS; (2) ABCpred; (3) BepiPred 1.0; (4) LBtope.

Indian population is the genotype 3 (60–70%). The aim of the study was to identify HLA Class II T cell epitopes targeting HCV genotype 3 in context with the HLA alleles prevalent in Indian population because vaccines targeting T-cell epitopes are considered to be more promising as they are capable of evoking a long-lasting immune response. In addition an antigen could escape the memory response generated by antibody due to antigenic drift. In the present study some linear B cell epitopes have also been identified which could be utilized in vaccine development and could elicit a humoral immune response in host.

Recently, Sheikh et al., [50] predicted the $CD8^+$ and $CD4^+$ T cell epitopes against influenza virus using an *in silico* approach. Alam et al., [51] have also predicted some promising T cell epitopes against Zika virus using an *in silico* strategy. These researches indicate the potential of this technology in the field of vaccinology. Recently Khan et al. [52] experimentally validated the*in silico* findings of predicted HLA Class I and Class II T cell epitopes targeting secretory protein (Ag85B) of *Mycobacterium tuberculosis*. They selected the epitopes with affinity (IC_{50} < 500 nm), however in the present study only those epitopes showing affinity < 50 nm have been selected thus making the results more stringent.

The criteria for selecting the HLA Class II T cell epitopes in the present study was that it should have least IC_{50} value, should be

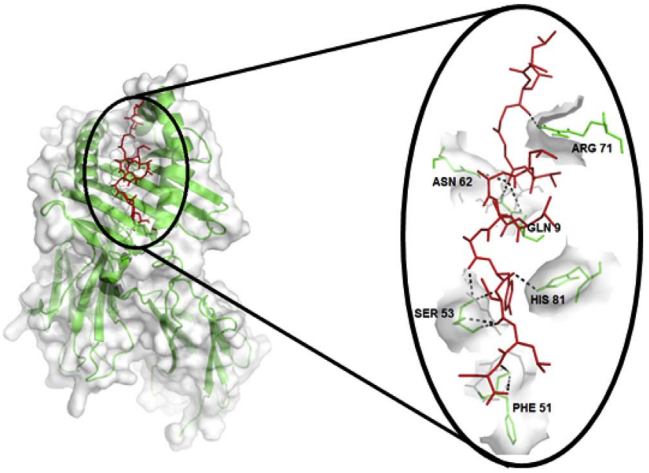

Fig. 4. Representation of the binding pattern of the peptide 4 (LLGGVLAALAAYCLS) from NS4a region with HLA DRB1*10:01 allele.

conserved, promiscuous and predicted by at least 3 out of 5 different servers used. Based on these criteria, a total of 28 HLA Class II T cell epitopes were predicted. In addition some HLA Class I T cell epitopes were also predicted. The core, NS3 and NS4 regions were predicted to have maximum numbers of T cell epitopes. However, the peptides belonging to NS4b region were predicted to be highly promiscuous having strong and intermediate affinity towards number of HLA alleles prevalent in Indian population. In addition, 15 linear B cell epitopes were also predicted. Further antigenicity of each peptide was also predicted, as such peptides are sometimes considered to be weak immunogens. Most of the peptides showed an antigenicity score above the threshold value 0.4 which is used as a cut off value for antigenicity prediction by VaxiJen v2.0 server. This server has already been utilized by several researchers for antigenicity determination of their predicted epitopes against number of pathogens like Zika Virus [51], Bronchitis Virus [53], influenza virus. Further, the study also determined the binding pattern of one of the peptide of NS4b region with the HLA alleles by molecular docking analysis using Autodock 4.0 tool. The peptide showed strong binding affinities to the HLA alleles with relevant hydrogen bonding.

The major challenge in peptide based vaccines is that they are not very immunogenic, thus the researchers propose that a suitable adjuvant is needed for such vaccines [45]. Thus the selection of a suitable adjuvant (delivery system) is the second most important criteria, just after the selection of epitopes in peptide based vaccines [54]. Currently several experimental adjuvants are known for their ability to induce immune responses like TLRs agonists, surface proteins on APCs which recognize PAMP and several other etc [55]. Other systems for synthetic vaccine delivery includes emulsions like Freund's complete and incomplete adjuvant, liposomes, virosomes, Immunostimulatory complexes (ISCOMs), polymetric particles like polylactides, polyglycolic acid or polyglycolides etc. Particulate systems like silicon dioxide nanoparticles, carbon nanotubes, gold nanoparticles, peptide non carriers, liposome polycation DNA (LPD) complex, liposomes and w/o emulsion are also used as adjuvants for peptide based vaccine delivery.

5. Conclusion

As HCV comes second to HIV among viruses in terms of heterogeneity in its genome, hence it is important to identify the conservancy in the circulating genotypes along with the HLA alleles prevalent in a particular area. To the best of our literature search, this is the first study to carry out identification of immundominant peptides targeting different proteins of Indian HCV genotype 3 using an *in-silico* approach. The epitopes from the NS4b region were predicted to be highly promiscuous covering most of the HLA Class II alleles prevalent in Indian population. It also contained some overlapping HLA Class I epitopes as well. This study has led to the identification of potentially immunogenic epitopes which should further be tested for their immunoreactivity using *in vitro* and *in vivo* approaches to support the *in-silico* findings.

Authors contribution

Varun Chauhan: Carried out *in silico* analysis.

Mini P Singh: Planned out the whole work and helped the author in writing the manuscript.

R K Ratho: Reviewed the whole manuscript and provided the necessary workstation for carrying out the present work.

Conflicts of interest

None.

Acknowledgement

We are thankful to Ms. Megha Sood, an IELTS trainer in Chandigarh, for reviewing the manuscript for making grammatical corrections.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.biologicals.2018.02.003.

Table 5
Affinities of peptides towards different HLA alleles after docking analysis.

Alleles	Peptides Docking energies (kcal/mol)			
	NS4b		NS4a	Core
	VVGVICAAILRRHVG (165)	SGIQYLAGLSTLPGN (42)	LLGGVLAALAAYCLS (4)	GGVARALAHGVRALE (143)
HLA DRB1*01:01	−6.6	−7.3	**−8.9**	−6.9
HLA DRB1*03:01	−6.9	−6.5	**−8.7**	−6.3
HLA DRB1*07:01	−6.3	−6.5	−7.0	−6.1
HLA DRB1*10:01	−7.8	−7.3	**−9.3**	−7.7
HLA DRB1*11:01	−5.9	−6.6	−6.7	−6.8
HLA DRB1*13:01	−7.3	**−8.5**	**−8.3**	−7.1
HLA DRB1*14:01	−6.0	−6.8	−7.8	−7.5
HLA DRB1*15:01	−6.1	−5.8	−7.7	−6.6

Note: The following are the docking scores of the peptides from NS4b at position 164 and 42, NS4a at position 4 and Core at position 143. The docking scores depicted in bold are the highest docking scores obtained.

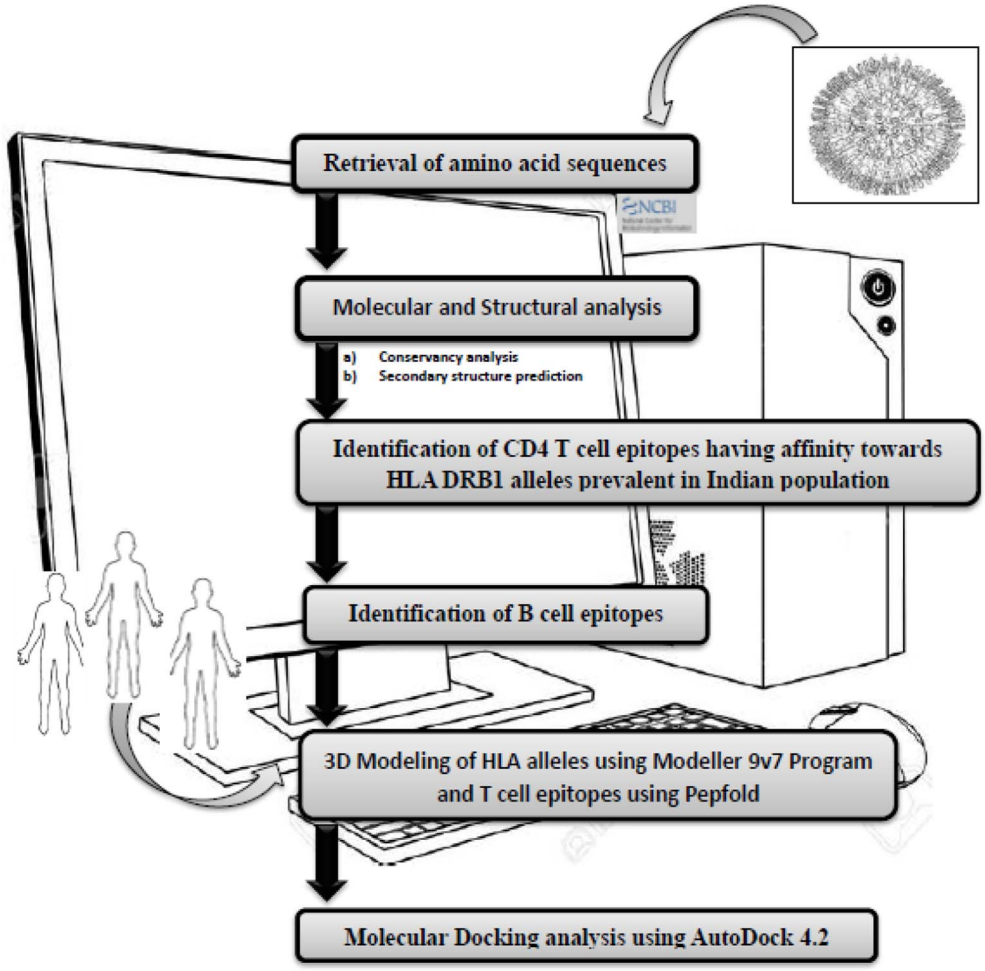

Fig. 5. Flowchart showing the strategy involved for epitopes identification in the present study.

References

[1] Manns MP, Wedemeyer H, Cornberg M. Treating viral hepatitis C: efficacy, side effects, and complications. Gut 2006;55:1350–9.

[2] Afdhal N, Reddy KR, Nelson DR, Lawitz E, Gordon SC, Schiff E, Nahass R, Ghalib R, Gitlin N, Herring R, Lalezari J. Ledipasvir and sofosbuvir for previously treated HCV genotype 1 infection. NEJM 2014;370(16):1483–93.

[3] Sulkowski MS, Gardiner DF, Rodriguez-Torres M, Reddy KR, Hassanein T, Jacobson I, Lawitz E, Lok AS, Hinestrosa F, Thuluvath PJ, Schwartz H. Daclatasvir plus sofosbuvir for previously treated or untreated chronic HCV infection. NEJM 2014;370(3):211–21.

[4] Fridell RA, Qiu D, Valera L, Wang C, Rose RE, Gao M. Distinct functions of NS5A in hepatitis C virus RNA replication uncovered by studies with the NS5A inhibitor BMS-790052. J Virol 2011;85(14):7312–20.

[5] Leroy V, Angus PW, Bronowicki JP, Dore G, Hezode C, Pianko S, Pol S, Stuart KA, Tse E, McPhee F, Bhore R. All-oral treatment with daclatasvir (DCV) plus sofosbuvir (SOF) plus ribavirin (RBV) for 12 or 16 weeks in HCV genotype (GT) 3-infected patients with advanced fibrosis or cirrhosis: the ALLY-3+ phase 3 study. Hepatology 2015:1860–75.

[6] Nelson DR, Cooper JN, Lalezari JP, Lawitz E, Pockros PJ, Gitlin N, Freilich BF, Younes ZH, Harlan W, Ghalib R, Oguchi G. All-oral 12-week treatment with daclatasvir plus sofosbuvir in patients with hepatitis C virus genotype 3 infection: ALLY-3 phase III study. Hepatology 2015;61(4):1127–35.

[7] Wang C, Jia L, O'Boyle DR, Sun JH, Rigat K, Valera L, Nower P, Huang X, Kienzle B, Roberts S, Gao M. Comparison of daclatasvir resistance barriers on NS5A from hepatitis C virus genotypes 1 to 6: implications for cross-genotype activity. Antimicrob Agents Chemother 2014;58(9):5155–63.

[8] Wyles DL. Antiviral resistance and the future landscape of hepatitis C virus infection therapy. J Infect Dis 2013;207(suppl 1):S33–9.

[9] Wang C, Valera L, Jia L, Kirk MJ, Gao M, Fridell RA. In vitro activity of daclatasvir on hepatitis C virus genotype 3 NS5A. Antimicrob Agents Chemother 2013;57(1):611–3.

[10] Gerlach JT, Diepolder HM, Jung MC, Gruener NH, Schraut WW, Zachoval R, et al. Recurrence of hepatitis C virus after loss of virus-specific CD4(+) T-cell response in acute hepatitis C. Gastroenterology 1999;117:933–41.

[11] zurWiesh JS, Lauer GM, Day CL, Kim AY, Ouchi K, Duncan JE, et al. Broad repertoire of the CD4+ Th cell response in spontaneously controlled hepatitis C virus infection includes dominant and highly promiscuous epitopes. J Immunol 2005;175:3603–13.

[12] Day CL, Lauer GM, Robbins GK, McGovern B, Wurcel AG, Gandhi RT, et al. Broad specificity of virus-specific CD4+ T-helper-cell responses in resolved hepatitis C virus infection. J Virol 2002;76:12584–95.

[13] Baumert TF, Fauvelle C, Chen DY, Lauer GM. A prophylactic hepatitis C virus vaccine: a distant peak still worth climbing. J hepatol 2014;61(1):S34–44.

[14] Pestka JM, Zeisel MB, Bläser E, Schürmann P, Bartosch B, Cosset FL, Patel AH, Meisel H, Baumert J, Viazov S, Rispeter K. Rapid induction of virus-neutralizing antibodies and viral clearance in a single-source outbreak of hepatitis C. PNAS 2007;104(14):6025–30.

[15] Osburn WO, Fisher BE, Dowd KA, Urban G, Liu L, Ray SC, Thomas DL, Cox AL. Spontaneous control of primary hepatitis C virus infection and immunity against persistent reinfection. Gastroenterology 2010;138(1):315–24.

[16] Fofana I, Fafi-Kremer S, Carolla P, Fauvelle C, Zahid MN, Turek M, et al. Mutations that alter use of hepatitis C virus cell entry factors mediate escape from neutralizing antibodies. Gastroenterology(e229) 2012;143:223–33.

[17] De Francesco R, Migliaccio G. Challenges and successes in developing new therapies for hepatitis C. Nature 2005;436(7053):953–60.

[18] Ashraf NM, Bilal M, Mahmood MS, Hussain A, Mehboob MZ. In-silico analysis of putative HCV epitopes against Pakistani human leukocyte antigen background: an approach towards development of future vaccines for Pakistani population. Infect Genet Evol 2016;43:58–66.

[19] Jiwani N, Gul RB. A silent storm: hepatitis C in Pakistan. J Pioneer Med Sci 2011;1(3):89.

[20] Klade CS, Wedemeyer H, Berg T, Hinrichsen H, Cholewinska G, Zeuzem S, Blum H, Buschle M, Jelovcan S, Buerger V, Tauber E, Frisch J, Manns MP. Therapeutic vaccination of chronic hepatitis C nonresponder patients with the peptide vaccine IC41. Gastroenterology 2008;134:1385–95.

[21] Yutani S, Komatsu N, Shichijo S, Yoshida K, Takedatsu H, Itou M, Kuromatu R, Ide T, Tanaka M, Sata M, Yamada A, Itoh K. Phase I clinical study of a peptide vaccination for hepatitis C virus-infected patients with different human leukocyte antigen-class I-A alleles. Cancer Sci 2009;100:1935–42.

[22] Yutani S, Ueshima K, Abe K, Ishiguro A, Eguchi J, Matsueda S, Komatsu N, Shichijo S, Yamada A, Itoh K, Sasada T. Phase II study of personalized peptide vaccination with both a hepatitis c virus-derived peptide and peptides from tumor-associated

antigens for the treatment of HCV-positive advanced hepatocellular carcinoma patients. J Immunol Res 2015;11:2015.
[23] Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinfo 2007;8:4.
[24] Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 1995;11:681–4.
[25] Andreatta M, Karosiene E, Rasmussen M, Stryhn A, Buus S, Nielsen M. Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. Immunogenetics 2015;67(11–12):641–50.
[26] Singh H, Raghava GPS. ProPred: prediction of HLA-DR binding sites. Bioinfo 2001;17(12):1236–7.
[27] Zhang GL, DeLuca DS, Keskin DB, Chitkushev L, Zlateva T, Lund O, Reinherz EL, Brusic V. MULTIPRED2: a computational system for large-scale identification of peptides predicted to bind to HLA supertypes and alleles. J Immunol Meth 2011;374(1):53–61.
[28] Nielsen M, Lund O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. BMC Bioinfo 2009;10:296.
[29] Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. BMC Bioinfo 2007;8:238.
[30] Agrawal S, Bhatnagar S, Bhardwaj U, Khan F, Sharma A, et al. Distribution of HLA class II antigens in three North Indian populations. Int J Hum Genet 2001;1(4):283–91.
[31] Agarwal S, Srivastava SK, Borkar M, Chaudhuri TK. Genetic affinities of north and northeastern populations of India:inference from HLA-based study. Tissue Antigens 2008;72(2):120–30.
[32] Middleton D, Menchaca L, Rood H, Komerofsky R. New allele frequency database. Tissue Antigens 2003;61(5):403–7http://www.allelefrequencies.net.
[33] Kankonkar SR, Shankarkumar U. HLA DRB1 gene study in different population groups from Mumbai, Maharashtra, India. Int J Hum Genet 2005;5(4):267–71.
[34] Kumar A, Yadav IS, Hussain S, Das BC, Bharadwaj M. Identification of immunotherapeutic epitope of E5 protein of human papillomavirus-16: an in silico approach. Biologicals 2015;43(5):344–8.
[35] Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics 2015;32(4):511–7. 29.
[36] EL-Manzalawy Y, Dobbs D, Honavar V. Predicting linear B-cell epitopes using string kernels. J Mol Recognit 2008;21:243–55.
[37] Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 2006;65(1):40–8.
[38] Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immunome Res 2006;2(1):2. 24.
[39] Singh H, Ansari HR, Raghava GPS. Improved method for linear B-Cell epitope prediction using Antigen's primary sequence. PLoS One 2013;8(5). e62216.
[40] Bui HH, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC Bioinfo 2007;8(1):361.
[41] Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234(3):779–815.
[42] Lamiable A, Thevenet P, Rey J, Vavrusa M, Derreumaux P, Tufféry P. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. Nucleic Acids Res 2016;29:29.
[43] Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS. Automated docking with selective receptor flexibility. J Comput Chem 2009;30:2785–91.
[44] Fernández MM, Guan R, Swaminathan CP, Malchiodi EL, Mariuzza RA. Crystal structure of staphylococcal enterotoxin I (SEI) in complex with a human major histocompatibility complex class II molecule. J Biol Chem 2006;281(35):25356–64.
[45] Skwarczynski M, Toth I. Peptide-based synthetic vaccines. Chem Sci 2016;7(2):842–54.
[46] Dhanda SK, Usmani SS, Agrawal P, Nagpal G, Gautam A, Raghava GP. Novel in silico tools for designing peptide-based subunit vaccines and immunotherapeutics. Brief Bioinfo 2016;25. bbw025.
[47] Kim KS, Park SA, Ko KN, Yi S, Cho YJ. Current status of human papilloma virus vaccines. Clin Exp Vaccine Res 2014;3:168–75.
[48] Tan PT, Khan AM, August JT. Highly conserved influenza A sequences as T cell epitopes-based vaccine targets to address the viral variability. Hum Vaccin 2011;7(4):402–9.
[49] Kwon SS, Kim NS, Yoo TJ. Vaccination with DNA encoding human T736 cell epitopes suppresses Der p induced allergic responses in mice. J Asthma 2005;737(42):119–25.
[50] Sheikh QM, Gatherer D, Reche PA, Flower DR. Towards the knowledge-based design of universal influenza epitope ensemble vaccines. Bioinformatics 2016;10. btw399.
[51] Alam A, Ali S, Ahamad S, Malik M, Ishrat R. From ZikV genome to vaccine: in silico approach for the epitope-based peptide vaccine against Zika virus envelope glycoprotein. Immunology 2016;149(4):386–99.
[52] Khan MK, Zaman S, Chakraborty S, Chakravorty R, Alam MM, Bhuiyan TR, et al. In silico predicted mycobacterial epitope elicits in vitro T-cell responses. Mol Immunol 2014;61:16–22.
[53] Bande F, Arshad SS, Hair Bejo M, Kadkhodaei S, Omar AR. Prediction and in silico identification of novel B-Cells and T-cells epitopes in the S1-Spike glycoprotein of M41 and CR88 (793/B) infectious bronchitis virus serotypes for application in peptide vaccines. Adv Bioinfo 2016;7:2016.
[54] Azmi F, Ahmad Fuaad AA, Skwarczynski M, Toth I. Recent progress in adjuvant discovery for peptide-based subunit vaccines. Human Vac Immunotherap 1 Mar 2014;10(3):778–96.
[55] Reed SG, Orr MT, Fox CB. Key roles of adjuvants in modern vaccines. Nat Med 1 Dec 2013;19(12):1597–608.