

RESEARCH

Open Access

# Structural analysis and epitope prediction of HCV E1 protein isolated in Pakistan: an in-silico approach

Sobia Idrees and Usman A Ashfaq\*

## Abstract

**Background:** HCV infection is a major health problem causing acute and chronic hepatitis. HCV E1 protein is a transmembrane protein that is involved in viral attachment and therefore, can serve as an important target for vaccine development. Consequently, this study was designed to analyze the HCV E1 protein sequence isolated in Pakistan to find potential conserved epitopes/antigenic determinants.

**Results:** HCV E1 protein isolated in Pakistan was analyzed using various bio-informatics and immuno-informatics tools including sequence and structure tools. A total of four antigenic B cell epitopes, 5 MHC class I binding peptides and 5 MHC class II binding peptides were predicted. Best designed epitopes were subjected to conservation analyses with other countries.

**Conclusion:** The study was conducted to predict antigenic determinants/epitopes of HCV E1 protein of genotype 3a along with the 3D protein modeling. The study revealed potential B-cell and T-cell epitopes that can raise the desired immune response against HCV E1 protein isolated in Pakistan. Conservation analysis can be helpful in developing effective vaccines against HCV and thus limiting threats of HCV infection in Pakistan.

**Keywords:** Hepatitis C Virus, T cell epitope, B cell epitope, 3D structure prediction, Vaccine

## Introduction

Hepatitis C Virus infection is a global health problem affecting 270 million people worldwide [1]. According to the World Health Organization, liver cancer by HCV caused approximately 308,000 annual deaths in 2004 [2]. The number of HCV infected individuals is increasing day by day, and there is variability in the prevalence reports of HCV in Pakistan but according to majority of studies, HCV is prevalent among 2.4-6.5% adults and among 0.44-1.6% of children [3]. From the prevalence analysis, clearly HCV genotype 3a is most common in Pakistan [4].

HCV is an RNA virus like dengue virus, West Nile virus and yellow fever virus belonging to the Flaviviridae family [5] and has a 9.5 kb genome with a positive-single stranded RNA that encodes a large polyprotein which is cleaved to produce four structural (Core, E1, E2 and P7)

and six non-structural proteins (NS2, NS3, NS4A, NS4B, NS5A, NS5B). These viral proteins are liable for viral replication and various cellular functions [5-8]. Among HCV structural proteins, envelope proteins play the primary role in viral entry. HCV envelope protein 1 (E1) is a transmembrane glycoprotein having a C-terminal domain responsible for membrane association and membrane permeability changes [9]. E1 acts as a fusogenic subunit of the HCV envelope and contains 4-5 N-linked glycans. As it is known that the interaction of the virion with various cell receptors results in HCV infection [10,11]. Therefore, it is important to target virus envelope proteins to stop viral entry. Although there is not much knowledge available about E1, but it is thought to be involved in intra-cytoplasmic virus-membrane fusion. Currently, the standard of care is pegylated interferon (PEG-INF) with ribavirin; this therapy gives 50% sustained virological response in genotype 1 and 80% for genotype 2 and 3 [12,13]. One of the top priorities in HCV infection should be the development of more effective therapies by developing antiviral compounds for infected patients.

\* Correspondence: [usmancemb@gmail.com](mailto:usmancemb@gmail.com)  
Department of Bioinformatics and Biotechnology, Government College University (GCU), Faisalabad, Pakistan

For designing effective inhibitors against envelope proteins, it is important to have knowledge of the epitopic regions/antigenic determinants of these glycoproteins. Bioinformatics analysis has opened new vistas to provide more insights into protein sequence and structural features. Both B-cell and T-cell epitopes/antigenic determinants are important in raising desired immune responses and the number of epitopes and modulation of immune recognition of antigens can be influenced by deglycosylation of viral glycoproteins [14]. This study was designed to perform immunoinformatic analysis on the HCV E1 glycoprotein isolated in Pakistan and to analyze antigenicity, hydrophobicity, surface accessibility and epitopic location of epitopes in HCV glycoprotein structure.

## Methods

### Protein retrieval and comparative modeling

The HCV E1 protein sequence was retrieved from NCBI protein database using the ID: ACN92051. It was ascertained that the three-dimensional structure of the protein was not available in Protein Data Bank (PDB). Therefore, the present study was designed to predict the 3D model and to predict epitopes of HCV E1 proteins isolated in Pakistan. Primary structure analysis was performed using the ProtParam online tool. The parameters computed by ProtParam [15] included the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) and secondary structure analysis was done using various online servers. Structure template with PDB ID 2VOV\_A having 43% identity was selected for the E1 protein. This template was used as a reference to determine the 3D structures of E1. Protein Structure Prediction Server (PS)<sup>2</sup> [16] predicted the homology model based on a package MODELLER. Moreover, Glycosylation sites of HCV E1 of Pakistani origin were found and their conservation with other regions of the world was also checked through Multiple Sequence Alignment. For this purpose, HCV E1 protein sequences isolated in different countries were retrieved from the NCBI protein database.

### Stereochemical analysis and model evaluation

Once the 3D model was generated, the Swiss-PdbViewer energy minimization test was applied to check for energy criteria in comparison with the potential of mean force derived from a large set of known protein structures. Structural evaluation and stereochemical analyses were performed using different evaluation and validation tools. Backbone conformation was evaluated by analyzing the Psi/Phi Ramachandran plot obtained from PROCHECK analysis. The Ramachandran plot of the phi/psi distribution in the model is developed using PROCHECK [17] for checking non-GLY residues at the disallowed regions. The

Z-score is indicative of overall model quality and is used to check whether the input structure is within the range of scores typically found in native proteins of similar size. The Z-score was determined by PROSA web tool [18]. The model was further evaluated through ERRAT [19]. Furthermore, visualization of the generated model was performed using UCSF Chimera 1.5.3. The model generated for protein was successfully submitted to the Protein model database (PMDB) having PMID PM0078432.

### T-cell epitope and B-cell epitope prediction

A systemic strategy was adapted to design potential T-cell and B-cell epitopes of HCV envelope protein. VaxiJen v2.0 online antigen prediction server was used for analyzing the antigenicity of the E1 protein [20]. Transmembrane topology of protein was checked using TMHMM [21]. B-cell epitopes were predicted using the BCPREDS online server using 75% of specific criteria for epitope prediction. All the predicted B-cell epitopes were checked from whether they were present in transmembrane regions or not using TMHMM results, and epitopes exposed on the surface of the membrane were selected and were subjected to further analysis. Antigenicity of selected epitopes were again checked using the Vexijen online server. DiscoTope server predicts discontinuous B-cell epitopes from protein three-dimensional structures. Disco Top 2.0 Server [22] was employed for discontinuous B-cell prediction using 3D structure of the HCV E1 protein of Pakistan. Furthermore, T-cell epitopes were screened. For this, Propred-1 which predicts epitopes for 47 MHC Class-I alleles and Propred, which predicts epitopes for 51 MHC Class-II alleles were utilized. Both servers cover a maximum number of HLA (Human Leukocyte antigens), therefore, are considered acceptable for predicting epitopes. Proteasome and immunoproteasome filters were set to a 5% threshold for MHC class I alleles. MHC binders that have proteosomal cleavage site at the C - terminal have greater chances to be T-cell epitopes [23].

### Epitope conservation analysis

Sequences of HCV E1 protein belonging to different regions of the world were retrieved from the NCBI database. A consensus sequence was drawn for each country, and all the consensus sequences were subjected to multiple sequence alignment using CLC workbench (data not shown). All the selected epitopes were checked for their conservation and variability by analyzing the multiple sequence alignment results and with the IEDB conservation analysis tool.

## Results

### Structural description of the model

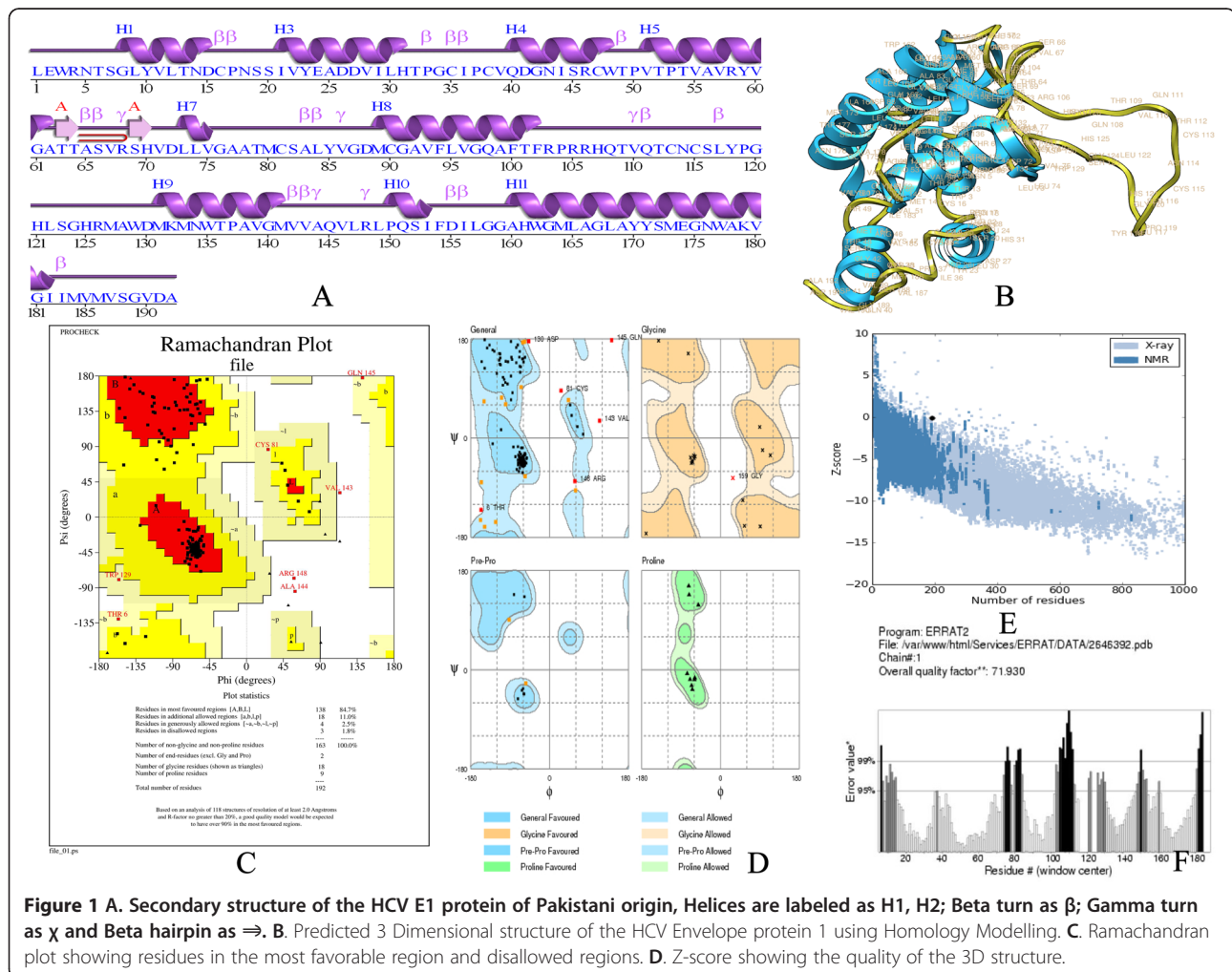
The present study was initiated to perform structure based sequence analysis studies on the HCV E1 protein

isolated in Pakistan. The protein sequence was retrieved using accession #: ACN92051 from the NCBI protein database. Primary structure analysis showed that the E1 protein had a molecular weight of 20830.1 Daltons and theoretical isoelectric point (PI) of 6.62. An isoelectric point below 7 indicates a negatively charged protein. The instability index (II) is computed to be 21.17. This classifies the protein as stable. The N-terminus of the sequence is considered to be L (Leu). The negative Grand average of hydropathicity (GRAVY) of 0.316 indicated that the protein was hydrophobic. Valine (V), Glycine (G), Alanine (A) and Leucine (L) were found in rich amounts in the protein. Secondary structure revealed that it had 34.9% alpha helices, 8.8%, beta turns, 23.96% extended strand and 32.81% coils (Figure 1A).

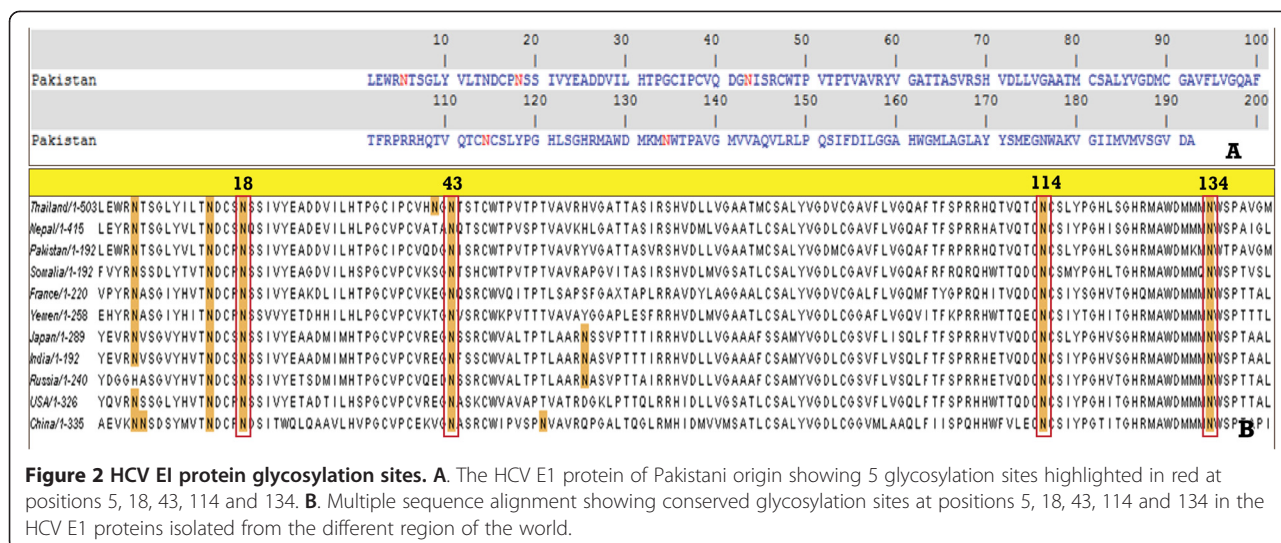
Protein 3D structure is very important in understanding the protein interactions, functions and their localization [24]. Homology modeling is the most common structure prediction method. To perform the homology modeling, the first and basic step is to find a best

matching template using similarity searching programs like PSI BLAST against a PDB database. Templates are selected based on their sequence similarity with query sequence. PDB ID 2VOV\_A was selected for homology modeling, which is an X-ray diffraction structure of the Rev-erb Beta with resolution of 1.35 Å. Both template and target protein sequences were used to predict the 3D structure of the target protein using Protein Structure Prediction Server (PS) <sup>2</sup> (Figure 1B).

The 3D structure of the protein showed that it had 49 hydrogen bonds. Quality and reliability of structure were checked by several structure assessment methods, including Z-score, ERRAT and Ramachandram plots. Procheck checks the stereochemical quality of a protein structure by analyzing residue-by-residue geometry and overall structure geometry. This tool was used to determine the Ramachandran plot to assure the quality of the model. The result of the Ramachandran plot showed 84.5% of residues in the favorable region (Figure 1C, 1D). The Z-score is indicative of overall model quality and is



**Figure 1 A. Secondary structure of the HCV E1 protein of Pakistani origin, Helices are labeled as H1, H2; Beta turn as  $\beta$ ; Gamma turn as  $\gamma$  and Beta hairpin as  $\Rightarrow$ .** B. Predicted 3 Dimensional structure of the HCV Envelope protein 1 using Homology Modelling. C. Ramachandran plot showing residues in the most favorable region and disallowed regions. D. Z-score showing the quality of the 3D structure.



used to check whether the input structure is within the range of scores typically found in native proteins of similar size. PROSA web was used to find the Z-score of the predicted structure. The Z - score of the protein was -0.11 (Figure 1E). Reliability of the model was further checked by ERRAT, which analyzes the statistics of non-bonded interactions between different atom types and plots the value of the error function versus position of a 9-residue sliding window, calculated by a comparison with statistics from highly refined structures. Results from ERRAT showed 71.930 overall model quality (Figure 1F). The Z-scores, Ramachandran plot and ERRAT results confirmed the quality of the homology model of the HCV E1 protein.

### Glycosylation site analysis

N-glycosylation sites were searched in the HCV E1 protein sequence using criteria as Asn-X-Ser or Asn-X-three sequences, where X is any amino acid residue. Four glycosylation sites were found at position 5, 18, 43, 114 and 134 (Figure 2A). To find a conserved glycosylation site in an HCV E1 protein of other countries, a multiple alignment using CLUSTALW was performed, and it was found that all glycosylation sites at position 5, 18, 43, 114 and 134 were conserved in E1 proteins of different countries as well as with Pakistan (Figure 2B).

**Table 1 Predicted B-cell epitopes**

Position	Epitope	Score
49	TPVTPVAVRYV	0.994
32	TPGCIPCVDGN	0.965
13	TNDCPNSSIVYE	0.87
153	IFDILGGAHWGM	0.702
96	VGQAFTFRPRRH	0.538
123	SGHRMAWDMKMN	0.292

### Epitope prediction

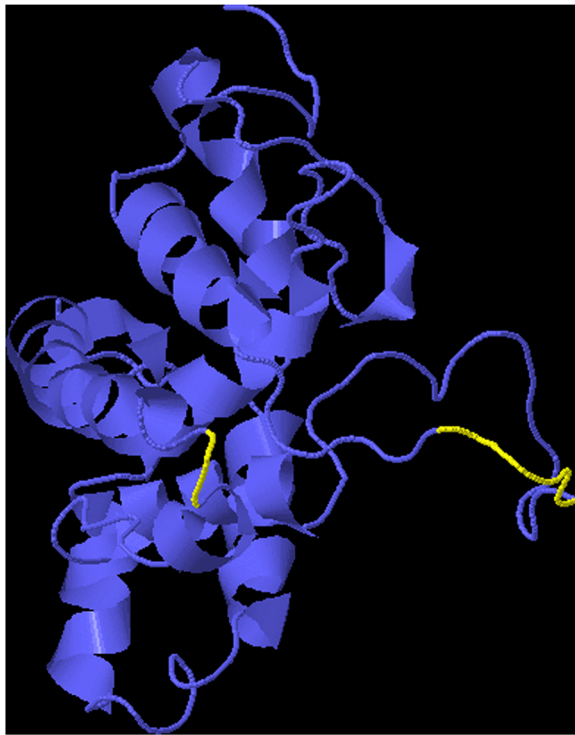
Overall antigenicity of E1 protein was predicted to be 0.5362 indicating it as a probable antigen. Transmembrane protein topology was checked using the TMHMM online tool, and was found that residues 1-73 presented outside while residues 74-96 were within the transmembrane region, and residues 97-192 were inside the core region of the protein.

### B-cell epitope prediction

B-cell epitopes are important for protection against virus infection. B-cell epitope prediction was performed using BCPRED server where criteria were set to have 75% specificity and 12 aa epitope length [25]. A total of six B-cell epitopes were predicted using a BCPRED server (Table 1). After checking the TMHMM results, it was found that epitope VGQAFTFRPRRH, with 0.538 BCPred score was in the transmembrane region while epitopes TPVTPVAVRYV, TPGCIPCVDGN, TNDCPNSSIVYE with 0.994, 0.965 and 0.87 scores, respectively, were exposed outside. Antigenicity of VGQAFTFRPRRH epitope was

**Table 2 Discontinuous epitopes predicted from the 3D structure of the E1 protein using DiscoTop online server**

Position	Residue	Contact number	Propensity score	DiscoTope score
109	THR	3	-2.545	-2.598
110	VAL	7	-2.183	-2.737
111	GLN	3	-1.235	-1.438
112	THR	1	-1.964	-1.853
113	CYS	1	-3.211	-2.957
114	ASN	9	-2.677	-3.404
155	ASP	1	-2.341	-2.187
156	ILE	1	-1.626	-1.554



**Figure 3 Predicted B-cell epitopic regions of the E1 protein 3D structure.** B-cell epitopic regions are shown in yellow color.

3D structure of the protein (Table 2). B-cells epitopes are shown in yellow color in the 3D structure of the E1 protein Figure 3.

#### T-cell epitope prediction

Propred-I (47 MHC Class-I alleles) and Propred (51 MHC Class-II alleles) were used to predict T-cell epitopes for the HCV E1 protein. ProPred1 is an online web tool for the prediction of peptides binding to MHC class-I alleles. The HCV E1 sequence was uploaded to the Propred server while selecting all the alleles, with a high scoring peptide threshold of 4%, and showing the top four epitopes in the tabular form along with proteasome and immunoproteasome filters. All the predicted epitopes were checked for their antigenicity and epitopes that were found to be antigenic in nature were used for further analysis (Table 3). Epitope MNWTPAVGM at position 154 was found to have the highest antigenicity among all epitopes assuring maximum binding affinity. The HCV E1 sequence was also used to predict MHC class II binding regions using the Propred online server (Table 4). Epitope YVGATTASV at position 30 was found to have the highest antigenicity ensuring maximum binding affinity. The HCV E1 protein structure with an epitope selected is shown in (Figure 4).

found to be 0.8539 and antigenicity of exo-membrane epitopes were 1.1421 for TPVTPTVAVRYV, 0.9738 for TPGCIPCVDGN indicating these epitopes as probable antigens while the antigenic score of TNDCPNSSIVYE was 0.2295 indicating it as a non-antigen, thereby, resulting in its exclusion. From the results, it can be inferred that these epitopes/antigenic determinants are important in raising the desired immune response. Moreover, the 3D structure of E1 was used to predict conformational discontinuous B-cells epitopes using the Disco Top 2.0 online server. A total of 8 B-cell epitopic locations were found from the

#### Epitope conservation and variability analysis

Moreover, the conservation of all predicted epitopes was checked by analyzing and comparing all the epitope sequences of the HCV E1 protein with E1 of other regions of the world. E1 sequences used in this study were from Somalia (AAF44733.1), Nepal (BAA04038.1), Canada (ABI23143.1), China (AAK95634.1), Japan (BAD06555.1), France (CAJ45644.1), India (AAG09116.1), Russia (CAD44972.1), USA (AAD21251.1) and Yemen (BAA07778.1) and were used for comparative studies through multiple alignment using ClustalW followed by verification with IEDB epitope conservation analysis

**Table 3 MHC class I binding peptides on the basis of antigenicity**

Starting position	Peptide	Allele	Antigenic score
83	ATTASVRSH	HLA-A1/HLA-A*1101/HLA-A3/HLA-A*3101/HLA-A*3302/HLA-B*5801	0.9061
154	MNWTPAVGM	HLA-A2/HLA-A*0201/HLA-A*3101/HLA-A20/HLA-B*2705/HLA-B*3501/HLA-B*5201/HLA-B*5301/HLA-B*5401/HLA-B*51/HLA-B7/HLA-B*0702/HLA-B8/HLA-Cw*0401/MHC-Ld	1.1593
160	VGMWAQVL	HLA-A*0205/HLA-A24/HLA-A20/HLA-A2.1/HLA-B14/HLA-B*2702/HLA-B*2705/HLA-B*3701/HLA-B*3801/HLA-B*3901/HLA-B*3902/HLA-B*4403/HLA-B*5101/HLA-B*5102/HLA-B*5103/HLA-B*5201/HLA-B*5301/HLA-B*51/HLA-B60/HLA-B62/HLA-B7/HLA-B8/HLA-Cw*0301/HLA-Cw*0401/HLA-Cw*0602/HLA-Cw*0702/MHC-Db/MHC-Dd/MHC-Kb/MHC-Kd/MHC-Kk/MHC-Ld	0.5535
178	LGGAHWGML	HLA-A*0205/HLA-A24/HLA-A2.1/HLA-B14/HLA-B*2705/HLA-B*3701/HLA-B*3901/HLA-B*3902/HLA-B*5101/HLA-B*5102/HLA-B*5103/HLA-B*5201/HLA-B*51/HLA-B60/HLA-B62/HLA-B7/HLA-B*0702/HLA-Cw*0301/HLA-Cw*0401/HLA-Cw*0602/HLA-Cw*0702/MHC-Db/MHC-Dd/MHC-Kb/MHC-Kd	0.6771
90	SHVDLLVGA	HLA-A*3302/HLA-B*3801/HLA-B*3901/HLA-B*5401/HLA-Cw*0702	0.9645

**Table 4 MHC class II binding peptides on the basis of antigenicity**

Starting position	Peptide	Allele	Antigenic score
12	LTNDCPNSS	DRB1_0305-309, DRB1_0311, DRB1_0401, DRB1_0421, DRB1_0426, DRB1_1107	0.4554
19	WTPVTPTVA	DRB1_0101, DRB1_408	0.7528
28	VRYVGATTASV	DRB1_0101, DRB1_0305, DRB1_0309, DRB1_0402, DRB1_0404, DRB1_0405, DRB1_0408, DRB1_0410, DRB1_0423, DRB1_0813, DRB1_1107	0.5239
30	YVGATTASV	DRB1_0101, DRB1_0305, DRB1_0309, DRB1_0401, DRB1_0402, DRB1_0404, DRB1_0405, DRB1_0408, DRB1_0410, DRB1_0421, DRB1_0423, DRB1_0426, DRB1_0701, DRB1_0703, DRB1_0801, DRB1_0802, DRB1_0813, DRB1_1101, DRB1_1114, DRB1_1120, DRB1_1128, DRB1_1302, DRB1_1305, DRB1_1307, DRB1_1321, DRB1_1323	1.0175
28	VRYVGATTA	DRB1-0102, DRB1-0306-0308, DRB1_0311, DRB1_1104, DRB1_1106, DRB1_1107, DRB1_1311, DRB1_1501, DRB1_1506, DRB5_0101, DRB5_0105	0.4463

resource [26]. Conservation analysis of epitopes showed conserved and variable residues of epitopes in the E1 sequences of other countries, and it was found that most of the predicted epitopes were conserved with the E1 sequence of Canada while having some conservation with other countries as well (Table 5).

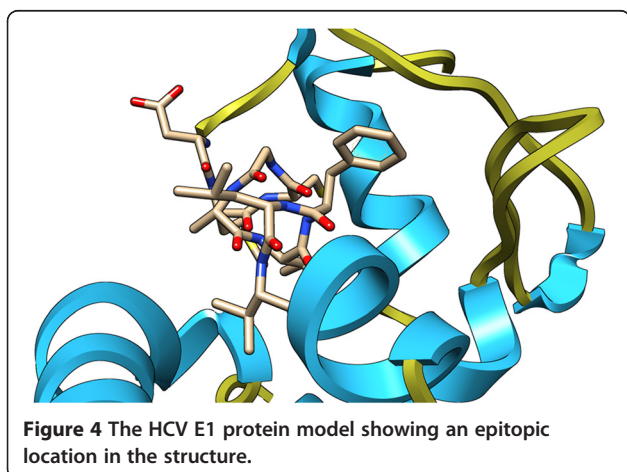
### Discussion

In this study, sequence and structure analysis, homology modeling and epitope analysis was performed on the HCV E1 protein isolated in Pakistan. We have used various sequence and structure analysis tools that helped in understanding of the sequence and its structure. Through primary structure analysis, amino acid composition of the HCV E1 glycoprotein was checked, and it showed that it has maximum Valine (V) residues and its N-terminus is a Leucine (L).

We used a homology modeling approach to predict the 3D structure of the HCV E1 protein of Pakistan. The predicted 3D structure will provide more insight into understanding the structure and function of this protein. Moreover, this structure can be used for drug designing or understanding the interactions between proteins. The HCV E1 protein was molecularly characterized using

various online servers, and it was observed that it had five glycosylation sites, and all of them were conserved in HCV E1 protein sequences of other countries. Clustal W multiple sequence alignment was used to determine the conservation and variability of HCV E1 protein belonging to different regions of the world, and it was determined that there were frequent variations at position 6 (Threonine), 11 (Valine), 17 (Proline), 32 (Threonine), 36 (Isoleucine), 40 (Glutamine), 41 (Aspartic Acid), 44 (Isoleucine), 45 (Serine), 46 (Arginine), 50 (Proline), 58 (Arginine), 59 (Tyrosine), 62 (Alanine), 67 (Valine), 77 (Alanine), 89 (Methionine), 96 (Valine), 103 (Arginine), 116 (Serine), 123 (Serine), 132 (Lysine), 136 (Threonine), 144 (Alanine), 145 (Glutamine), 152 (Serine), 153 (Isoleucine), 157 (Leucine), 158 (Glutamine), 164 (Methionine), 174 (Glutamic Acid), 181 (Glutamine), 182 (Isoleucine), 185 (Valine), 187 (Valine) in the HCV E1 protein sequences. All other residues of the HCV E1 protein were conserved in all sequences.

As a part of the present study, we predicted B-cell and T-cell epitopes of the HCV E1 protein using different online tools. Only four B-cell epitopes were found to be antigenically effective, and it can be inferred that these epitopes/antigenic determinants are important in raising the desired immune response. Using 3D structure of the E1 protein, eight B-cell epitopic locations were identified. All the predicted B-cell epitopes were checked for their localization in the protein structure, and it was found that the majority of predicted epitopes were in the outside region of the protein. T-cell epitopes were predicted using Propred I and Propred online servers and their antigenicity was found using the Vexijen online server. It was found that the MHC class I binding peptide MNWTPAVGM and the MHC class II binding peptide YVGATTASV had maximum antigenicity ensuring maximum binding affinity. Furthermore, all the selected epitopes were checked for their conservation with other countries of the world, and it was found that most of the epitopes were conserved among Pakistan and Canada, suggesting that these E1 epitopes of these two countries may be evolutionary related. Moreover, all the epitopes



**Figure 4** The HCV E1 protein model showing an epitopic location in the structure.

**Table 5 Conservation and variability analysis of B-cell and T-cell epitopes in comparison with HCV E1 proteins of other regions**

Peptide	India	Russia	Japan	USA	China	Nepal	Yemen	France	Canada	Somalia
VGQAFTFRPRRH	<b>VS</b> QLFTFS <b>SP</b> RRH	VSQLFTFS <b>SP</b> RRH	<b>IS</b> QLFTFS <b>SP</b> RRH	VGQLFTFS <b>SP</b> RRH	<b>AA</b> QLFI <b>IS</b> PXHH	VGQAFTFS <b>SP</b> RRH	VGQ <b>VIT</b> FK <b>PR</b> RH	VGQ <b>MFT</b> Y <b>RP</b> RQH	VGQAFTFRPRRH	VGQ <b>AFR</b> FR <b>QR</b> QH
TPVTPTVAVRYV	<b>VAL</b> TPTLA <b>ARNA</b>	<b>VAL</b> TPTLA <b>ARNA</b>	<b>VAL</b> TPTLA <b>ARNS</b>	<b>VAV</b> APT <b>VAT</b> RDG	<b>IPV</b> SP <b>NI</b> AV <b>QQP</b>	TPV <b>SPT</b> VAV <b>KHL</b>	<b>KPV</b> TPTVAV <b>AYG</b>	<b>VQI</b> TPTLS <b>APSF</b>	TPVTPTVAVRYV	TPVTPTVAV <b>RAP</b>
TPGCIPCVDGN	TPGC <b>VPCV</b> REGN	TPGC <b>VPCV</b> QEDN	TPGC <b>VPCV</b> REGN	<b>SP</b> GC <b>VPCV</b> REGN	<b>VP</b> GC <b>VPC</b> E <b>KV</b> GN	<b>LP</b> GC <b>VPCV</b> ATAN	<b>LP</b> GC <b>VPCV</b> KTGN	TPGC <b>VPCV</b> KEGN	TPGCIPCVDGN	<b>SP</b> GC <b>VPCV</b> K <b>SGN</b>
TNDCPNSSIVYE	TNDC <b>SN</b> SSIVYE	TNDC <b>SN</b> SSIVYE	TNDC <b>SN</b> SSIVYE	TNDCPNSSIVYE	TNDC <b>SN</b> DSIT <b>WQ</b>	TNDC <b>SN</b> Q <b>SIV</b> YE	TNDCPNSS <b>V</b> WYE	TNDCPNSSIVYE	TNDCPNSSIVYE	TNDCPNSSIVYE
ATTASVRSH	<b>VP</b> TT <b>IR</b> RH	<b>VP</b> TA <b>IR</b> RH	<b>VP</b> TT <b>IR</b> RH	<b>LP</b> TT <b>QL</b> RRH	ALTRGLRTH	ATTASIRSH	<b>AP</b> LES <b>FR</b> RH	<b>AX</b> T <b>AP</b> L <b>RRA</b>	ATTASVRSH	<b>VIT</b> ASIRSH
MNWTPAVGM	MN <b>W</b> S <b>P</b> T <b>AAL</b>	MN <b>W</b> S <b>P</b> T <b>TAL</b>	MN <b>W</b> S <b>P</b> T <b>AAL</b>	MN <b>W</b> S <b>P</b> T <b>TAL</b>	MN <b>W</b> S <b>P</b> T <b>ATM</b>	MN <b>W</b> S <b>P</b> A <b>I</b> GL	MN <b>W</b> S <b>P</b> T <b>TTL</b>	MN <b>W</b> S <b>P</b> T <b>TAL</b>	<b>N/A</b>	<b>QN</b> W <b>S</b> P <b>T</b> V <b>SL</b>
VGMWAQVLR	<b>AAL</b> W <b>S</b> Q <b>LL</b> RI	<b>TAL</b> W <b>S</b> Q <b>LL</b> RI	<b>AAL</b> V <b>AS</b> Q <b>L</b> FRI	<b>TAL</b> W <b>A</b> Q <b>LL</b> RV	<b>AT</b> M <b>IL</b> A <b>Y</b> AMRI	<b>I</b> GL <b>A</b> V <b>S</b> H <b>L</b> MRL	<b>TT</b> LL <b>L</b> AQ <b>I</b> MRI	<b>TAL</b> L <b>M</b> AQ <b>L</b> LRI	<b>N/A</b>	<b>V</b> S <b>L</b> I <b>V</b> AQ <b>V</b> LRL
LGGAHWGML	<b>VAG</b> AHWG <b>L</b>	<b>VAG</b> AHWG <b>V</b> L	<b>VAG</b> AHWG <b>V</b> L	<b>IAG</b> AHWG <b>V</b> L	<b>IS</b> GAHWG <b>V</b> M	<b>IAG</b> AHWG <b>V</b> M	<b>VAG</b> GHWG <b>V</b> L	<b>VAG</b> GHWG <b>V</b> L	<b>N/A</b>	<b>L</b> V <b>G</b> S <b>H</b> WG <b>V</b> L
SHVDLLVGA	<b>RH</b> V <b>D</b> LLVGA	<b>RH</b> V <b>D</b> LLVGA	<b>RH</b> V <b>D</b> LLVGA	<b>RH</b> I <b>D</b> LLV <b>G</b> S	<b>TH</b> I <b>D</b> M <b>W</b> V <b>M</b> S	SHV <b>D</b> M <b>L</b> VGA	<b>RH</b> V <b>D</b> L <b>M</b> VGA	<b>RA</b> V <b>D</b> Y <b>L</b> A <b>G</b> G	SHVDLLVGA	SHV <b>D</b> L <b>M</b> V <b>G</b> S
LTNDCPNSS	<b>VT</b> NDC <b>SN</b> SS	<b>VT</b> NDC <b>SN</b> SS	<b>VT</b> NDC <b>SN</b> SS	<b>VT</b> NDCPNSS	<b>VT</b> NDC <b>SN</b> DS	LTNDC <b>SN</b> Q <b>S</b>	<b>IT</b> NDCPNSS	<b>VT</b> NDCPNSS	LTNDCPNSS	<b>VT</b> NDCPNSS
WTPVTPTVA	<b>WV</b> AL <b>T</b> PTLA	<b>WV</b> AL <b>T</b> PTLA	<b>WV</b> AL <b>T</b> PTLA	<b>WV</b> A <b>V</b> APTVA	<b>WIP</b> V <b>SP</b> NI <b>A</b>	WTPV <b>S</b> P <b>T</b> V <b>A</b>	<b>WK</b> P <b>V</b> T <b>P</b> T <b>V</b> A	<b>WV</b> Q <b>I</b> T <b>P</b> T <b>L</b> S	WTPVTPTVA	WTPVTPTVA
VRYVGATTASV	<b>AR</b> NAS <b>V</b> P <b>IT</b> TI	<b>AR</b> NAS <b>V</b> P <b>IT</b> AI	<b>AR</b> NS <b>N</b> V <b>P</b> ITTI	<b>TR</b> D <b>G</b> K <b>L</b> P <b>I</b> T <b>Q</b> L	<b>VQ</b> Q <b>P</b> G <b>A</b> L <b>T</b> R <b>G</b> L	<b>VK</b> H <b>L</b> G <b>A</b> T <b>T</b> A <b>S</b> I	<b>V</b> A <b>Y</b> G <b>S</b> A <b>P</b> L <b>E</b> S <b>F</b>	<b>AP</b> S <b>F</b> G <b>A</b> X <b>T</b> A <b>P</b> L	VRYVGATTASV	<b>VR</b> A <b>P</b> G <b>V</b> I <b>T</b> A <b>S</b> I
YVGATTASV	<b>NA</b> S <b>V</b> P <b>IT</b> TI	<b>NA</b> S <b>V</b> P <b>IT</b> AI	<b>NS</b> N <b>V</b> P <b>IT</b> TI	<b>DG</b> K <b>L</b> P <b>I</b> T <b>Q</b> L	<b>Q</b> P <b>G</b> A <b>L</b> T <b>R</b> G <b>L</b>	<b>H</b> L <b>G</b> A <b>T</b> T <b>A</b> S <b>I</b>	<b>Y</b> G <b>S</b> A <b>P</b> L <b>E</b> S <b>F</b>	<b>S</b> F <b>G</b> A <b>X</b> T <b>A</b> P <b>L</b>	YVGATTASV	<b>AP</b> G <b>V</b> I <b>T</b> A <b>S</b> I
VRYVGATTA	<b>AR</b> NAS <b>V</b> P <b>IT</b>	<b>AR</b> NAS <b>V</b> P <b>IT</b>	<b>AR</b> NS <b>N</b> V <b>P</b> IT	<b>TR</b> D <b>G</b> K <b>L</b> P <b>I</b> T	<b>VQ</b> Q <b>P</b> G <b>A</b> L <b>T</b> R	<b>VK</b> H <b>L</b> G <b>A</b> T <b>T</b> A	<b>V</b> A <b>Y</b> G <b>S</b> A <b>P</b> L <b>E</b>	<b>AP</b> S <b>F</b> G <b>A</b> X <b>T</b> A	VRYVGATTA	<b>VR</b> A <b>P</b> G <b>V</b> I <b>T</b> A

Variable residues in epitopes are shown in bold.

showed some conservation with all other countries but there were frequent variations at some points.

## Conclusion

To develop effective vaccines it is important to target multiple antigenic components of the virus, thus directing the immune system to protect the host from the virus. Therefore, this study was conducted to predict antigenic determinants/epitopes of the HCV genotype 3a E1 protein along with the 3D protein modeling. The study revealed potential B-cell and T-cell epitopes that can raise the desired immune response to the HCV E1 protein isolated in Pakistan. For diagnosing HCV genotype 3a, these epitopes are highly useful and can also help in developing successful vaccines against HCV 3a infection to save the Pakistani population from potential HCV threats.

## Competing interests

All authors have no institutional or financial competing interests.

## Authors' contributions

UAA designed the study, and SI wrote the manuscript. UAA and SI performed all in-silico work, and UAA critically reviewed the manuscript. All the authors read and approved the final manuscript.

## Authors' information

Sobia Idrees (MPhil student), Usman A Ashfaq (PhD molecular Biology and Group leader, Human Molecular Biology Group, Department of Bioinformatics and Biotechnology, GCU, Faisalabad.

Received: 26 September 2012 Accepted: 4 April 2013

Published: 10 April 2013

## References

1. Kim JL, Morgenstern KA, Griffith JP, Dwyer MD, Thomson JA, Murcko MA, Lin C, Caron PR: Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. *Structure* 1998, **6**:89–100.
2. World Health Organization. Hepatitis C Fact Sheet: <http://www.who.int/mediacentre/factsheets/fs164/en/>.
3. Jafri W, Subhan A: Hepatitis C in Pakistan: magnitude, genotype, disease characteristics and therapeutic response. *Trop Gastroenterol* 2008, **29**:194–201.
4. Idrees M, Riazuddin S: Frequency distribution of hepatitis C virus genotypes in different geographical regions of Pakistan and their possible routes of transmission. *BMC Infect Dis* 2008, **8**:69.
5. Suzuki R, Suzuki T, Ishii K, Matsuura Y, Miyamura T: Processing and functions of Hepatitis C virus proteins. *Intervirology* 1999, **42**:145–152.
6. Kato N: Molecular virology of hepatitis C virus. *Acta Med Okayama* 2001, **55**:133–159.
7. Ashfaq UA, Ansar M, Sarwar MT, Javed T, Rehman S, Riazuddin S: Post-transcriptional inhibition of hepatitis C virus replication through small interference RNA. *Virology* 2011, **8**:112.
8. Ashfaq UA, Javed T, Rehman S, Nawaz Z, Riazuddin S: An overview of HCV molecular biology, replication and immune responses. *Virology* 2011, **8**:161.
9. Ciccaglione AR, Costantino A, Marcantonio C, Equestre M, Geraci A, Rapicetta M: Mutagenesis of hepatitis C virus E1 protein affects its membrane-permeabilizing activity. *J Gen Virol* 2001, **82**:2243–2250.
10. Burlone ME, Budkowska A: Hepatitis C virus cell entry: role of lipoproteins and cellular receptors. *J Gen Virol* 2009, **90**:1055–1070.
11. Ashfaq UA, Masoud MS, Khaliq S, Nawaz Z, Riazuddin S: Inhibition of hepatitis C virus 3a genotype entry through Glanthis Nivalis Agglutinin. *Virology* 2011, **8**:248.
12. Munir S, Saleem S, Idrees M, Tariq A, Butt S, Rauff B, Hussain A, Badar S, Naudhani M, Fatima Z, et al: Hepatitis C treatment: current and future perspectives. *Virology* 2010, **7**:296.
13. Ashfaq UA, Khan SN, Nawaz Z, Riazuddin S: In-vitro model systems to study Hepatitis C Virus. *Genet Vaccines Ther* 2011, **9**:7.
14. Fournillier A, Wychowski C, Boucreux D, Baumert TF, Meunier JC, Jacobs D, Muguet S, Depla E, Inchauspe G: Induction of hepatitis C virus E1 envelope protein-specific immune response can be enhanced by mutation of N-glycosylation sites. *J Virol* 2001, **75**:12088–12097.
15. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF: Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* 1999, **112**:531–552.
16. Chen CC, Hwang JK, Yang JM: (PS)2: protein structure prediction server. *Nucleic Acids Res* 2006, **34**:W152–W157.
17. Laskowski RA, MacArthur MW, Moss DS, Thornton JM: PROCHECK - a program to check the stereochemical quality of protein structures. *J App Cryst* 1993, **26**:283–291.
18. Wiederstein M, Sippl MJ: ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 2007, **35**:W407–W410.
19. Colovos C, Yeates TO: Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* 1993, **2**:1511–1519.
20. Doytchinova IA, Flower DR: VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* 2007, **8**:4.
21. Krogh A, Larsson B, Von Heijne G, Sonnhammer EL: Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 2001, **305**:567–580.
22. Kringelum JV, Lundegaard C, Lund O, Nielsen M: Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLoS Comput Biol* 2012, **8**:e1002829.
23. Somvanshi P, Singh V, Seth PK: In Silico Prediction of Epitopes in Virulence Proteins of Mycobacterium Tuberculosis H37Rv for Diagnostic and Subunit Vaccine Design. *J Proteomics Bioinform* 2008, **1**:143–153.
24. Idrees S, Ashfaq UA: A brief review on dengue molecular virology, diagnosis, treatment and prevalence in Pakistan. *Genet Vaccines Ther* 2012, **10**:6.
25. EL-Manzalawy Y, Dobbs D, Honavar V: Prediction of linear B-cell epitopes using string kernels. *J Mol Recognit* 2008, **21**:243–255.
26. 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 Bioinformatics* 2007, **8**:361.

doi:10.1186/1743-422X-10-113

Cite this article as: Idrees and Ashfaq: Structural analysis and epitope prediction of HCV E1 protein isolated in Pakistan: an in-silico approach. *Virology Journal* 2013 **10**:113.

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)

