

RESEARCH

Open Access

# Promiscuous prediction and conservancy analysis of CTL binding epitopes of HCV 3a viral proteome from Punjab Pakistan: an *In Silico* Approach

Abida Shehzadi<sup>1\*</sup>, Shahid ur Rehman<sup>1</sup>, Muhammad Idrees<sup>2\*</sup>

## Abstract

**Background:** HCV is a positive sense RNA virus affecting approximately 180 million people world wide and about 10 million Pakistani populations. HCV genotype 3a is the major cause of infection in Pakistani population. One of the major problems of HCV infection especially in the developing countries that limits the limits the antiviral therapy is the long term treatment, high dosage and side effects. Studies of antigenic epitopes of viral sequences of a specific origin can provide an effective way to overcome the mutation rate and to determine the promiscuous binders to be used for epitope based subunit vaccine design. An *in silico* approach was applied for the analysis of entire HCV proteome of Pakistani origin, aimed to identify the viral epitopes and their conservancy in HCV genotypes 1, 2 and 3 of diverse origin.

**Results:** Immunoinformatic tools were applied for the predictive analysis of HCV 3a antigenic epitopes of Pakistani origin. All the predicted epitopes were then subjected for their conservancy analysis in HCV genotypes 1, 2 and 3 of diverse origin (worldwide). Using freely available web servers, 150 MHC II epitopes were predicted as promiscuous binders against 51 subjected alleles. E2 protein represented the 20% of all the predicted MHC II epitopes. 75.33% of the predicted MHC II epitopes were (77-100%) conserve in genotype 3; 47.33% and 40.66% in genotype 1 and 2 respectively. 69 MHC I epitopes were predicted as promiscuous binders against 47 subjected alleles. NS4b represented 26% of all the MHC I predicted epitopes. Significantly higher epitope conservancy was represented by genotype 3 i.e. 78.26% and 21.05% for genotype 1 and 2.

**Conclusions:** The study revealed comprehensive catalogue of potential HCV derived CTL epitopes from viral proteome of Pakistan origin. A considerable number of predicted epitopes were found to be conserved in different HCV genotype. However, the number of conserved epitopes in HCV genotype 3 was significantly higher in contrast to its conservancy in HCV genotype 1 and 2. Despite of the lower conservancy in genotype 1 and 2, all the predicted epitopes have important implications in diagnostics as well as CTL-based rational vaccine design, effective for most population of the world and especially the Pakistani Population.

## Background

Family Flaviviridae comprises small enveloped pathogens classified in three genera: *Flavivirus*, *Pestivirus*, and *Hepacivirus*. Members of these genera cause various diseases in humans and other animals such as birds, horses and pigs. The only genera *Flavivirus* contain more than 70 members

including Hepatitis C Virus (HCV), Dengue virus, West Nile virus and tick-borne encephalitis virus [1-3].

HCV is a positive sense RNA virus affecting approximately 180 million people world wide and rate of infection in Pakistani population is about 10 million [4,5]. HCV genome contributes about 9400 nucleotides that encode single polyprotein of approximately 3010 to 3033 amino acids in length [6]. This single polyprotein is processed by viral as well as host proteases into three structural proteins (i.e. core, E1 and E2) and four non-structural proteins (i.e. NS2, NS3, NS4, and NS5A) [7].

\* Correspondence: abida@cemb.edu.pk; idreeskhan@cemb.edu.pk  
<sup>1</sup>Bioinformatics Division, Centre of Excellence in Molecular Biology, University of the Punjab, Lahore, Pakistan  
<sup>2</sup>Division of Molecular Virology, Centre of Excellence in Molecular Biology, University of the Punjab, Lahore, Pakistan  
Full list of author information is available at the end of the article

HCV mainly spreads via blood supply, reuse of glass syringes and needles, unsterilized medical equipment, use of tooth brushes of HCV patients, etc [7] and causes of acute and chronic infections [8]. Clinical demonstrations of acute Hepatitis C Viral infection include Jaundice, Fever, Myalgia, Fatigue, Lethargy, Increased ALT, Anorexia and Fulminant hepatic failure [7]. About 80% of HCV infected individuals develop chronic infections [9]. Chronic liver infections develop chronic hepatitis, cirrhosis and hepatocellular carcinoma within a period of 10, 20 and 30 years respectively followed by viral infection [10,11]. Out of 70-80% chronically infected individuals, 20% develop cirrhosis and 1-5% individuals suffer from final stage of liver diseases [12]. Hepatic steatosis is the accumulation of lipids in hepatocytes and is reported for the cause of cirrhosis [13] with the more severe cases being reported in patients infected with HCV genotype 3a [14]. The prevalence of steatosis in Pakistani population is about 61.5-65.5% compared with 32.8-81.2% in western countries [15]. The percentage of males infected with HCV chronic liver stage is higher than females with the age of patients between 40-50 years [5].

HCV is classified into six genotypes each having various subtypes [16-18]. These genotypes are distributed differently in various parts of the world with the genetic variance between them is about one third. The genotypes 1, 2 and 3 have world wide distribution. But the significant differences are observed in subtype distribution. Subtype 1a is mostly found in North America and Europe followed by 2b and 3a. Subtype 1b is frequently found in South East Europe and Tunisia and 2c in North Italy. Genotype 4 is mainly restricted to Middle East and Central Africa and genotype 5 in South Africa. Genotype 6 is distributed throughout South East Asia and also being isolated from Hong Kong and Vietnam [17]. The most frequent HCV genotypic distribution in Pakistan is 3a [49.05%] followed by 3b [17.66%] [19]. The knowledge of HCV distribution is crucial for treatment therapy and vaccination because of its predictive value in terms of response to antiviral therapy and vaccination. Effective responses to antiviral therapy are normally associated with genotype 2 and 3 in comparison to any other genotype [20].

HCV replicates at about  $10^{12}$  new HCV viruses/day. Replication is carried out by RNA dependent RNA polymerase. RNA polymerase lacks the "proofreading" ability that ensures the high mutation rate of about 8-18 mutations in genomic RNA/year [21,20]. Such a high mutation rate limits the treatment therapy and vaccination. The current treatment therapy for HCV is INF alpha along with ribavirin limited to about 50% population [22]. Although the response rate is not much deterring, but high dosage, long-term treatment and side effects limits the usage [23,21]. There is the possibility that

after next few years, new antiviral agents such as inhibitors of the viral protease, helicase or polymerase will further improve the response rate of the current therapeutic agents. However, antiviral therapy is not affordable in most developing countries, where the prevalence of HCV is generally the highest. Thus, given the huge reservoir of HCV worldwide, the development of an effective vaccine may be the cheapest way to control disease associated with HCV infection.

Development of an effective HCV vaccine requires understanding of immune response. Viral immune response is associated with Major Histocompatibility complex protein (MHC) and T lymphocytes/T cell. MHC are classified into 2 broad categories, MHC I and MHC II [24]. MHC initially recognizes the viral antigenic epitopes and presents to T lymphocytes for degradation. MHC I presents the antigenic epitopes to CD8+ T cells and MHC II presents to CD4+ T cells for viral degradation [25,26]. CD8 T cells also referred to as cytotoxic T cells (CTL or Tc), limit viral infections by initial recognizing and their subsequent killing infected cells and secreting cytokines. CD4 T referred to as helper cells or Th cells and provides growth factors and signals for generation and maintenance of CD8 T cells [27]. T cells recognize the antigens only when they are associated with MHC, surface glycoprotein exposed on surface of all vertebrate cells. The selection of T cell epitopes is also important because these are linear and hence easy to synthesize.

A particular vaccine developed against HCV can't be effective for Pakistani population because of variations in HCV genomic sequences and distribution with regard to geographical area. Since a large number of Pakistani population is infected by HCV3a and number of patients enrolled in public and hospitals is increasing day by day. So there is a current need to develop a vaccine against HCV in particular to HCV3a that will cover approximately maximum Pakistani population. The current vaccines are DNA vaccine, Peptide vaccine and epitopic vaccines. Epitopes are the small antigenic segments of viral proteins and causes infections in the host. Epitopic vaccines provide more potent and controlled immune response and eliminates the potential lethal effects of the use of whole viral proteins [28]. Promiscuous epitopes (epitopes capable of binding maximum number of HLA alleles) may overcome the population coverage. Secondly the conserved epitopes reduces antigen escape associated with the viral mutation [29]. So the present study was designed for the prediction of promiscuous epitopes and to analyze their conservancy in general population. Any mutation in the peptide/epitope will lower the conservancy, so it was hypothesized to analyze the pI value of the mutated amino acid residue, that if remain in the range as was in original epitope provides the likeliness of that particular epitope to be used for epitopic vaccine

design having an effective control over viral mutation, immune response with minimum side effects.

## Methods

### Sequence Retrieval and Analysis

The sequence of fully sequenced HCV 3a genome and protein of Pakistani origin was retrieved from NCBI [GU294484]. The number of individual bases in the genome i.e. the number of adenine; cytosine, guanine and thymine were calculated from DDBJ database. The molecular weight of proteins, percentage of highly repeated amino acid and the least repeated amino acid in the viral protein was calculated by using sequence and search analysis tool at PIR database (<http://pir.georgetown.edu/>).

### Epitope Prediction

Promiscuous epitopes of HCV 3a viral proteins were predicted for HLA I and HLA II binding alleles using freely available immunoinformatics tools such as ProPred I, and ProPred respectively. In comparison to other epitope prediction tools, Propred 1 and Propred cover maximum number of Human Leukocyte antigens i.e. HLA and being used for epitopic prediction for HBV and tuberculosis. ProPred1 allows the user to predict antigenic epitopes for 47 MHC I alleles and ProPred allows epitopes prediction for 51 MHC II alleles. Predictions through these tools can be carried out at various thresholds from 1 to 10%. The algorithms designed for the working of these tools are based on linear coefficients of matrices. Maximum of the matrices were retrieved from BIMASS where the score of each peptide is calculated in multiplication and/or sum up manner. For example the score of following peptide "PACDP-GRAA" can be calculated by using following equation:

$$\text{Score} = P(1) \times A(2) \times C(3) \times D(4) \times P(5) \times G(6) \times R(7) \times A(8) \times A(9)$$

$$\text{Score} = P(1) + A(2) + C(3) + D(4) + P(5) + G(6) + R(7) + A(8) + A(9)$$

Where P (1) is score of P at position 1.

Only the promiscuous epitopes with score higher than the chosen threshold score were assigned as predicted epitopes for the selected HLA alleles [30]. For the following study the default threshold i.e. 4% was used where the sensitivity and specificity are nearly the same for most of the HLA alleles available in ProPred1 and ProPred server. Moreover, MHC I alleles were predicted by keeping the proteosome and immunoproteosome filters on at 5% threshold because most of the MHC binders having a proteosomal cleavage site at C-terminal have higher likelihood to be T-cell epitopes [31]. The predicted promiscuous epitopes were positioned in the table in a decreasing order of their score.

### Epitope Conservancy Analysis

All the predicted epitopes of HCV 3a proteins of Pakistani origin were subjected for worldwide conservancy analysis among HCV genotype 1, 2 and 3. 5 sequences against each HCV protein (used for epitope prediction) were retrieved from NCBI randomly. The predicted epitopes of HCV 3a (Pakistani origin) along with 5 selected sequences of individual genotypes (genotype 1, 2 and 3; one at a time) were submitted to epitope conservancy analysis tool available at IEDB database ([http://tools.immuneepitope.org/tools/conservancy/iedb\\_input](http://tools.immuneepitope.org/tools/conservancy/iedb_input)). All the epitopes having 77-100% conservancy were selected while rejecting the epitopes having variation at the anchor residues. The anchor residues in the predicted epitopes were highlighted by making it bold. The epitopes that were 100% conserved in the selected proteins of the 3 viral genotypes 1, 2 and 3 were also fully bold. Epitopes with 88/77% conservancy were with single or double amino acid variation respectively and to highlight them bold format was used in the conservancy column against each genotype.

Asteric sign (\*) indicates that one out of five selected sequences either does not respond to epitope conservancy or have conservancy lower than 77%. Double asteric sign (\*\*) indicates that only one sequence responds for 77-100% conservancy to the selected epitope.

### Validation of varied amino acids using pI value

The Peptides with single or double amino acid variation were analyzed for their hydrophatic characteristics or pI value [32]. The pI gives the information that the varied residue retained the amino acid group or diverted from its normal group in a particular peptide under consideration and thus provides information to be used or their rejection. All the varied amino acid residue with diverted group (with considerable change of pI value) were separated from other using superscript "D" for single variation and "DD" for diverted group for doubly varied residues. The superscript "D" in doubly varied residues of particular peptides represents the partial variation i.e. one of the varied residue retained the amino acid group while other residue shifted the amino acid group by a considerable change of pI value.

## Results

HCV 3a genome of Pakistani origin comprises 9474 bp with GC content 2622 and 2700 respectively. The GC contents are 12.35% higher than AT contents. The genome encodes a polyprotein that subsequently get fragmented into structural and non structural protein of obvious molecular weight. The envelope protein E2 comprises highest molecular weight 38755.3 KDa (Table 1).

**Table 1 It comprises the data of HCV genome size, Proteins, Molecular weight and %age of highly repeated and least repeated amino acid residues in individual bases**

Bases	No.	Proteins	aa Number	Mol. Wt.	Highly repeated aa	% of repetition	Least repeated aa	% of repetition
Total bp	9474	Capsid	114	12985.8	R	18.4	C/F	0.9
A	1974	Core	75	7638.88	L	16	E/K/M/Y	1.3
C	2700	E1	190	20643.9	V	11.1	E	1.1
G	2622	E2	350	38755.3	L	13.1	K	1.4
T	2178	NS3	149	15423.6	A/G	11.4	N	0.7
		NS4a	54	5751.69	L/V	14.8	H/M/T/W	1.9
		NS4b	194	20167.5	A	13.4	C	0.4
		NS5a-1a	62	6700.72	G	14.5	E/D	1.6
		NS5a-1b	101	11224.6	P	11.9	K/W	1

Leucine (L) a neutral nonpolar amino acid residue has the highest percent of repetition (13.1%) in E2 protein. The least repeated residue of E2 is a basic polar Lysine (K) (1.4%). The shortest segment viral protein is NS4a (5751.69 KDa molecular weight) comprising 54 amino acid residues. Leucine (L) and Valine (V) have highest percentage of repetition (14.8) and Histidine (H), Methionine (M), Threonine (T) and Tryptophan (W) are the least repeated amino acid residues (1.9%). The molecular weight of other viral proteins and percent repetition of their amino acid residue for were listed in Table 1. The percentage of amino acid residues gives an out look for their pI value and their probability of incidence in the antigenic epitopes.

F (Phenylalanine), I (Isoleucine), L (Leucine), M (Methionine), V (Valine), W (Tryptophan) and Y (Tyrosine) were mainly the anchor residues for MHC II predicted epitopes and are nonpolar in nature. Total 150 epitopes were predicted against 51 alleles of MHC II (Table 2). The highest number of epitopes was represented by E2 protein comprising 20% of all MHC II predicted epitopes. VFLLNPCGL, FVILVFLLL, WHINSTVLH, FNLLDVPKA, LELINTHGS, VQYLYGVGS are the promiscuous binders of 45-50 MHC II alleles. E2 is followed by NS2 and NS4B proteins representing 14.66% of the predicted MHC II epitopes. In case of NS2 VRAHVLVRL, VILLTSLLY and VRLCMFVRS are the best binders both in term of score and the HLA allele coverage (50-51 MHC II alleles). FFNILGGWV, VNLLPAILS and VVNLLPAIL are the best binders of NS4b protein both in terms of HLA coverage (41 HLA coverage for the first epitope and 51 for the next 2 epitopes) and binding efficiency. LVVGVICAA, FNILGGWVA, WQKLEAFWH, IQY-LAGLST and VVGVICAAL are also the epitopes of good quality covering 31 to 35 HLA alleles available in ProPred. For the NS5a\_1a only three epitopes (MRLAGPRTC, FISCQKGYK and VVSTRCPG) were predicted as promiscuous binders with the binding score higher than the selected threshold. Out of these three epitopes

MRLAGPRTC is capable of binding all the HLA alleles available in ProPred server while FISCQKGYK and VVSTRCPG bind 22 and 25 HLA alleles respectively. The predicted promiscuous binders against other proteins were also summarized in table 2.

Total 69 epitopes were predicted as promiscuous epitopes for MHC I alleles. The anchor residues in case of MHC I are quite varying both in amino acid residues and also in their nature. Mostly represented anchor residues are neutral nonpolar and neutral polar. However, quite small percentage of anchor residues were also acidic polar and basic polar in nature. The highest number of MHC I binding epitopes were represented by NS4b protein comprising 26% of all MHC I predicted epitopes. NFVSGIQYL epitope of NS4b is the best promiscuous binder of highest binding score. NS4b is followed by NS2, E2 and NS3 proteins representing 20.28% (NS2 epitopes) and 11.59% (for E2 and NS3). In case of NS2, 14 promiscuous epitopes were predicted with varying binding efficiency. GSRDGVILL, DGVILLTSL, WAAAGLKDL and LQVWVPLL are the good binders both in term of score and the HLA allele coverage (21, 28, 27 and 28 alleles respectively). E2 predicted epitopes covers 20 to 28 HLA alleles except the PLLHSTTEL epitope that covers only 11 HLA alleles but with highest binding efficiency. NS3 epitopes covers 8 to 25 HLA alleles and were also ranked on the basis of their binding efficiency predicted by the score. The least represented epitopes were by NS5a\_1a. It comprises only one epitope (HVKNQSMRL) as predicted promiscuous binders for 16 MHC I binding alleles. The promiscuous binders of MHC I for other proteins were also predicted and summarized in table 3.

Out of total 150 predicted MHC II epitopes, 75.33% were (77-100%) conserve in genotype 3 (Table 1) against the randomly selected viral proteins. Out of 75.33% conserved peptides of genotype 3, 71.68% peptides were 100% conserve while 22.12% peptides were having single residue variation (88% epitope conservancy). Only the

**Table 2 Predicted HLA II epitopes HCV Proteins of Pakistani origin and their conservancy in Genotype 1, 2 and 3 worldwide**

Epitope start Position	Predicted T-cell epitopes	HLA alleles	HCV genotype 1	HCV Genotype 2	HCV Genotype 3
<b>Capsid</b>					
43	LGVRATRKA	23	LGVRATRKT <sup>D</sup>	LGVRATRKT	LGVRATRKT <sup>D</sup>
36	LPRRGPRLG	15	<b>LPRRGPRLG</b>	<b>LPRRGPRLG</b>	<b>LPRRGPRLG</b>
106	WGPNDPRRR	16	WGPTDPRRR	WGPTDPRHR <sup>D</sup>	WGPNDPRRR
34	YVLPRRGPR	24	YLLPRRGPR	YLLPRRGPR	YVLPRRGPR
21	VKFPGGGQI	8	<b>VKFPGGGQI *</b>	<b>VKFPGGGQI</b>	<b>VKFPGGGQI</b>
35	VLPRRGPRL	9			VLPRRGPRL
45	VRATRKAASE	25	VRATRKTSE <sup>D</sup>	VRATRKTSE <sup>D</sup>	VRATRKTSE <sup>D</sup>
30	VGGVYVLP	39	VGGVYLLPR *	VGGVYLLPR	VGGVYVLP
15	IRRPQDVKF	6			IRRPQDVKF
95	WLLSPRGSR	28	<b>WLLSPRGSR</b>	<b>WLLSPRGSR</b>	<b>WLLSPRGSR</b>
29	IVGGVYVLP	3	IVGGVYLLP *	IVGGVYLLP	IVGGVYVLP
82	WPLYGNEGC	10	<b>WPLYGNEGC</b>	<b>WPLYGNEGC</b>	<b>WPLYGNEGC</b>
85	YEGNECGWA	11	<b>YEGNECGWA</b>	<b>YEGNECGWA *</b>	<b>YEGNECGWA</b>
33	VYVLPRRGP	1	VYVLPRRGP	VYVLPRRGP	VYVLPRRGP
<b>Core</b>					
61	FLLALLSCL	50	FLLALLSCL	FLLALLSCI	FLLALLSCL
64	LALLSCLIH	45	LALLSCLTV <sup>DD</sup>		LALLSCLIH *
15	FADLMGYIP	41	<b>FADLMGYIP</b>	<b>FADLMGYIP</b>	<b>FADLMGYIP</b>
24	LVGAPVGGV	44	LVGAPLGGA		LVGAPVGGV *
63	LLALLSCLI	36	LLALLSCLT <sup>D</sup>	LLALLSCIT <sup>D</sup>	LLALLSCLI *
62	FLLALLSCL	24	FLLALLSCL	FLLALLSCI	FLLALLSCL *
32	VARALAHGV	10		VARALAHGV	VARALAHGV
21	YIPLVGAPV	28	YIPLVGAPL	YIPVVGAPL	YIPLVGAPV
19	MGYIPLVGA	26	MGYIPLVGA	MGYIPVGA	MGYIPLVGA
<b>E1</b>					
58	YVGATTASI	41			YVGATTASI *
140	MVAHILRL	39			MVAHILRL*
2	WRNTSGLYV	27			WRNTSGLYV
138	VGMVAHIL	28			
56	VKYVGATTA	21			VRKYVGATTA <sup>D</sup> *
9	YVLTNARSN	31			YVLTNDCSN <sup>DD</sup>
161	WGVLAGLAY	15	WGVLAGMAY	WGVVFGLAY	WGVLAGLAY
93	FLVGQAFTF	11	FLVGQLFTF		FLVGQAFTF
181	IIMVMFSGV	91			IIMVMFSGV
130	MMMNWSPAV	35	MMMNWSPTA <sup>D</sup>		MMMNWSPAM
134	WSPAVGMV	6			WSPAMGMV *
132	MNWSAVGM	14			MNWSAVGM *
169	YYTMQGNWA	18			YYSMQGNWA
47	WTPMTPTVA	21			WTPVPTVA *
172	MQGNWAKVA	25	MVGNWAKVL <sup>D</sup>	MQGAWAKVI <sup>D</sup>	WTPVPTVA <sup>D</sup> *
145	ILRLPQTLF	19			ILRLPQTLF
<b>E2</b>					
122	MLPHHRPW	3			
151	VFLLNPCGL	48			
337	WEFVILVFL	4			WEFVILVFL

**Table 2 Predicted HLA II epitopes HCV Proteins of Pakistani origin and their conservancy in Genotype 1, 2 and 3 worldwide (Continued)**

339	FVILVFLLL	46			FVILVFLLL
35	WHINSTVLH	41			
342	LVFLLLADA		LLFLLLADA	LLFLLLADA	LVFLLLADA
100	VLLAYAPRP	50			
198	FRPLLPURL	47			
218	VRLGALVDT	12			
62	FNLLDVPKA	45			
26	LELINTHGS	46			
57	FYYHKFNLL	12	FYYHKFNSS <sup>D</sup>		FYYHKFNST <sup>DD</sup>
83	VGPLDRCQH	26			
58	YYHKFNLLD	24			
286	LLHSTTELA	17	LLHSTTEWA		LLHSTTELA
129	VWGTDPK	14	VWGTTDKL <sup>DD</sup>	VWGTTDRL <sup>DD</sup> *	VWGTDAK
320	VQYLYGVGS	46	VQYLYGVGS		VQYLYGVGS
159	LLWGGLGG	14			
293	LAILPCSFT	7			LAILPCSFT
335	LKWEFVILV	4			LKWEFVILV
322	YLYGVGSGM	5	YLYGVGSSI <sup>D</sup>		YLYGVGSGM
300	FTPMPALST	17			FTPMPALST
245	FYTVQGEDV	4			
18	IVRGPEQRL	26			
100	VLLAYAPRP	4			
257	VWHRFTAAC	19	VEHRLTAAC <sup>D</sup> *		
206	LLQETSRGH	8			
1	YITGGTAAR	8			
267	WTRGERCDI	10			WTRGERCEI
310	IHLHQNIVD	11	IHLHQNIVD *		IHLHQNIVD
<b>NS2</b>					
101	VRAHVLVRL	51			VRAHVLVRL
62	VILLTSLLY	50			VILLTSLLY *
73	LVFDIAKLL	24	LVFDITKLL <sup>D</sup> *	LVFDITKLL <sup>D</sup> *	LIFDITKLL <sup>D</sup>
153	LKDLAVATE	7			LKDLAVATE *
113	FVRSVTGGK	37			
130	VGRWFNTYL	11			VGRWFNTYL *
123	FQMAILSVG	31			FQMILHVG <sup>D</sup>
137	YLYDHLAPM	21			YLYDHLAPM
74	VFDIAKLLIA	23	VFDITKLLLA <sup>D</sup> *	VFDITKLLLA <sup>D</sup>	
107	LVRLCMFVR	36			LVRLCMLVR
108	VRLCMFVRS	51			VRLCMLVRS *
89	YFVRAHVLV	33			YFVRAHVLV
11	ILVLFGFFT	15			
37	YAICRCESA	18		IINGLPVSA <sup>D</sup>	YTICRCESA <sup>D</sup> *
33	WWNQYAICR	8			WWNQYTICR <sup>D</sup>
185	ILCGLPVSA	10	IINGLPVSA *	IINGLPVSA	ILCGLPVSA
145	MQHWAAAGL	18			MQHWAAAGL
50	VPPLLARGS	21			VPSLLARGS <sup>D</sup> *
88	LYLIQAAIT	35			LYLIQTAIT <sup>D</sup> *

**Table 2 Predicted HLA II epitopes HCV Proteins of Pakistani origin and their conservancy in Genotype 1, 2 and 3 worldwide (Continued)**

158	VATEPVIFS	14	VAVEPWF <sup>D</sup>	VAVEPVFS	VATEPVIFS
37	YAICRCESA	19			YTICRCESA <sup>D</sup> *
175	WGADTAACG	11	WGADTAACG *	WGADTAACG *	WGADTAACG
<b>NS3</b>					
4	VQVLSTATQ	46	VQIVSTATQ	VQVLS <sup>SVTQ</sup> <sup>D</sup>	
43	LQMYTNVDQ	42			
129	VCTRGVAKA	21	VCTRGVAKA	VCARGVAKS <sup>DD</sup> *	
24	WTVYHGAGS	13	WTVYHGAGT	WTVYHGAGN	
84	VIPARRRGD	18	VIPVRRRGD *		
138	LQFIPVETL	45			
140	FIPVETLST	43	FIPVENLGT <sup>D</sup>		
6	VLSTATQTF	19	IVSTATQTF		
53	LVGWAPPAG	29	LVGWAPQAG	LVGWSP <sup>PPG</sup> <sup>D</sup>	
27	YHGAGSRTL	22	YHGAGRTI		
14	FLGTTLGGV	10			
77	LVTREADVI	25	LVTRHADVI <sup>D</sup> *	LVTRNADVI <sup>D</sup> *	
98	LSPRPLACL	12		LSPRPLSTL <sup>D</sup>	
124	IFRAAVCTR	44			
<b>NS4a</b>					
23	VVIVGHIEL	43	VVIVGRIL <sup>DD</sup>	VVIVGRIV <sup>DD</sup>	VVIVGHIEL
3	WVLLGGVLAA	43	WVLVGGVLAA	WVLVGGVLAA	WVLLGGVLAA
4	VLLGGVLAAL	40	VLVGGVLAAL	VLVGGVLAAL	VLLGGVLAAL
38	VPDKEVLYQ	11			VPDKEVLYQ *
24	VIVGHIELG	8			VIVGHIELG
10	LAALAAYCLS	8	LAALAAYCLT *	LAALAAYCLS	LAALAAYCLS
16	YCLSVGCW	6		YCLSTGCW <sup>D</sup>	YCLSVGCW
26	VGHIELGGK	9			VGHIELGGK
25	IVGHIELGG	29			IVGHIELGG
20	VGCWIVGH	15			VGCWIVGH
9	VLAALAAYC	9	VLAALAAYC*	VLAALAAYC	VLAALAAYC
29	IELGGKPAL	14			IELGGKPAL
<b>NS4b</b>					
81	FFNILGGWV	41			FFNILGGWV
153	VNLLPAILS	51	VNLLPAILS	VNLLPAILS	VNLLPAILS
152	VVNLLPAIL	51			
39	WNFVSGIQY	16	WNFISGIQY	WNFISGIQY	WNFVSGIQY
165	LWGVICAA	35	LWGVVCAA	LWGVVCAA	LWGVICAA
82	FNILGGWVA	32	FNILGGWVA	FNILGGWVA	FNILGGWVA
81	FFNILGGWV	5			FFNILGGWV
63	LMAFAASVT	9	LMAFTASIT <sup>D</sup>	LMAFTA <sup>AVT</sup> <sup>DD</sup>	LMAFTASVT <sup>D</sup>
27	WQKLEAFWH	35		WQKLEVFWA <sup>D</sup>	WQKLEAFWH *
167	VGVICAALL	11	VGVC <sup>CAIL</sup>	VGVC <sup>CAIL</sup>	VGVICAAIL
45	IQYLAGLST	35	IQYLAGLST	IQYLAGLST	IQYLAGLST
64	MAFAASVTS	23	MAFTASITS	MAFTA <sup>AVTS</sup> <sup>DD</sup>	MAFTASVTS <sup>D</sup>
84	ILGGWATH	24	ILGGWAAQ <sup>DD</sup>	ILGGWAAQ <sup>DD</sup>	ILGGWATH
103	VVSLAGAA	10		VGAGLAGAA <sup>D</sup>	VVSLAGAA
166	VGVICAAL	31	VGVVCAAI	VGVVCAAI	VGVICAAIL

**Table 2 Predicted HLA II epitopes HCV Proteins of Pakistani origin and their conservancy in Genotype 1, 2 and 3 worldwide (Continued)**

85	LGGWVATHL	3	LGGWVAAQL <sup>DD</sup>	LGGWVAAQL <sup>DD</sup>	LGGWVATHL
60	VASLMAFAA	15			VASLMAFTA <sup>D</sup>
41	FVSGIQYLA	8	FISGIQYLA	FISGIQYLA	FVSGIQYLA
139	FKIMGGELP	21		FKIMSGEVP <sup>D</sup>	FKIMGGEFP *
9	LQRATQQQA	14			LQRATQQQA *
122	LDILAGYGA	6			LDILAGYGA *
104	VSLAGAAI	3			VSLAGAAI
<b>NS5a_1a</b>					
39	MRLAGPRTC	51	MRIVGPRTC *	FISCQKGYR <sup>D</sup> *	MRLAGPRTC*
3	FISCQKGYK	22	FFSCQRGYK <sup>DD</sup> *		FISCQKGYK *
19	VWSTRCPG	25			VMSTRCPG *
<b>NS5a_1b</b>					
73	LLRDEITFV	20	LLRDEVTFQ <sup>D</sup> *	LLRDEVTFQ <sup>D</sup> **	LLRDEITFV *
16	WRVAANSYV	33	WRVAAEEYV <sup>DD</sup> *	WRVAASEYV <sup>D</sup>	WRVAANSYV
55	FTEVDGVR	4	FTELDGVR*	FTEVDGVR **	FTEVDGVR
80	FWGLNSYA	25			FWGLNSYA *
32	FHYITGATE	16			FHYITGATE
61	VRLHRYAPP	27	VRLHRYAPA*		VRLHRYAPP *
87	YAIGSQLPC	20	YVVGSQLPC *	VRLHRYAPA <sup>D</sup> **	YAIGSQLPC *
23	YVEVRRVGD	14	YVEVTRVGD <sup>D</sup> *	YVEVTRVGD <sup>D</sup> **	YVEVRRVGD

Bold amino acid residues in T-cell Epitope column indicates the anchor residues.

Bold individual amino acid residues in HCV Genotype 1, 2 and 3 columns indicated the variation in peptide in comparison to the predicted epitope.

\*Indicates that one of the protein sequence selected for epitope conservancy either does not respond or have conservancy lower than 70%.

\*\* Indicates that only one of the protein sequence from selected sequences respond to epitope conservancy.

<sup>D</sup> Indicates that amino acid residue in case of single/double variation diverted their group compared to primary epitope using pl value.

<sup>DD</sup> Indicates that both amino acid residues in case of double variation diverted their group compared to primary epitope using pl value.

**Table 3 Predicted HLA I epitopes HCV Proteins of Pakistani origin and their conservancy in Genotype 1, 2 and 3 worldwide**

Epitope start Position	Predicted T-cell epitopes	HLA alleles	HCV genotype 1	HCV Genotype 2	HCV Genotype 3
<b>Capsid</b>					
38	RRGPRLGVR	9	RRGPRLGVR	RRGPRLGVR	RRGPRLGVR
35	VLPRRGPR	25			VLPRRGPR
<b>Core</b>					
55	PGCSFSIFL	8	PGCSFSIFL	PGCSFSIFL	PGCSFSIFL *
41	RALEDGINF	20		RVLEDGVNF **	RALEDGINF *
7	VIDTLTCGF	15	VIDTLTCGF	VIDTLTCGF *	VIDTLTCGF *
35	ALAHGVRAL	24	ALAHGVRVL	ALAHGVRVL	ALAHGVRAL *
24	LVGAPVGGV	18	LVGAPLGGGA *		LVGAPVGGV *
26	GAPVGGVAR	9	GAPLGGAAAR *	GAPLGGVAR	GAPVGGVAR *
<b>E1</b>					
135	SPAVGMWA	14			SPAMGMWA *
86	GDVCGAVFL	19	GDLCCGSVFL <sup>D</sup>	GDVCGAVMI	GDMCGAVFL *
144	HILRLPQTL	22			HILRLPQTL *
156	IAGAHWVGL	27	IAGAHWVGL		IAGAHWVGL
64	ASIRGHVDL	25			ASIRSHVDL <sup>D</sup>
<b>E2</b>					
285	PLLHSTTEL	11		PLLHSTTEW <sup>D</sup>	PLLHSTTEL



**Table 3 Predicted HLA I epitopes HCV Proteins of Pakistani origin and their conservancy in Genotype 1, 2 and 3 worldwide (Continued)**

305	ALSTGLIHL	25	ALT <b>T</b> GLIHL		ALSTGLIHL
227	CSFTMPAL	20	CSFT <b>T</b> L <b>P</b> AL <sup>D</sup> *		CSFTMPAL
71	QQLQAHHFL	27			
157	CGLLVGGGL	28			
212	RGHIQPVRL	24			
6	TAARGGQGL	25			
157	CGLLVGGGL	28			
<b>NS2</b>					
172	VITWGADTA	6			VITWGADTA *
75	FDIAKLLIA	12	FDIT <b>K</b> LLIA <sup>D</sup>	FDIT <b>K</b> LLIA <sup>D</sup> *	FDIT <b>K</b> LLIA <sup>D</sup>
70	YPSLVFDIA	15			YPSLIFDIT <sup>DD</sup>
57	GSRDGVILL	21	GGRDA <b>V</b> ILL <sup>D</sup> **	GGRDA <b>V</b> ILL <sup>D</sup> *	GSRDGVILL
60	DGVILLTSL	26			DGVILLTSL *
148	WAAAGLKDL	27		WAA <b>S</b> GLRDL <sup>DD**</sup>	WAAAGLKDL *
50	VPPLLARGS	11			VPPLLARGS *
46	LQVWPPLL	28	LHW <b>V</b> WPPLN <sup>DD</sup>	LHW <b>V</b> WPPLN <sup>DD</sup> *	LQVWPPLL *
117	VTGGKYFQM	16			V <b>V</b> GGKYFQM <sup>D</sup> *
65	LTSLLYPSL	23			LTSLLYPSL
6	TLGAGILVL	48			TLGAG <b>V</b> LVL *
73	LVFDIAKLL	31	LVFDIT <b>K</b> LL <sup>D</sup>	LVFDIT <b>K</b> LL <sup>D</sup> *	LIFDIT <b>K</b> LL <sup>D</sup>
145	MQHWAAAGL	26			MQHWAAAGL
178	DTAACGDIL	21	DTAACGD <b>I</b>	DTAACGD <b>I</b> <sup>D</sup> *	DTAACGDIL
<b>NS3</b>					
119	GHVAGIFRA	8	GH <b>A</b> VAGIFRA *	GH <b>V</b> GLFRA *	
27	YHGAGSRTL	14	YHGAG <b>R</b> TI	YHGAG <b>N</b> KTLD <sup>D</sup>	
128	AVCTRGVAK	8	AVCTRGVAK	AVCTRGVAK *	
57	PAPPGAKSL	11	PAP <b>Q</b> GARSL <sup>DD</sup> *	P <b>S</b> PPG <b>T</b> KSL <sup>DD</sup>	
98	LSPRPLACL	25		LSPR <b>P</b> STL <sup>D</sup>	
95	ASLLSPRPL	24			
130	CTRGVAKAL	20	CTRGVAK <b>V</b>		
7	LSTATQTFL	24			
<b>NS4a</b>					
3	WVLLGGVLA	11	WV <b>V</b> GGVLA	WV <b>V</b> GGVLA	WVLLGGVLA
23	WVIVGHIEL	21	WVIV <b>R</b> IIL <sup>DD</sup>	WVIV <b>R</b> IIL <sup>DD</sup>	WVIVGHIEL
10	LAALAAYCL	24	LAALAAY <b>C</b> L *	LAALAAYCL	LAALAAYCL
5	LLGGVLAAL	27	LVGGVLAAL	LVGGVLAAL	LLGGVLAAL
<b>NS4b</b>					
96	PQSSSAFW	6			PQSSSAFW
40	NFVSGIQYL	30	NFVSGIQYL	NFVSGIQYL	NFVSGIQYL
81	FFNILGGWV	13			FFNILGGWV
46	QYLAGLSTL	17	QYLAGLSTL	QYLAGLSTL	QYLAGLSTL
102	FVVSGLAGA	9		FV <b>G</b> AGLAGA <sup>D</sup>	FVVSGLAGA
54	LPGNPAVAS	14	LPGNPAIAS	LPGNPAIAS	LPGNPAVAS
161	SPGALWGV	14	SPGAL <b>V</b> GV	SPGAL <b>V</b> GV	SPGAL <b>V</b> GV
141	IMGGELPNA	7			IMGGE <b>F</b> PTA <sup>D</sup> *
164	ALWGVICA	11	ALWGV <b>C</b> A	ALWGV <b>C</b> A	ALWGVICA

**Table 3 Predicted HLA I epitopes HCV Proteins of Pakistani origin and their conservancy in Genotype 1, 2 and 3 worldwide (Continued)**

117	LGRVLLDIL	22	LGKVLVDIL <sup>D*</sup>	LGKVLVDIL <sup>D</sup>	LGKVLVDIL <sup>D*</sup>
59	AVASLMAFA	9	AIASLMAFT <sup>D</sup>	AIASLMAFT <sup>D</sup>	AVASLMAFT <sup>D</sup>
152	VVNLPAIL	15			
113	GIGLGRVLL	24			GIGLGVLL <sup>D*</sup>
56	GNPAVASLM	12	GNPAIASLM	GNPAIASLM	GNPAVASLM
52	STLPGNPAV	21	STLPGNPAI	STLPGNPAV	STLPGNPAV
85	LGGWVATHL	26	LGGWVAAQL <sup>DD</sup>	LGGWVAAQL <sup>DD</sup>	LGGWVATHL
145	ELPNAEDVW	11			
99	SSAFVWSGL	22			SSAFVWSGL
<b>NS5a_1a</b>					
33	HVKNGSMRL	16	HVKNGSMRI *	HVKNGSMRI **	HVKNGSMRL
<b>NS5a_1b</b>					
49	VPAAEFFTE	6	VPAAEFFTE *	VPAAEFFTE **	VPAAEFFTE
79	TFWGLNSY	10	TFQVGLNQY <sup>D*</sup>	TFTVGLNSF <sup>D*</sup>	TFTVGLNSY <sup>D*</sup>
76	DEITFWGL	19	DEVTFQVGL <sup>D*</sup>	DEVTFVGL <sup>D*</sup>	DEITFMVGL *

Bold amino acid residues in T-cell Epitope column indicates the anchor residues.

Bold individual amino acid residues in HCV Genotype 1, 2 and 3 columns indicated the variation in peptide in comparison to the predicted epitope.

\*Indicates that one of the protein sequence selected for epitope conservancy either does not respond or have conservancy lower than 70%.

\*\* Indicates that only one of the protein sequence from selected sequences respond to epitope conservancy.

<sup>D</sup> Indicates that amino acid residue in case of single/double variation diverted their group compared to primary epitope using pI value.

<sup>DD</sup> Indicates that both amino acid residues in case of double variation diverted their group compared to primary epitope using pI value.

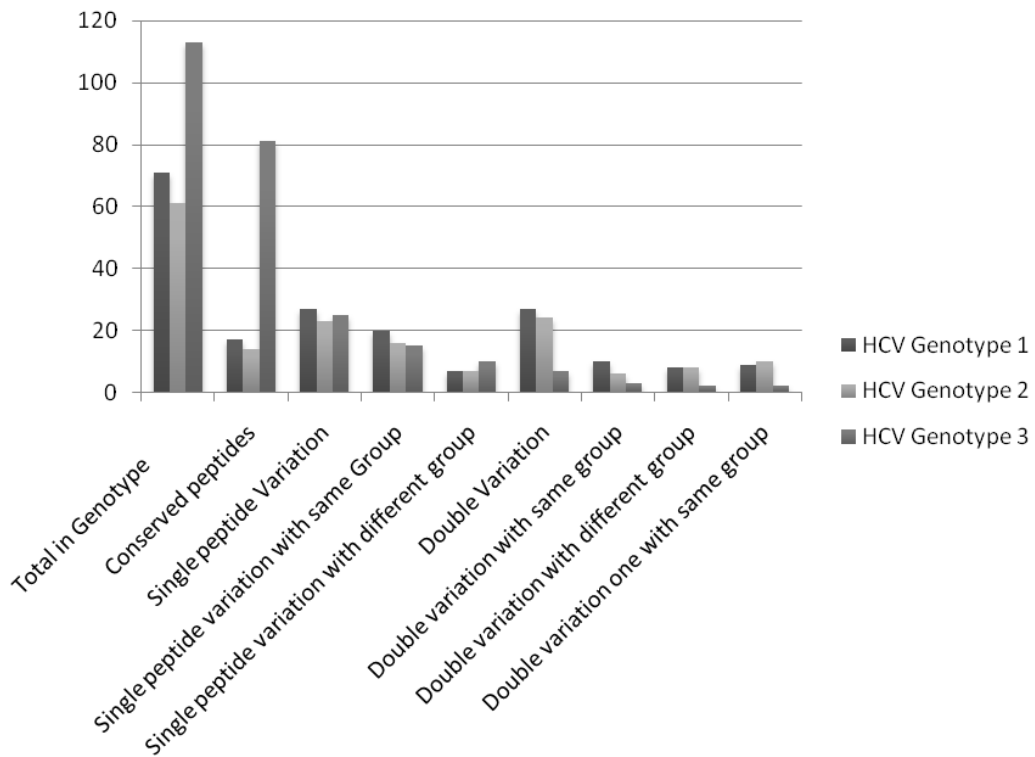
40% peptides of singly varied residues diverted their amino acid group and the pI value while 60% singly varied residues retained the amino acid group as was in the predicted epitope of HCV 3a proteins. 6.19% peptides comprised the 77% epitope conservancy because of double residue variation in the peptides of general population in contrast to predicted epitopes of HCV 3a of Pakistani origin. Out of 6.19%, doubly varied amino acid residues 42.85% peptides retained their amino acid group and nearly same pI value as in case of predicted epitope while 28.57% peptides were having partial group diversion and 28.57% (of doubly varied amino acid residues) peptides diverted their amino acid group because of considerable variation in the pI value. Similar data was also obtained for the HCV genotype 1 and 2 consisting 47.33% and 40.66% conservancy respectively. However, in contrast to genotype 3, only 23.94% predicted epitopes were 100% conserve in randomly selected sequences of genotype 1 and 22.95% in genotype 2. Their rate of single/double residue variation was also predicted and expressed as figure 1.

Out of total 69 predicted MHC I epitopes, 78.26% were (77-100%) conserve in genotype 3 (Table 2) against the randomly selected viral proteins. Out of 78.26% conserved peptides of genotype 3, 72.22% peptides were 100% conserve while 22.22% peptides were having single residue variation (88% epitope conservancy). 40.66% peptides of singly varied residues retained the amino

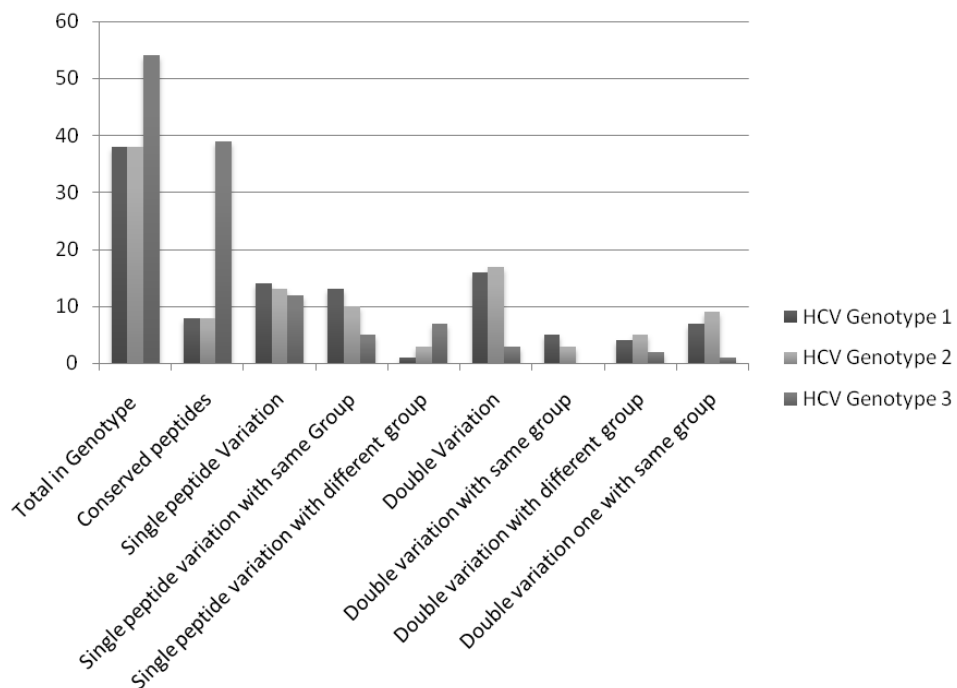
acid group as was in the predicted epitope of HCV 3a proteins while 58.33% singly varied residues diverted their amino acid group and the pI value. 5.5% peptides comprised the 77% epitope conservancy because of double residue variation in the peptides of general population in contrast to predicted epitopes of HCV 3a of Pakistani origin. Out of 5.5%, doubly varied amino acid residues 66.66% peptides were having partial group diversion and 33.33% (of doubly varied amino acid residues) peptides diverted their amino acid group because of considerable variation in the pI value. Similar data was also obtained for the HCV genotype 1 and 2 consisting 55.07% conservancy. However, in contrast to genotype 3, only 21.05% predicted epitopes were 100% conserve in randomly selected sequences of genotype 1 and 2. Their rate of single/double residue variation was also predicted and expressed as figure 2.

## Discussion

The modern technique for control of HCV infection is a vaccine preparation that can specifically induce antibody-mediated immunity. The rapid advancements in the computational methodologies and immunoinformatics/immuno-bioinformatics provide new strategies for the synthesis of antigen specific epitopic vaccine against infectious agents such as viruses and pathogens. Epitopic vaccine against HIV, malaria and tuberculosis provided promising results and supported the defensive



**Figure 1** A comparative analysis of HCV 3a Predictive epitopes against MHC II alleles and their conservancy analysis in Genotype 1, 2 and 3 worldwide.



**Figure 2** A comparative analysis of HCV 3a Predictive epitopes predicted against MHC I and their conservancy analysis in Genotype 1, 2 and 3 worldwide.

and therapeutic uses of these vaccines [33]. Thus in the present study, a new systematic immunoinformatics approach was applied for the predicted antigenic epitopes of HCV 3a proteins of Pakistani origin followed by diversity and conservancy in other genotypes (1,2 and 3) in randomly selected HCV sequences from NCBI and mainly belong to Thailand, Cuba, UK, USA, China, Japan, France, Italy and Germany. The immunogenic epitopes identified were nanomers and could be used diagnostically to detect HCV specific CTL responses in the patients and after vaccination. A CTL based HCV vaccine might not efficient enough to prevent from infection but it might protect the body from the disease. The analysis showed that the minimal number of epitopes required to represent the complete antigenicity of the whole proteins are significantly smaller than required to represent full length proteins. The majority of the epitopes reported here had intermediate to high HLA binding affinity.

By the use of an efficient CTL based epitope delivery technology; the predicted epitopes could eventually become vaccines in their own or fused as polytopes. The design of the HCV vaccine using conserved epitopes can avoid viral mutation and thus provides more efficient results. The study shows that the predicted epitopes were highly conserved in HCV genotype 3 and also but less conserved in genotype 1 and 2 both for MHC I and MHC II. Moreover, to ensure the viral detection at all stages of its intracellular evolution we have used all the viral proteins. Therefore, the total number of predicted epitopes were also maximized in correspond to the number of covered proteins used for the analysis.

#### Abbreviations

HCV: hepatitis C virus; HLA: human leukocyte antigen; MHC: major histocompatibility complex; CTL: cytotoxic T lymphocytes.

#### Author details

<sup>1</sup>Bioinformatics Division, Centre of Excellence in Molecular Biology, University of the Punjab, Lahore, Pakistan. <sup>2</sup>Division of Molecular Virology, Centre of Excellence in Molecular Biology, University of the Punjab, Lahore, Pakistan.

#### Authors' contributions

AS and SH designed the study. AS performed the immunoinformatics analysis and drafted the manuscript. MI critically reviewed the manuscript. All authors have read and approved the final manuscript.

#### Competing interests

The authors declare that they have no competing interests.

Received: 25 January 2011 Accepted: 8 February 2011

Published: 8 February 2011

#### References

1. Qi R, Zhang L, Chi C: Biological characteristics of dengue virus and potential targets for drug design. *Acta Biochim Biophys Sin* 2008, **40**:91-101.

2. Blitvich JB, Fernandez-Salas I, Contreras-Cordero FJ, Marlenee LN, Gonzalez-Rojas IJ, Komar N, Gubler JD, Calisher HC, Beaty JB: Serologic Evidence of West Nile Virus Infection in Horses, Coahuila State, Mexico. *Emerging Infectious Diseases* 2003, **9**(7):853-856.
3. Moennig V: Introduction to classical swine fever: virus, disease and control policy. *Veterinary Microbiology* 2000, **73**:93-102.
4. Stiffler DJ, Nguyen M, Sohn AJ, Liu C, Kaplan D, Seeger C: Focal Distribution of Hepatitis C Virus RNA in Infected Livers. *PLoS ONE* 2009, **4**(8):1-7.
5. Raja SN, Janjua AK: Epidemiology of hepatitis C virus infection in Pakistan. *J Microbiol Immunol Infect* 2008, **41**:4-8.
6. Raja SN, Singh NN, Janjua AK, Najam-Ul-Haq R: Hepatitis C virus infection: An enigma continues. *Medicine Today* 2006, **4**(3):93-101.
7. Das RB, Kundu B, Khandapkar R, Sahni S: Geographical distribution of hepatitis C virus genotype in India. *Indian J Pathol Microbiol* 2002, **45**(3):323-328.
8. Rychłowska M, Bieńkowska-Szewczyk K: Hepatitis C- new developments in the studies of the viral life cycle. *Acta Biochimica Polonica* 2007, **54**(4):703-715.
9. Jobarteh Modou, Malfroy Marine, Peterson Ingrid, Jeng Adam, Ramu Sarge-Njie, Alabi Abraham, Peterson Kevin, Cotten Matt, Hall Andrew, Sarah Rowland-Jones, Whittle Hilton, Tedder Richard, Jaye Assan, Mendy Maimuna: Seroprevalence of hepatitis B and C virus in HIV-1 and HIV-2 infected Gambians. *Virology Journal* 2010, **7**:230-239.
10. Re DV, Caggiari L, Vita DS, Mazzaro C, Lenzi M, Galli M, Monti G, Ferri C, Zignego LA, Gabrielli A, Sansonno D, Dammacco F, Libra M, Sacchi N, Talamini R, Spina M, Cannizzaro R, Guidoboni M, Dolcetti R: Genetic insights into the disease mechanisms of type II mixed cryoglobulinemia induced by hepatitis C virus. *Digestive and Liver Disease* 2007, **39**(1):65-71.
11. Wazir I, Wazir F, Javed M, Saeed M, Najeib-ul-Haq Khan H: Effect of vitamin "e" supplements in therapy of chronic hepatitis c: a histological study. *Gomal Journal of Medical Sciences* 2008, **6**(2):81-86.
12. Jia Y, Wei L, Jiang D, Wang J, Cong X, Fei R: Antiviral action of interferon- $\alpha$  against hepatitis C virus replicon and its modification by interferon- $\gamma$  and interleukin-8. *J Gastroenterology and Hepatology* 2007, **22**:1278-1285.
13. Yoon JE, Hu K: Hepatitis C Virus (HCV) Infection and Hepatic Steatosis. *Int J Med Sci* 2006, **3**(2):53-56.
14. Minakari M, Sameni KF, Shalmani MH, Molaee M, Zali M: Hepatic Steatosis in Iranian Patients with Chronic Hepatitis C. *Med Princ Pract* 2008, **17**:126-130.
15. Zubair A, Jamal S, Mubarak A: Morphometric Analysis of Hepatic Steatosis in Chronic Hepatitis C Infection. *The Saudi Journal of Gastroenterology* 2009, **15**(1):11-4.
16. Ali Amjad, Habib Ahmad, Ali Ijaz, Sheema Khan, Gulshan Zaidi, Muhammad Idrees: Prevalence of active hepatitis c virus infection in district mansehra Pakistan. *Virology Journal* 2010, **7**:334-338.
17. Ramia S, Eid-Fares J: Distribution of hepatitis C virus genotypes in the Middle East. *International Journal of Infectious Diseases* 2006, **10**:272-277.
18. Ali Amjad, Habib Ahmed, Muhammad Idrees: Molecular epidemiology of Hepatitis C virus genotypes in Khyber Pakhtoonkhaw of Pakistan. *Virology Journal* 2010, **7**:203-209.
19. Idrees M, Riazuddin S: Frequency distribution of hepatitis C virus genotypes in different geographical regions of Pakistan and their possible routes of transmission. *BMC Infectious Diseases* 2008, **8**(69):1-9, 69.
20. Lauer G, Walker B: Hepatitis C Virus Infection. *N Engl J Med* 2001, **345**(1):41-52.
21. Jawaid A, Khuwaja KA: Treatment and Vaccination for Hepatitis C: Present and Future. *J Ayub Med Coll* 2008, **20**(1):129-133, Abbottabad.
22. Pan Q, Henry DS, Metselaar JH, Scholte B, Kwekkeboom J, Tilanus WH, Janssen ALH, Laan WJL: Combined antiviral activity of interferon- $\alpha$  and RNA interference directed against hepatitis without affecting vector delivery and gene silencing. *J Mol Med* 2009, **87**:713-722.
23. Gozlan J, Lacombe K, Gault E, Raguin G, Girard P: Complete cure of HBV-HDV co-infection after 24 weeks of combination therapy with pegylated interferon and ribavirin in a patient co-infected with HBV/HCV/HDV/HIVq. *Journal of Hepatology* 2009, **50**:432-434.
24. Tan Lei, Lu Huijun, Dan Zhang, Mingyao Tian, Hu Bo, Wang Zhuoyue, Jin Ningyi: Protection against H1N1 influenza challenge by a DNA vaccine expressing H3/H1 subtype hemagglutinin combined with MHC class II-restricted epitopes. *Virology Journal* 2010, **7**:363-376.

25. Reche AP, Reinherz LE: **PEPVAC: a web server for multi-epitope vaccine development based on the prediction of supertypic MHC ligands.** *Nucleic Acids Research* 2005, **33**:138-142.
26. Zhang LG, Khan MA, Srinivasan NK, August TJ, Brusic V: **MULTIPRED: a computational system for prediction of promiscuous HLA binding peptides.** *Nucleic Acids Research* 2005, **33**:172-179.
27. Robinson LH, Amara RR: **T cell vaccines for microbial infections.** *Nature Medicine Supplement* 2005, **11**(4):25-32.
28. Groot DSA, Sbai H, Aubin SC, Mcmurry J, Martin W: **Immuno-informatics: Mining genomes for vaccine components.** *Immunology and Cell Biology* 2002, **80**:255-269.
29. Lund O, Nielsen M, Kesmir C, Petersen GA, Lundegaard C, Worning P, Sylvester-Hvid C, Lamberth K, Røder G, Justesen S, Buus S, Brunak S: **Definition of supertypes for HLA molecules using clustering of specificity matrices.** *Immunogenetics* 2004, **55**(12):797-810.
30. Singh H, Raghava SPG: **ProPred1: prediction of promiscuous MHC Class-I binding sites.** *Bioinformatics* 2003, **19**(8):1009-1014.
31. Somvanshi P, Singh V, Seth KP: **In Silico Prediction of Epitopes in Virulence Proteins of Mycobacterium Tuberculosis H37Rv for Diagnostic and Subunit Vaccine Design.** *Journal of Proteomics & Bioinformatics* 2008, **1**(3):143-153.
32. Kyte J, Doolittle FR: **A simple method for displaying the hydrophobic character of a protein.** *Journal of Molecular Biology* 1982, **157**(1):105-132.
33. Khan MA, Miotto O, Heiny TA, Salmon J, Srinivasan NK, Nascimento MJE, Marques JETA, Brusic V, Tan WT, August TJ: **A systematic bioinformatics approach for selection of epitope-based vaccine targets.** *Cellular Immunology* 2006, **244**:141-147.

doi:10.1186/1743-422X-8-55

**Cite this article as:** Shehzadi *et al.*: 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**:55.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

