Journal of Biomolecular Structure and Dynamics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tbsd20

Core amino acid substitutions in HCV-3a isolates from Pakistan and opportunities for multi-epitopic vaccines

Ayyaz Khan, Mehboob Nawaz, Saeed Ullah, Irshad Ur Rehman, Abbas Khan, Shoaib Saleem, Nasib Zaman, Zabta Khan Shinwari, Muhammad Ali & Dong-Qing Wei

To cite this article: Ayyaz Khan, Mehboob Nawaz, Saeed Ullah, Irshad Ur Rehman, Abbas Khan, Shoaib Saleem, Nasib Zaman, Zabta Khan Shinwari, Muhammad Ali & Dong-Qing Wei (2020): Core amino acid substitutions in HCV-3a isolates from Pakistan and opportunities for multi-epitopic vaccines, Journal of Biomolecular Structure and Dynamics, DOI: 10.1080/07391102.2020.1850353

To link to this article: https://doi.org/10.1080/07391102.2020.1850353

JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS
https://doi.org/10.1080/07391102.2020.1850353

Core amino acid substitutions in HCV-3a isolates from Pakistan and opportunities for multi-epitopic vaccines

Ayyaz Khan[a], Mehboob Nawaz[a], Saeed Ullah[b], Irshad Ur Rehman[c], Abbas Khan[d], Shoaib Saleem[e], Nasib Zaman[f], Zabta Khan Shinwari[a,g], Muhammad Ali[a] and Dong-Qing Wei[d,h,i]

[a]Department of Biotechnology, Quaid-i-Azam University Islamabad, Islamabad, Pakistan; [b]Saidu Group of Teaching Hospital, Swat, Pakistan; [c]Center of Biotechnology and Microbiology, University of Peshawar, Peshawar, Pakistan; [d]Department of Bioinformatics and Biological Statistics, School of Life Sciences and Biotechnology, Shanghai Jiaotong University, Shanghai, China; [e]National Center for Bioinformatics, Quaid-i-Azam University Islamabad, Islamabad, Pakistan; [f]Center of Biotechnology and Microbiology, University of Swat, Swat, Pakistan; [g]Pakistan Academy of Sciences, Islamabad, Pakistan; [h]State Key Laboratory of Microbial Metabolism, Shanghai-Islamabad-Belgrade Joint Innovation Center on Antibacterial Resistances, Joint Laboratory of International Cooperation in Metabolic and Developmental Sciences, Ministry of Education and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, P.R. China; [i]Peng Cheng Laboratory, Shenzhen, Guangdong, P.R China

Communicated by Ramaswamy H. Sarma

ABSTRACT

Hepatitis C virus (HCV), which infected 71 million worldwide and about 5%–6% are from Pakistan, is an ssRNA virus, responsible for end-stage liver disease. To date, no effective therapy is available to cure this disease. Hence, it is important to study the most prevalent genotypes infecting human population and design novel vaccine or small molecule inhibitors to control the infections associated with HCV. Therefore, in this study clinical samples ($n = 35$; HCV-3a) from HCV patients were subjected to Sanger sequencing method. The sequencing of the core gene, which is generally considered as conserved, involved in the detection, quantitation and genotyping of HCV was performed. Multiple mutations, that is, R46C, R70Q, L91C, G60E, N/S105A, P108A, N110I, S116V, G90S, A77G and G145R that could be linked with response to antiviral therapies were detected. Phylogenetic analysis suggests emerging viral isolates are circulating in Pakistan. Using *ab initio* modelling technique, we predicted the 3D structure of core protein and subjected to molecular dynamics simulation to extract the most stable conformation of the structure for further analysis. Immunoinformatic approaches were used to propose a multi-epitopes vaccine against HCV by using core protein. The vaccine constructs consist of nine CTL and three HTL epitopes joined by different linkers were docked against the two reported Toll-like receptors (TLR-3 and TLR-8). Docking of vaccine construct with TLR-3 and TLR-8 shows proper binding and *in silico* expression of the vaccine resulted in a CAI value of 0.93. These analyses suggest that specific immune responses may be produced by the proposed vaccine.

ARTICLE HISTORY
Received 12 May 2020
Accepted 9 November 2020

KEYWORDS
Hepatitis C virus; core protein; amino acid substitution; phylogenetics; protein modelling; epitopic vaccine

Introduction

Hepatitis C virus (HCV) is one of the viruses (A, B, C, D and E), which cause liver inflammation in humans (Choo et al., 1989). HCV causes acute hepatitis and can sometimes lead to chronic hepatitis known to be the primary cause of hepatic cancer and cirrhosis (Hoofnagle, 2002; Perz et al., 2006). HCV was also termed as 'Viral time bomb' because it is asymptomatic nature, have high prevalence, low diagnosis and complexity in treatment (Csete et al., 2008).

The genotypes and prevalence of HCV infection vary in different geographic regions. According to WHO, about 71 million individuals have been living with chronic HCV infection worldwide and every year, 399,000 individuals die of HCV induced liver cancer and liver cirrhosis (World Health Organization, 2017). In Pakistan, HCV is exceptionally prevalent, about 5%–6% of the population is infected (Umer & Iqbal, 2016). One in every 20 Pakistanis is HCV-infected (Al Kanaani et al., 2018) and genotype 3a is the most prevalent (Afridi et al., 2013; Attaullah et al., 2011; Idrees & Riazuddin, 2008), followed by untypable (Afzal et al., 2013; S. Aziz et al., 2010; Grebely et al., 2014).

The HCV genome is about 9.6 kb, with a single open reading frame (ORF) encoding a single polyprotein of 3010 amino acids (Choo et al., 1989; Reed & Rice, 2000), which is post-translationally cleaved by virus and host-specific protease into structural (Core C, Envelope E1 and E2) and non-structural proteins (P7, NS2, NS3, NS4A, NS4B, NS5A and NS5B; Scheel & Rice, 2013).

Core gene of HCV is generally considered conserved, involved in the detection, quantitation (Okazaki et al., 2008) and genotyping of HCV (Idrees, 2008; Ohno et al., 1997). Core protein positioned at the polyprotein N-terminus, a 191 amino acid protein, multimerizes to form the viral capsid shell. The core protein is of various sizes (17–23 kDa), among

CONTACT Muhammad Ali alibiotech01@gmail.com; Dong-Qing Wei dqwei@sjtu.edu.cn
Supplemental data for this article can be accessed online at https://doi.org/10.1080/07391102.2020.1850353

which the 21 kDa (P21) is the most prevalent form (Yasui et al., 1998). Core protein initially cleaved by host signal peptidase (SP) at the endoplasmic reticulum (Moradpour et al., 2007) and subsequent removal of C-terminal E1 signal sequence by signal peptide peptidase along with transmembrane region eventually yields mature core protein (Okamoto et al., 2008). There are two distinct functional domains of mature core protein, that is, domain 1 (1–177aa) and domain 2 (118–177aa; McLauchlan, 2000). Domain 1 (1–177aa) is hydrophilic, having majority of amino acids positively charged and involved in RNA binding. N-terminus of domain 1 comprises immunodominant antigenic B-cell epitope (Jolivet-Reynaud et al., 1998; Ménez et al., 2003). Domain 2 (118–177) constitutes the C-terminal, is hydrophobic and function in association with the membrane of endoplasmic reticulum and lipid droplets of mammalian cells (Akuta et al., 2005; Boulant et al., 2005; Shavinskaya et al., 2007).

The immune response against HCV infection does not show a fixed pattern. However, in many viral infections, the specific immune response is based on dendritic and macrophage cells that present viral protein to helper T cells, cytotoxic T cells and B cells. Antibodies produced by B-cell clear virus circulating in the blood and prevent from reinfection, for example, antibodies produced against hepatitis B virus (HBV) surface antigen are very crucial for clearance of HBV virion. T cells ($CD4^+$) have specific receptors, which recognize peptides derived from proteolytically cleaved viral proteins via class II MHC molecules. $CD8^+$ T-cell response in HCV-infected patients is defective (Golden-Mason & Rosen, 2006; Liang et al., 2000; S. T. Khan, Karges, et al., 2018; Scarselli et al., 2002). Before the discovery of direct-acting antiviral agents (DAAs), ribavirin (RBV) and pegylated interferon (Peg-IFN) remained the standard treatment for HCV infection; however, it had lower response rates and adverse reactions to these drugs happened in a substantial percentage of patients (Chen et al., 2016; Hunt & Pockros, 2013; Manns et al., 2001; Spengler, 2018). Multiple reasons, however, such as the excessive genetic variability such as R70Q, R70H and L91M reported to cause resistance to IFN therapy in HCV genome and the possible risks of testing killed or live-attenuated vaccines in clinical trials, are significant obstacles to the development of a successful HCV vaccine.

Immunoinformatic methods are a promising alternative for recognizing and designing extremely immunogenic HCV multi-epitope vaccine (Memarnejadian et al., 2009). As a result, a vast number of study attention has now shifted to understanding an anti-HCV vaccine design based on immunoinformatics (Ikram et al., 2018). An appropriate HCV multi-epitope vaccine should include immunogenic epitopes that can produce effective responses from $CD4^+$, $CD8^+$ T and B cells (Sabet et al., 2014).

This study aimed to perform nucleotide sequencing of the core region of HCV, its phylogenetic analysis and core protein amino acid substitutions in chronically HCV-infected patients (genotype 3a). We also predicted the three-dimensional (3D) structure of core protein, performed molecular dynamics (MD) simulation and designed multi-epitopic vaccine. The constructed vaccine was further assessed by methods of molecular docking and validated through *in silico* expression.

Materials and methods

Sample collection

Positive serum samples of hepatitis C were obtained from patients visiting Lahore, Peshawar, Karak and AJK hospitals from August 2018 to May 2019. From all respondents, written permission was obtained. A total of 372 HCV-positive patients' sera were collected, in which ($n = 169$) 52.5% were females and ($n = 153$) 47.5% were males. The mean age of the patients was 38 ± 9.79 years (range: 10–70).

Inclusion criteria

Serum samples positive for viral RNA (by PCR) were included. HCV chronic infection patients, treated with different antiviral therapies (sofosbuvir 400 mg, daclatasvir 60 mg and RBV) received therapy for more than 48 weeks, having complete demographic characteristics and were willing to cooperate in the study were included. Only HCV genotype 3a isolates were included in the study.

Exclusion criteria

Those patients who had co-infections (i.e. HBV, HAV and/or HDV) and did not have history of treatment were not included. Genotypes other than 3a were excluded.

HCV RNA extraction and cDNA synthesis

HCV RNA was extracted from 200 µl serum sample using GF-1 Nucleic acid Extraction kit (Vivantis Technologies, Selangor, Malaysia) according to the manufacturer protocol. Complementary DNA (cDNA) was synthesized using RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) and random primers (Thermo Fisher Scientific) in a 20 µl reaction mixture which include: 10 µl of extracted RNA, 1 µl random hexamer primers, 1 µl double distilled water, 4 µl 5× reaction buffer, 2 µl dNTP (10 mM), 1 µl RevertAid M-MuLV RT (200 U/µl) and 1 µl RiboLock (20 U/µl). All the reagents were mixed gently and incubate on 25 °C for 5 min, followed by 45 °C for 60 min and finally reaction was terminated at 70 °C for 5 min. The quantity of cDNA was assessed through NanoDrop spectrophotometer (Thermo Fisher Scientific) and the presence of HCV RNA was confirmed by nested PCR for 5′-UTR region of HCV.

Genotyping

Type-specific HCV genotyping was performed as described earlier with little amendments (Idrees, 2008).

Amplification and sequencing of core gene

HCV-3a core gene was amplified through nested PCR using OneTaq Master Mix (New England BioLabs). The sequences

Table 1. Name of primers, sequences, nucleotide position (reference: NZL1) and size of the amplified product.

S. No.	Primer	Oligonucleotide sequence 5′–3′	Nucleotide position[a]	PCR products size
1.	ExFCore	GGATCCTGCAACATGAGCACACTTCC	343–363	~573bp
2.	ExRCore	CTCGAGAGACGTGCCCGCCACTCT	896–915	
3.	InFCore	ATGAGCACACTTCCTAAACCTCA	339–362	
4.	InRCore	ACTGGCTGCTGGATGAATTAAGC	912–890	

[a]Reference NZL1 genome (GenBank accession number D17763).

of primers are given in Table 1. First-round PCR was performed in 25 µl reaction mixture containing: double distilled water 5 µl, OneTaq Master Mix 12.5 µl, ExFCore primer (10 µM) 1 µl, ExRCore primer (10 µM) 1 µl, cDNA 5 µl and DMSO 0.5 µl. Thermocycling condition was; initial denaturation 95 °C for 2 min: 35 cycles (95 °C for 30 s, 56 °C for 40 s, 72 °C for 1 min) final extension was 72 °C for 10 min and then hold on 4 °C for infinity. The second round nested PCR amplification was performed using 4 µl of the first-round product and 1 µl of each nested primers (InFCore and InRCore). The thermocycling conditions were same as in first-round PCR only annealing temperature was reduced to 53 °C. The amplified product of 573 bp, was purified from agarose gel using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific), and nucleotide sequencing was performed through the Sanger sequencing method.

Phylogenetic analysis

For homology with other sequences, 35 sequences (core gene HCV 3a) of present work were searched in NCBI gene bank using nucleotide BLAST. For determination of regional distribution sequences of core gene, HCV 3a from different countries namely, India, UK, China, the USA, Iran, Indonesia, Canada and Thailand were retrieved from the GenBank NCBI. Sequences were aligned using MEGAX (clustalW; Kumar et al., 2018). Phylogenetic trees were constructed among present work sequences and with sequences retrieved from above-mentioned countries. BEAST software version 1.8.4 was used for phylogenetic analysis, through Bayesian likelihood algorithm (Drummond & Rambaut, 2007). The SRD06 substitution model with relaxed lognormal clock was used. The analysis was run 700 million generations and after every 70,000 generations trees were sampled which result in 7000 trees. For tree annotation, TreeAnnotator v.1.8.3 was used through burn-in of 1000 trees. The tree was visualized through FigTree (v.1.4.2; http://tree.bio.ed.ac.uk/software/figtree/).

Alignment of core protein sequences

The nucleotide sequences of all samples were translated to amino acid sequences using CLC sequence viewer (version 8) software and online server ExPASy (https://web.expasy.org/translate/; Artimo et al., 2012). The amino acid sequence alignment was performed using MEGAX (Kumar et al., 2018). The sequences were aligned with two reference protein sequence of HCV isolate NZL1 (GenBank protein ID: BAA04609; Sakamoto et al., 1994) and K-3a (UniProt ID: Q81495; McLauchlan, 2000).

Protein structure prediction

The crystallographic three-dimensional (3D) structure of the HCV core protein has not been reported yet, so *ab initio* modelling was used. The primary amino acid sequence of the core protein was searched in Universal Protein Resource (UniProt) to confirm it and to obtain a sequence of core protein with high annotation score. BLAST search in UniProt results in the most suitable amino acid sequence of core protein genotype 3a (Q81495), which was used as a template for protein structure prediction. Initiator methionine removed by the host, so amino acid sequence without methionine initiator was used for protein structure prediction. For structure prediction, the amino acid sequence of reference core protein (Q81495) was submitted to the Robetta server *de novo* structure prediction methods (http://robetta.bakerlab.org/; Kim et al., 2004) and I-TASSER (Zhang, 2008) Quark-based *ab initio* modelling (https://zhanglab.ccmb.med.umich.edu/QUARK/; Xu & Zhang, 2013). Among the predicted models, the best model was carefully chosen based on PROCHECK (Laskowski et al., 1996), Ramachandran plot (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php; Biasini et al., 2014), ProSA-web (Wiederstein & Sippl, 2007), ERRATv2.0 (Colovos & Yeates, 1993), Verify3D (Bowie et al., 1991) and PDBsum (Laskowski et al., 2018) evaluations.

Impact of the reported substitutions on the structural stability

To determine the impact of these substitutions on the stability of the protein we used mCSM server (Pires et al., 2014), which used signature-based method to compute the stability index upon substitution. The server accepts a PDB format as input.

Molecular dynamics simulation

MD simulation was conducted for the selected model of Robetta server (model 3) using AMBER 14 package (Pearlman et al., 1995; Salomon-Ferrer et al., 2013). Hydrogen and sodium (Na^+) ions added for system counter neutralization with the help of 'tleap' Amber package. SANDER module (AMBER 14) was used for energy minimization of the selected core protein. To remove constraints atoms the SANDER module was applied at two stages (about 6000 steps in each stage; Salomon-Ferrer et al., 2013). The SHAKE and particle mesh Ewald (PME) method with a 10 Å cut-off radius of non-bond contacts used for long-term interaction. Using the 310 K (36.85 °C) Langevin temperature and constant pressure (1 atm) with isotropic molecule-based scaling was estimated for 10,000 ps time equilibrium, followed by a total of 100 ns

simulation. After every 2.0 ps time scale, MD trajectory sampling was performed. Analysis such as RMSD (root mean square deviation), RMSF (root mean square fluctuation), radius of gyration and B-factor was calculated using CPPTRAJ and PYTRAJ in AMBER 14. The following equation was solved after 100 ns to calculate the stability of protein structure.

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=0}^{N}[m_i \times (X_i \times \ Y_i)^2]}{M}}$$

where N is atom number, m_i is particle mass i, X_i is the co-ordinate vector for target atom i, Y_i is the co-ordinate vector for reference atom i and M is the total mass. All $m_i = 1$ and $M = N$ if the RMSD is not mass-weighted.

Cytotoxic T-lymphocyte (CTL) epitope prediction

For CTL epitopes of the primary amino acids, the sequence of HCV core protein was submitted to an online server NetCTL1.2 (http://www.cbs.dtu.dk/services/NetCTL/; Larsen et al., 2007). NetCTL1.2 combines three approaches to predict the T-cell epitopes such as C-terminal cleavage score, transporter-associated protein score (TAP) for processing the antigen-MHC complex to the surface of the cell and MHC-I binding peptide prediction efficiency. The prediction threshold 0.75 was set as default. CTL epitopes with high MHC binding affinity were shortlisted for further analysis. The CTL epitopes of HCV core protein were predicted for all the available serotypes of HLA (human leukocyte antigen).

Helper T-cell (HTL) epitope prediction

For HTL epitope prediction, an online server module Immune Epitope Database (IEDB; http://www.iedb.org/; Vita et al., 2015) was used. The online server (IEDB) is based on the affinity of epitope with receptors, and the epitope can be ranked based on IC_{50} value, which is an integrated function in the server and assigns a specific value to each epitope. Peptides with high affinity must have IC_{50} value <50 nM. The IC_{50} value less than 500 (<500) nM shows an intermediate affinity. The percentile rank is inversely related to binding affinity of the epitope to its receptor. HTL Epitopes against seven human alleles, namely HLA-DRB3*02:02, HLA-DRB1*07:01, HLA-DRB3*01:01, HLA-DRB1*15:01, HLA-DRB4*01:01, HLA-DRB5*01:01 and HLA-DRB1*03:01 were predicted.

IFN-gamma inducing peptide predictions

To test the ability of epitopes in provoking immune responses (Th1) and IFN-γ production, predicted epitopes were analysed by an online server IFN epitope server (http://crdd.osdd.net/raghava/ifnepitope/index.php; Dhanda et al., 2013) using predict option. Motif and SVM hybrid was designated as the method and IFN-gamma (IFN-γ) versus other cytokines as a prediction model. IFNs minimize host damage and have a defensive character against infectious diseases. In cancer IFNs used in chemotherapy (Green et al., 2016; Hiramatsu et al., 2015), enteroviral myocarditis treatment (Kühl et al., 2012) and for hepatitis treatments (Lin et al., 2016). In addition, IFNs also used as adjuvants in combination with antigen for multiple diseases in vaccine designing (Toporovski et al., 2010) like influenza (James et al., 2007), HPV (Öhlschläger et al., 2009) and HIV (Abaitua et al., 2006; Jiang et al., 2007).

B-cell epitope prediction

B-cell epitopes were predicted using ABCpred server (http://crdd.osdd.net/raghava/abcpred/; El-Manzalawy et al., 2008). ABCpred server using the artificial neural network (machine learning-based method) to predicts linear B-cell epitopes using fixed-length patterns. This is the first server which is based on recurrent neural network and predicts B-cell epitopes with an accuracy of 65.93%.

Vaccine construct

To join the CTL epitope with the selected three HTL epitopes AAY linkers (Sabourin et al., 2007) were used. AAY linkers between two epitopes are required for the separation of epitopes which is essential for effective working. In addition, a β-defensin amino acid sequence was added at the *N*-terminal of final vaccine construct with the help of EAAAK linker as an adjuvant. β-defensin helps in the recruiting of naïve T cells and immature dendritic cells at the infection site.

Antigenicity and allergenicity

To predict antigenicity of the final vaccine construct an online tool VaxiJen (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html; Doytchinova & Flower, 2007) was also used with the cut-off of a score of as much as 0.5. Vaccine or epitope with an antigenicity greater than 0.5 was considered as antigenic. The server uses the physicochemical properties of the protein sequences and predicts the antigenicity and allergenicity.

Molecular docking

Immune cells bear immune receptors on their surface that play a crucial role in inducing a robust immune response. Molecular docking of the final vaccine with Toll-like receptors (TLR-3; PDB ID: 2A0Z) and (TLR-8; PDB ID: 3W3J) was performed using an online server PatchDock (https://bioinfo3d.cs.tau.ac.il/PatchDock/; Schneidman-Duhovny et al., 2005).

Codon optimization and *in silico* cloning

JCAT tool was used for reverse translation and optimization of codons for host *E. coli* (K12 strain) to obtain a high level of protein expression. JCAT results include CAI (codon adaptation index) and GC content of the query sequence. NdeI

Table 2. Gender-wise distribution of HCV genotypes ($n = 322$).

Genotype	Subtype	Male	Female	Total	Percentage
1	1a	7	0	7	2.2
	1b	9	12	21	6.5
2	2a	3	0	3	0.9
	2b	7	9	17	5.0
3	3a	57	72	129	40
	3b	15	12	27	8.4
4		6	16	22	6.8
Untypable		46	48	94	29.2
Mixed genotypes		3	0	3	0.9
Total		153	169	322	100

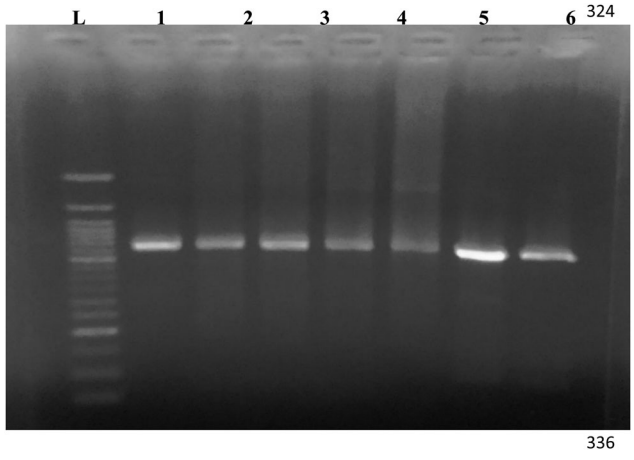

Figure 1. Typical agarose gel electrophoresis photograph of HCV 3a core gene amplified products. Lane 'L' shows 50 bp DNA size marker ladder (England Biolabs); Lane 1 shows positive control. Lanes 2, 3, 4, 5, 6 and 7 show 573-bp amplified HCV 3a genotype core gene.

and XhoI restriction sites were added to the reverse translated sequence of the protein. SnapGene tool was used for cloning of the reverse translated sequence into *E. coli* strain K12, pET-28a(+) vector was used to carry the sequence to the host. All these analyses were performed using *in silico* tools.

Results

A total of 372 HCV serum samples were collected from Lahore, Peshawar, Karak and AJK. Out of 372 samples, 322 were found positive for HCV RNA. The patients were of different ages ranging from 17 to 66 years in which 52.5% ($n = 169$) were females, and 47.5% ($n = 153$) were males. Out of 372 samples (155; 41.66%) were collected from Lahore Punjab (85; 22.84%) from Peshawar (92; 24%) from Karak KP and (40; 10.75%) from AJK. HCV genotype 3a was found to be dominant 129 (40.00%) in both genders, followed by untypable samples 94 (29.19%). The genotypes of other samples were: subtype 1a ($n = 70$), 1b ($n = 21$), 2a ($n = 3$), 2b ($n = 17$), 3b ($n = 27$), genotype 4 ($n = 22$) and mixed genotypes ($n = 3$). The demographic characteristics of all enrolled patients are given in Table 2.

Nucleotide sequencing of core gene

In this work, a total of 35 sequences of core gene of HCV 3a (about 573 bp in size) were obtained that were resistant to antiviral therapy (combination therapy of either Sofosbuvir + RBV or Sofosbuvir + Daclatasvir). The sequencing of core gene also confirmed our genotyping results. The

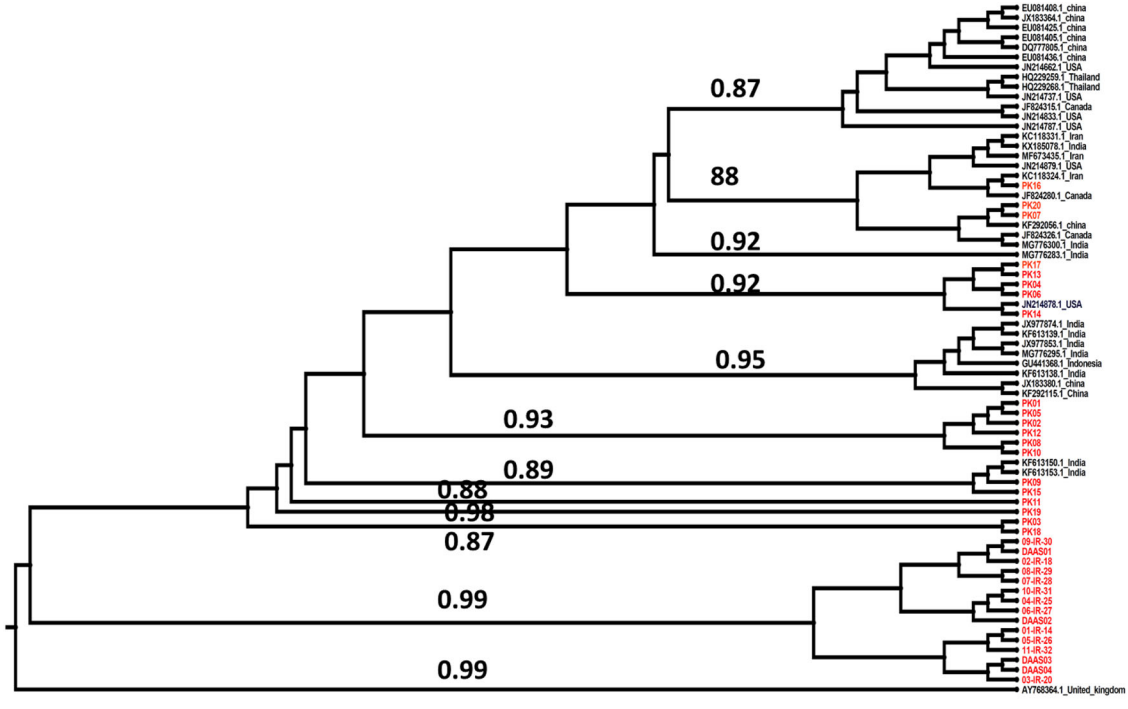

Figure 2. Phylogenetic relationship among HCV isolates of the current study with worldwide isolates. The phylogenetic tree was constructed BEAST package 1.8.4. Red colour indicates isolates of the current study, whereas black colour shows reference isolates retrieved from NCBI.

agarose gel representing the 573 base pairs of core gene HCV 3a is represented in Figure 1.

Phylogenetic analysis

In the current work, 35 HCV-3a sequences of core gene were obtained, whereas further 37 sequences were retrieved from the gene bank NCBI as reference sequences from various countries, that is, Canada, India, Indonesia, the USA, Thailand and UK. The phylogenetic tree(s) were constructed. The phylogenetic analysis shows between present-study sequences show emerging viral isolates. All the isolates were clustered into six different groups. The isolates PK09 and PK18 were at the phylogenetic tree's root and were considered viral ancestor, whereas the remaining are relatively emerging viral isolates (Supplementary Figure S1).

Table 3. Main amino acid substitutions found in the hepatitis C virus core protein of genotype 3a.

Position	Wild type	Variants
46	R	C ($n = 5$; 14.28%)
70	R	Q ($n = 2$; 5.71%)
91	L	C ($n = 14$; 40%)
60	G	E ($n = 2$; 5.71%)
105	N/S	A ($n = 2$; 5.71%)
108	P	A ($n = 5$; 14.28%)
110	N	I ($n = 2$; 20%)
116	S	V ($n = 4$; 11.42%)
90	G	S ($n = 1$; 2.85%)
77	A	G ($n = 7$; 20%)
145	G	R ($n = 5$; 14.28%)

n is the number of patients having variant amino acid.

The recently reported sequences were analysed phylogenetically with reference sequences (HCV 3a) from various countries including Indonesia, China, UK, Thailand, Japan, India, the USA, Canada and other countries. These countries reference sequences have been acquired from the NCBI database. Current study isolate (PK16) clustered with Irani sequence (KC118324.1_Iran), whereas isolates (PK20) and (PK07) show homology with sequences of China, India and Canada. Isolate (PK14) was clustered with another isolate reported from the USA. The study demonstrates that most of our present HCV 3a core gene isolates clustered together. The taxa of recent study viral isolates are given in red, whereas other worldwide isolates are presented in black colour as shown in Figure 2.

Amino acid substitutions in the core protein

Main amino acid substitutions that were found in core protein (genotype 3a) are R46C, R70Q, L91C, G60E, N/S105A, P108A, N110I, S116V, G90S, A77G and G145R (Table 3).

Protein modelling and validation

The structure of HCV core protein has not been reported yet, so the structure was predicted using *ab initio* modelling was used to predict the 3D structure from the amino acid sequence of reference core protein (Q81495). Robetta server was used, which uses *de novo* structure prediction method (http://robetta.bakerlab.org/). For *ab initio* modelling, Quark modelling (https://zhanglab.ccmb.med.umich.edu/QUARK/) server was used. The scores predicted by different structure

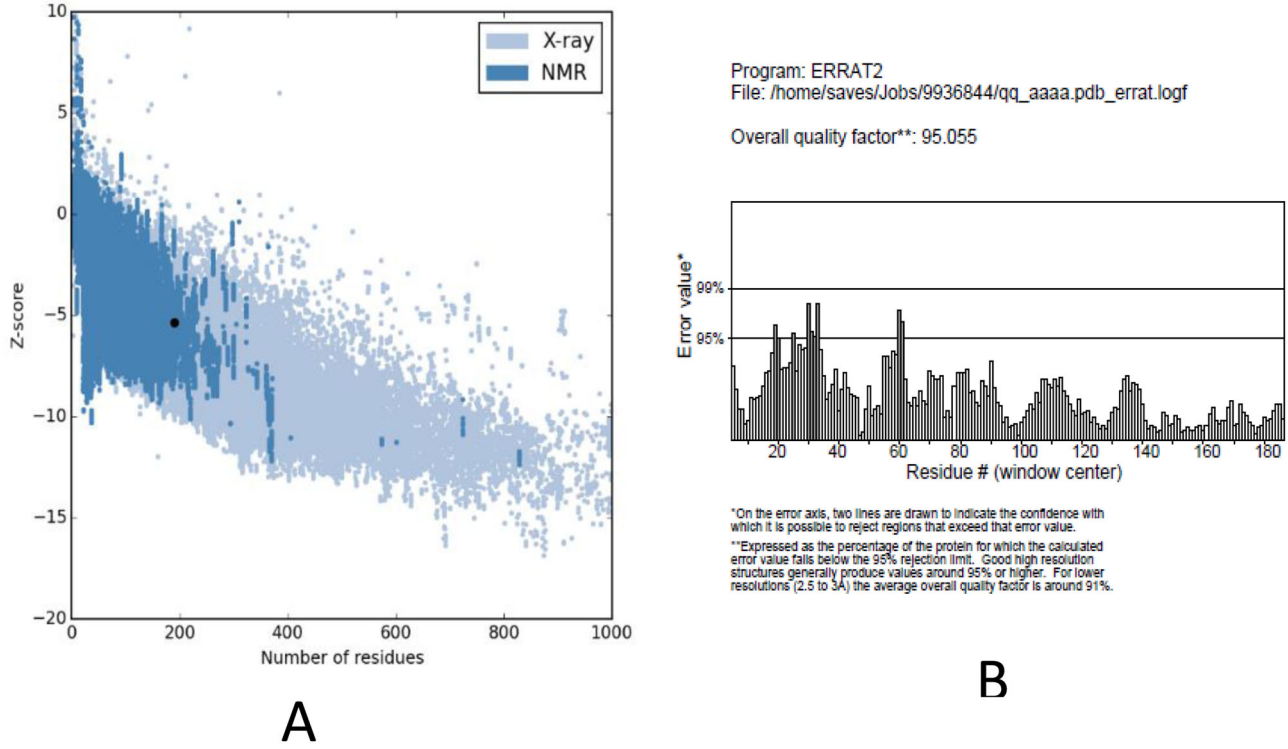

Figure 3. The structural evaluation results generated by the ERRAT and ProSA-web server. (A) *Z*-score analysis of model 3 using ProSA-web, (B) ERRAT plot of model 3.

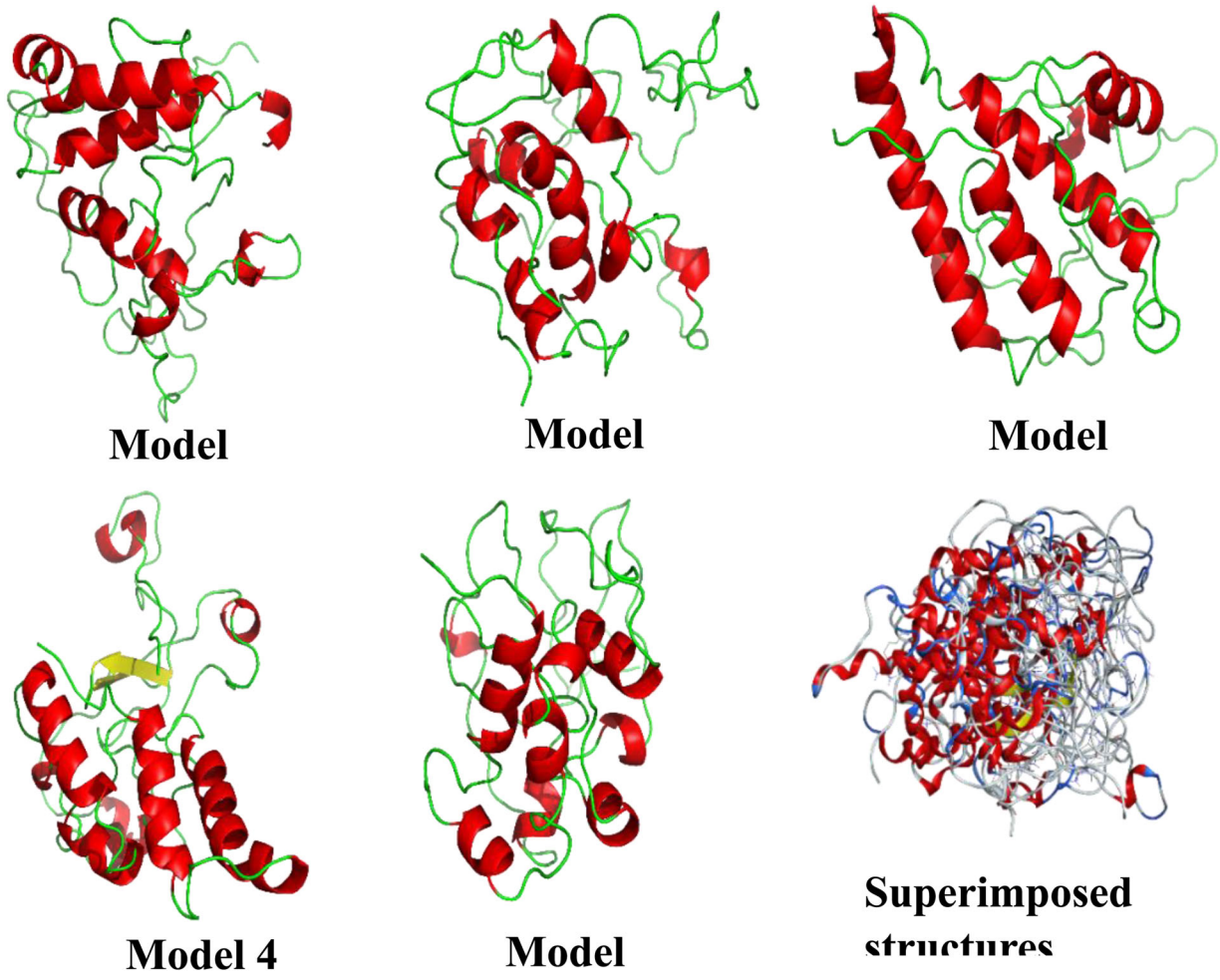

Figure 4. The 3D protein models of the core protein generated by the Robetta web server. The figure shows the top five models generated and the last figure is the superimposed of the five predicted models. For visualization, PyMOL software package was used.

Table 4. Structural details of protein model, that is, ERRAT, Verify3D, *Z*-scores and PDB sum evaluations for models generated by Robetta server.

				PDB sum evaluations				
Model	ERRAT Quality factor	Verify3D 3D-1D score >0.2 (%)	Pro-SA web *Z*-score	Most favoured regions	Additional allowed regions	Generously allowed regions	Disallowed regions	G-Factors
Model 1	89.9441	87.89%	−5.68	88.2 %	11.8%	0	0	0.27
Model 2	80.5556	78.42%	−5.34	82.6%	17.4%	0	0	0.08
Model 3	95.0549	85.79%	−5.35	92.4%	7.6%	0	0	0.25
Model 4	86.7403	83.16%	−6.84	90.3%	9.7%	0	0	0.27
Model 5	81.2155	84.21%	−5.73	86.8%	11.8%	0.7%	0.7%	0.15

Table 5. The results generated by the mCSM server. The specified impact is given in the last column.

PDB_FILE	Chain	RES_POS	RSA	PRED_DDG	Outcome
Core Protein	A	R46C	43	−0.617	Destabilizing
Core Protein	A	G60E	106.6	−0.429	Destabilizing
Core Protein	A	R70Q	38.6	−0.823	Destabilizing
Core Protein	A	A77G	68.7	−0.649	Destabilizing
Core Protein	A	G90S	70.6	−0.111	Destabilizing
Core Protein	A	L91C	0	−1.533	Destabilizing
Core Protein	A	N105A	111.2	−0.145	Destabilizing
Core Protein	A	P108A	60.4	−0.897	Destabilizing
Core Protein	A	N110I	48	0.095	Stabilizing
Core Protein	A	S116V	57.6	0.253	Stabilizing
Core Protein	A	G145R	108.3	−0.261	Destabilizing

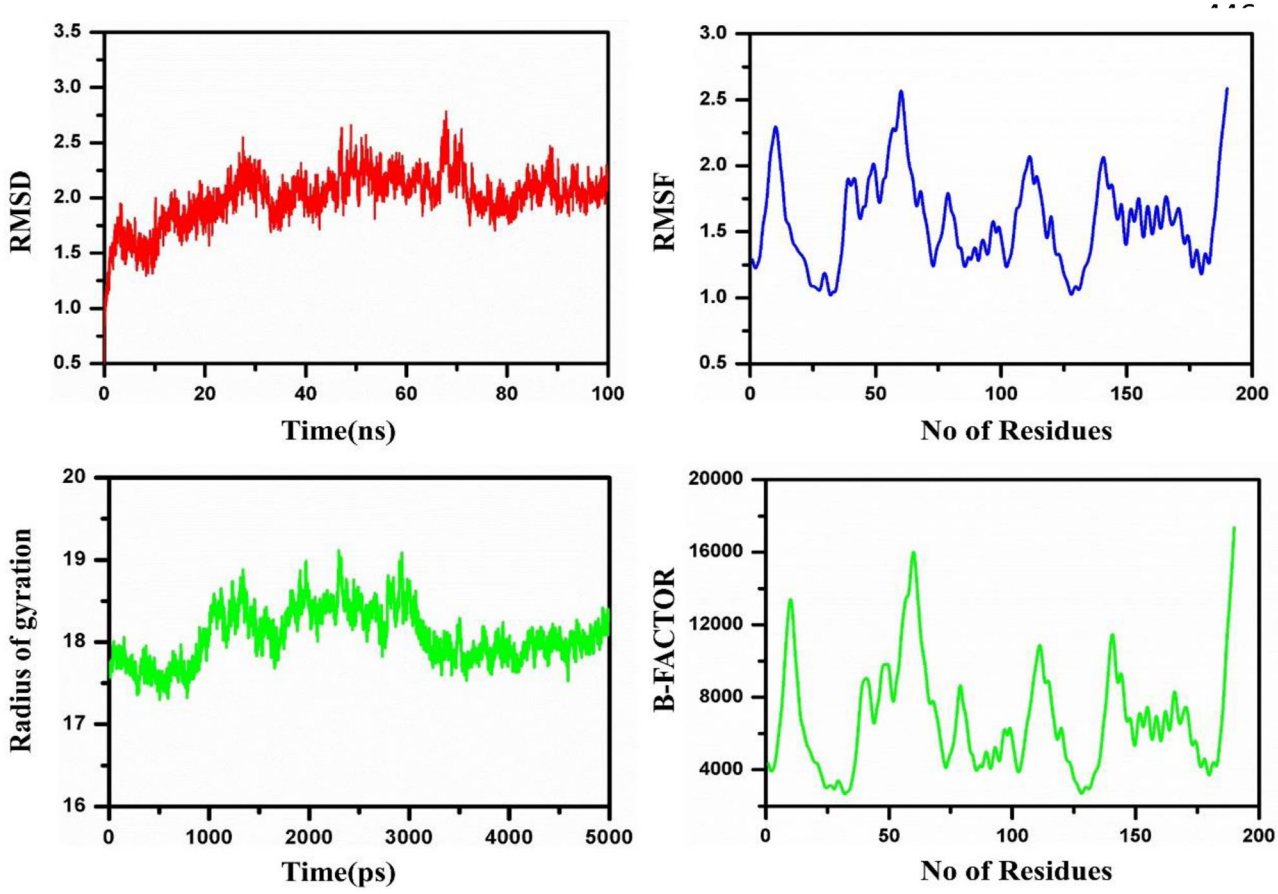

Figure 5. Molecular dynamics simulation study of HCV core protein structure. (A) Root mean square deviation of the core protein for the time duration of 100 ns. (B) Root mean square fluctuation representation of the core protein side chains for the same time duration. (C) and (D) are showing the radius of gyration and B-factor of the core protein structure.

Table 6. Cytotoxic T-lymphocyte (CTL) epitopes (viral epitopes) predicted by NetCTL.

Core peptide	Peptide start	Peptide end	Percentile rank
RLGVRATRK	42	50	0.2
DLMGYIPLV	131	139	0.3
FSIFLLALF	173	181	0.36
DPRRRSRNL	110	118	0.22
FLLALFSCL	176	184	0.3
VRATRKTSE	45	53	0.34
ALAHGVRAL	149	157	0.38
FLLALFSCL	176	184	0.4
SWAQPGYPW	74	82	0.54
VLPRRGPRL	35	43	0.58
FSIFLLALF	173	181	0.72
IYVGVYVLP	29	37	0.72
GGAARALAH	144	152	0.74
PNDPRRRSR	108	116	0.85
YIPLVGAPL	135	143	1

evaluation servers are tabulated in Table 4. Model 3 was selected as the final structure for further analyses. The ERRAT plot and *Z*-score analysis using ProSA-web of model 3 is given in Figure 3. The models generated through Robetta servers are given in Figure 4.

Impact of the substitutions on the structure of protein

The mCSM server predicted the stability index of each substitution. The more the negative score the more destability. Among the total of 11 substitutions, 9 were reported to destabilize the structure while the only two amino acid substitutions stabilized the structure. The reported generated by the mCSM server including the RSA and predicted DDG is given in Table 5.

Dynamics of the modelled 3D structure

To verify the stability and test the residual fluctuation, the final structure of the core protein (model 3) was subjected to MD simulation. RMSF and RMSD were calculated for protein structure and given in Figure 5. For the time period of 100 ns, RMSD of the protein and RMSF of all the residues of the side chain were examined to determine whether the structure is stable or not. RMSD of the system was 2.0 Å. RMSF was found within acceptable range excluding for a rare residue that showed more significant fluctuation due to long loops. On the other hand, the radius of gyration and B-factor are also in uniformity with the RMSD and RMSF results showing compactness and flexibility of the protein residues. This structure was used as input for all the other processes.

CTL and HTL epitopes prediction

NetCTL1.2 server was used for CTL epitopes prediction. The server predicted 17 T-cell epitopes, among which nine

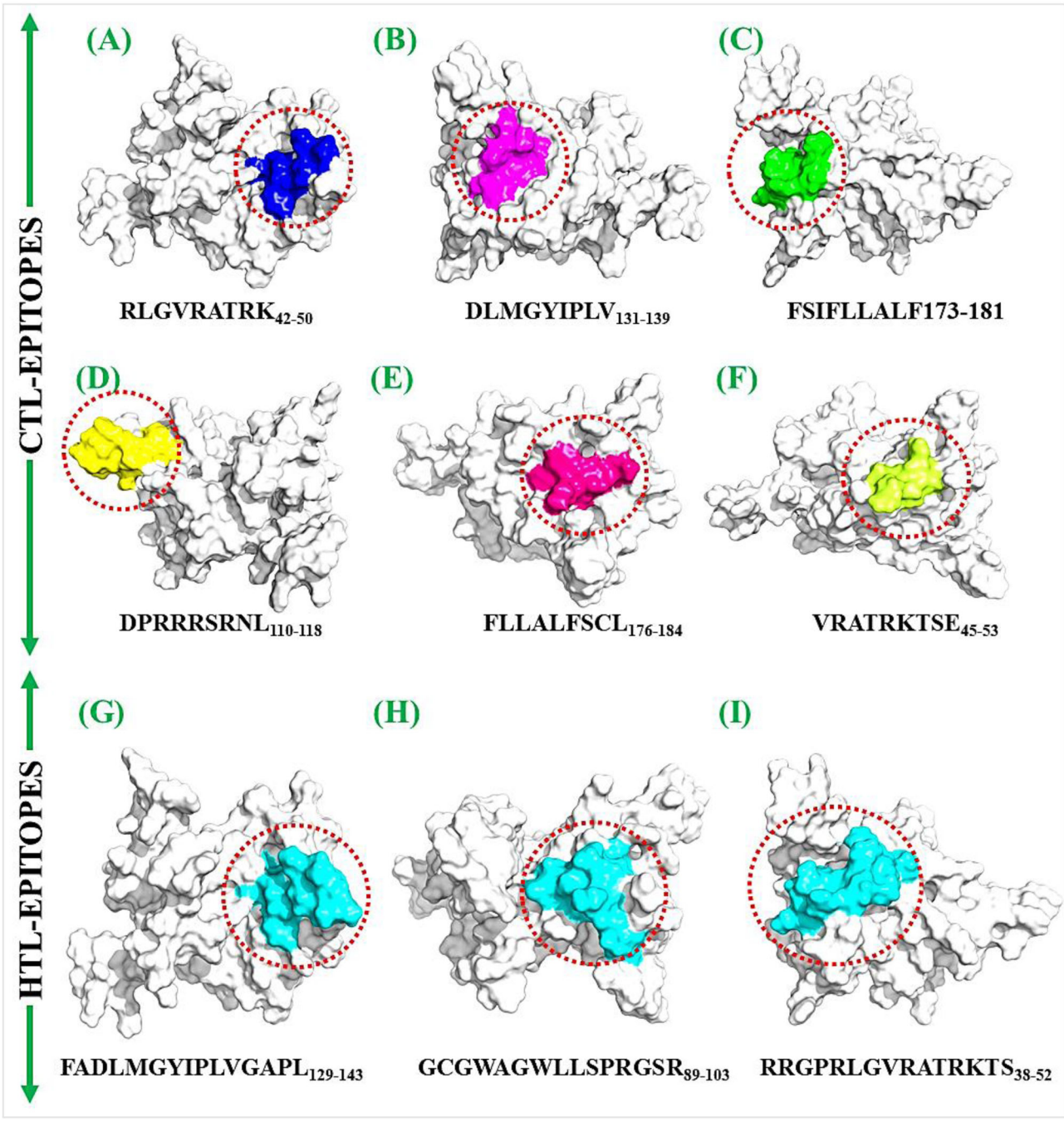

Figure 6. The predicted T-cell and HTL epitopes using the defined threshold. (A) $RLGVRATRK_{42–50}$, (B) $DLMGYIPLV_{131–139}$, (C) $FSIFLLALF_{173–181}$, (D) $DPRRRSRNL_{110–118}$, (E) $FLLALFSCL_{176–184}$ and (F) $VRATRKTSE_{45–53}$ are CTL epitopes, whereas (G) $FADLMGYIPLVGAPL_{129–143}$, (H) $GCGWAGWLLSPRGSR_{89–103}$ and (I) $RRGPRLGVRATRKTS_{38–52}$ are the predicted HTL epitopes.

Table 7. Predicted epitopes, which can provoke immune responses.

S. No.	Epitope name	Sequence	Method	Result	Score
1	Epitope_1	RLGVRATRK	MERCI	Positive	2
2	Epitope_3	FSIFLLALF	MERCI	Positive	1
3	Epitope_4	DPRRRSRNL	MERCI	Positive	13
4	Epitope_6	VRATRKTSE	MERCI	Positive	1
5	Epitope_8	GCGWAGWLLSPRGSR	MERCI	Positive	5
6	Epitope_9	RRGPRLGVRATRKTS	MERCI	Positive	2
7	Epitope_2	DLMGYIPLV	SVM	Positive	0.53463897
8	Epitope_5	FLLALFSCL	SVM	Positive	0.50347475
9	Epitope_7	FADLMGYIPLVGAPL	SVM	Positive	0.34206139

Epitope number 1–6 is CTL epitope, whereas number 7–9 are HTL epitope.

epitopes were selected satisfying the criteria of percentile rank score >0.5. Table 6 shows the CTL epitopes.

In the same way, IEDB MHC-II module was used for prediction of HTL epitopes. Epitopes for the query protein was predicted against seven human HLAs given as HLA-DRB1*03:01, HLA-DRB1*07:01, HLA-DRB1*15:01, HLA-DRB3*01:01, HLA-DRB3*02:02, HLA-DRB4*01:01, HLA-DRB5*01:01. These epitopes were predicted and were positioned at different locations within core protein RLGVRATRK$_{42–50}$, DLMGYIPLV$_{131–139}$, FSIFLLALF$_{173–181}$, DPRRRSRNL$_{110–118}$, FLLALFSCL$_{176–184}$, VRATRKTSE$_{45–53}$, ALAHGVRAL$_{149–157}$, FLLALFSCL$_{176–184}$ and FSIFLLALF$_{173–181}$, respectively. Three epitopes were selected based on percentile rank, their non-allergenic nature and antigenicity. Duplicates were removed as they were predicted against different HLAs. The epitopes selected to construct the final vaccine are shown in Figure 6.

IFN-gamma inducing peptide predictions

The predicted epitopes were screened for immune response effect using IFNepitope server, which predicted that the epitopes we predicted are potential to provoke the immune response and the results are tabulated in Table 7.

Table 8. B-cell epitope prediction by ABCpred prediction server.

Rank	Sequence	Start position	Score
1	QPGYPWPLYGNEGCGW	77	0.92
2	SRPNWAPNDPRRRSRN	102	0.90
3	AGWLLSPRGSRPNWAP	93	0.88
3	TRKTSERSQPRGRRKP	48	0.88
4	GGVIYVGVYVLPRRGP	26	0.87
5	YVLPRRGPRLGVRATR	34	0.86
5	IPLVGAPLGGAARALA	136	0.86
6	QPRGRRKPIPKARRSE	56	0.79
7	LPKPQRKTKRNTIRRP	3	0.78
8	PIPKARRSEGRSWAQP	63	0.77
9	LAHGVRALEDGINFAT	150	0.75
9	LGKVIDTLTCGFADLM	118	0.75
10	RLGVRATRKTSERSQP	42	0.74
10	DPRRRSRNLGKVIDTL	110	0.74
11	PQDVKFPGGGVIYVGV	18	0.73
12	EDGINFATGNLPGCSF	158	0.72
13	LGGAARALAHGVRALE	143	0.71
14	TLTCGFADLMGYIPLV	124	0.70
15	SFSIFLLALFSCLIHP	172	0.64
16	KTKRNTIRRPQDVKFP	9	0.60

B-cell epitope prediction

An online server ABCpred was used for prediction of B-cell epitopes (Saha & Raghava, 2006). Sixteen epitopes were predicted having a score higher than 0.5. The epitopes obtained from this server are tabulated in Table 8.

Vaccine construct

A vaccine sequence was constructed by joining the only CTL epitope to the rest of HTL epitopes. AAY linker was used to join CTL epitopes to the HTL epitopes. The same linker was placed between the HTL epitopes as well. Besides, 45-amino-acid adjuvant β-defensin (GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCRRKKK) was introduced with the aid of another linker (EAAAK) at the *N*-terminal of the vaccine structure. The schematic representation of the final vaccine construct is given in Figure 7. With the addition of the adjuvant and linkers, the final length of the vaccine candidate was 167AA including 45AA an adjuvant.

Antigenicity and allergenicity

The antigenicity of the final vaccine structure was evaluated at a threshold of 0.9882% using the VaxiJen server (Doytchinova & Flower, 2008) by selected bacteria model. The score indicates the antigenic nature of the sequence. The final vaccine structure with a score of −0.67 was observed to be non-allergenic, while the default threshold was −0.4.

Molecular docking of vaccine with T-cell receptors (TLR-8 and TLR-3)

Molecular docking was performed against TLR-8 (PDB ID: 3W3J) and TLR-3 (PDB ID: 2A0Z) shown in Figure 8. The

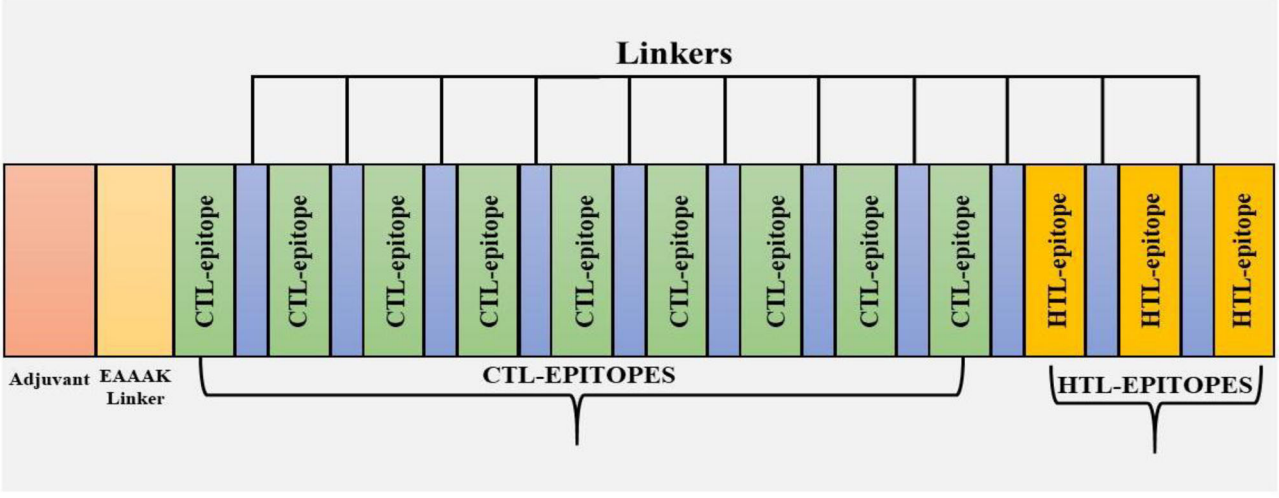

Figure 7. Schematic illustration of the vaccine construct consists of epitopes joined together by linkers and an adjuvant. A total of nine CTL epitopes joined by the AAY linkers, whereas the only three epitopes are joined by the GPGPG linker. The adjuvant was added through the EAAK linker.

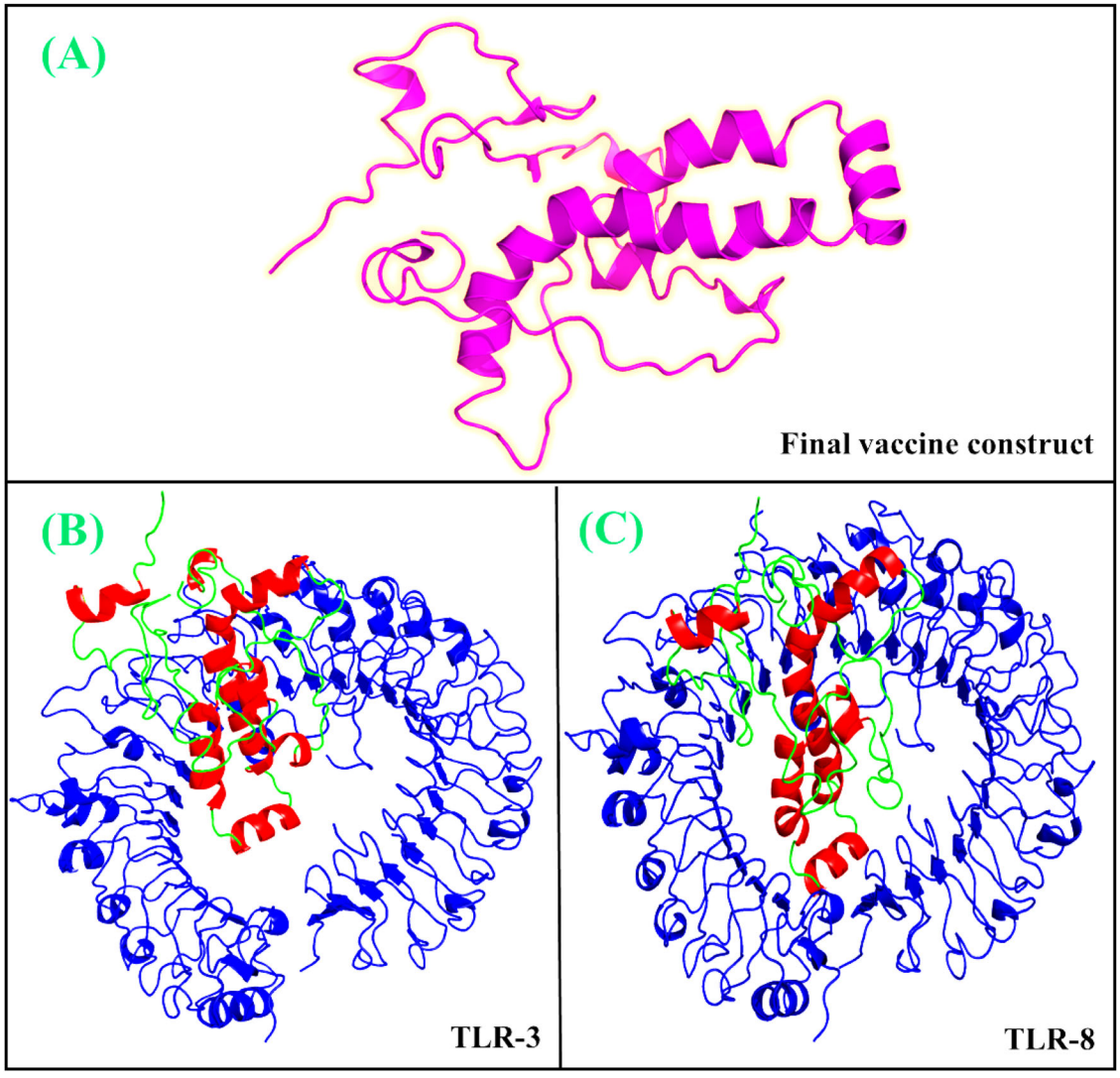

Figure 8. (A) Final vaccine construct docked against TLR-3 and TLR-8. (B) TLR-3 and (C) TLR-8 (receptors) are shown in blue, whereas red and green colours represent the multi-epitope subunit vaccine.

server predicted 10 interaction models based on the protein surface's geometry and electrostatic complementarity. To refine and re-score molecular docking solutions, the FireDock (Fast Interaction Refinement in Molecular Docking) server was used (Mashiach et al., 2008). The best scoring model was selected and the server predicted van der Waals (−46.81), partial electrostatics (20.73) and additional estimations of the binding free energy (−35.09) for TLR-3 while for TLR-8 van der Waals interactions (−27.89), partial electrostatics (3.0) and additional estimations of the binding free energy (−17.22) for TLR-8 was reported. Docking of TLR-3 revealed hydrogen bonds such as LYS200[NZ] with LEU21[O] at a distance of 2.85Ð, ASN230[ND2] with LEU98[O] at a distance of 2.42Ð, TYR383[OH] with GLY147[O] at a distance of 2.22Ð, HIS359[ND1] with ARG149[O] at a distance of 3.54Ð and ASN230[OD1] with ALA100[N] at a distance of 3.68Ð. In case of docking with TLR-3 revealed multiple hydrogen bonding such as ARG696[NH1] with ALA99[O] at a distance of 2.68Ð, ARG643[NH2] with SER103[OG] at a distance of 2.71Ð and CYS267[O] form and hydrogen bond with LYS50[NZ] of 2.56Ð, respectively. The docked models of the vaccine construct with TLR-3 TLR-8 are given in Figure 8.

Codon optimization and *in silico* cloning

For the optimization of codons for *E. coli* strain K12, Java codon adaptation tool was used. After optimization of codons, the nucleotide length was 207, CAI of optimized nucleotide sequence was 0.93, and the average GC content of the sequence was 64.25% showing better expression possibility of the protein in its host *E. coli*. Between 30% and 70%, the percentage range of GC content considered optimal. Finally, restriction cloning was carried out using Snap gene tool after addition of XhoI and NdeI restriction sites in pET28a(+) vector shown in Figure 9.

Discussion

HCV infection is a global issue and has a considerable effect on morbidity and mortality due to genetic variability makes

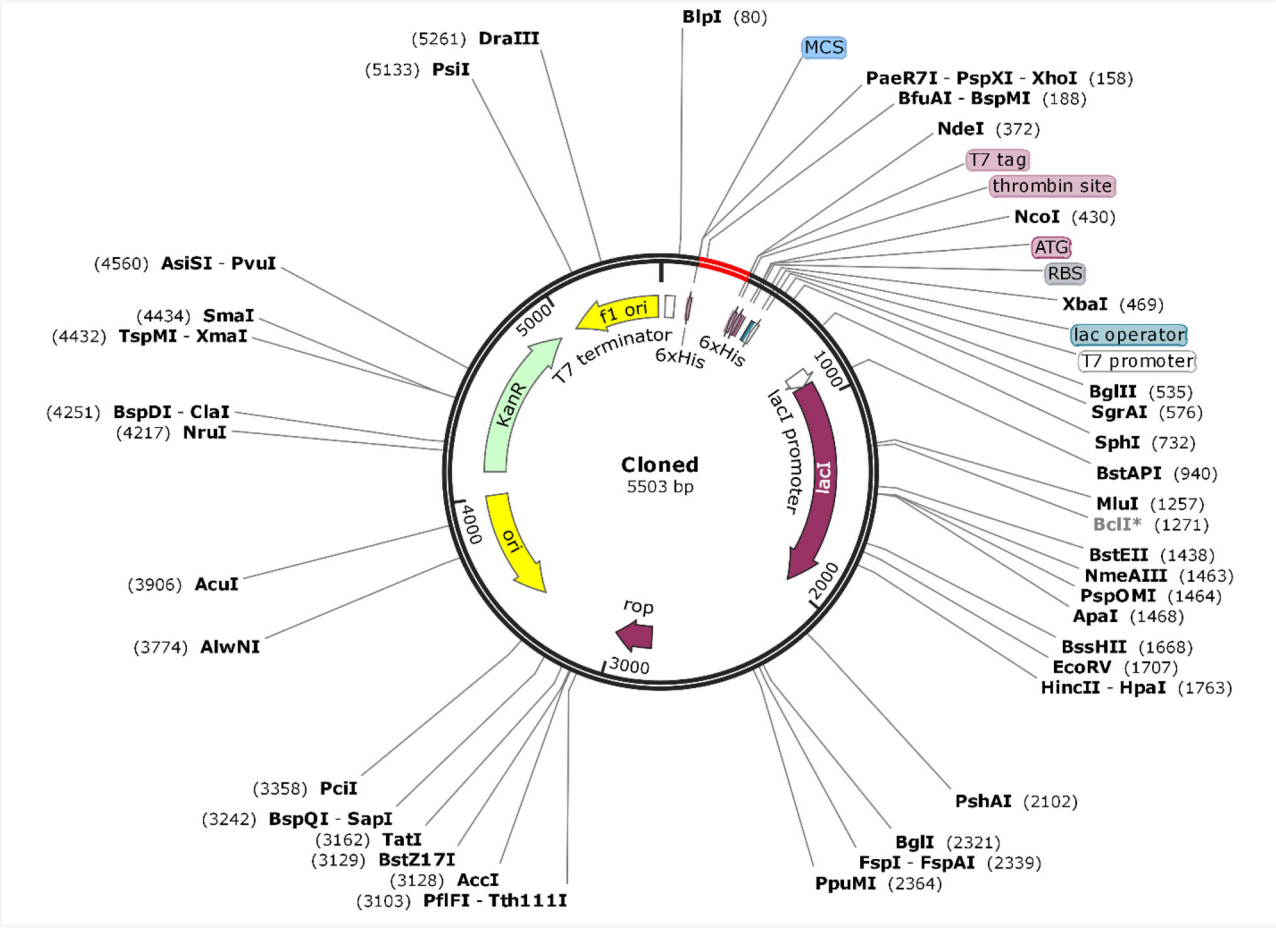

Figure 9. *In silico* restriction cloning of the final vaccine construct into pET28a(+) expression vector where purple part representing the vaccine insert and black circle showing the vector.

this virus challenging to treat. The 69th world health assembly in 2016 ratified the Global Health Sector Strategy (GHSS) to eradicate HCV by 2030, and introduced global target for the management and care of HCV infection by reducing 90.00% new cases, leading to 65.00% reduction in mortality, and providing access to affordable treatment to 80.00% of chronic HCV infection (World Health Organization, 2017). All over the world analysis of the clinical profile of HCV-positive patients, knowledge about genotype diversity, distribution pattern is imperative for the development of a strategy to control and treat HCV infection.

Distribution pattern of HCV genotype varies among different nations and regions within the same country. In this study, 3a genotype was found to be most common followed by untypable. The predominance of genotype 3a (40%) in the current study is in accordance with previous studies (H. Aziz et al., 2013; S. Khan, Attaullah, et al., 2011). Our finding is also confirmed by other studies conducted, and its associates were found genotype 3a is the most prevalent genotype in Pakistan (Al Kanaani et al., 2018). After 3a genotype untypable sample (29.2%) make second-highest proportion in this study and these findings are in agreement with the previously reported study of Afzal et al. (2014). For elimination of HCV infection in Pakistan, there is dire need to resolve the mystery of untypable genotypes through sequencing-based approach of genotype and molecular cloning techniques to fulfil the limitations in old diagnostic assays (M. Ali et al., 2014).

Core gene of HCV 3a isolates was sequenced and analysed for phylogenetic and evolutionary relationships (Kichatova et al., 2018). Phylogenetic analysis of nucleotide sequences of core gene with each other shows that our isolates clustered into different groups/clusters and our isolates are clustered with the isolates from Iran, the USA, China, India and Canada. This homology may be due to migration of Pakistani peoples for labour purpose and trade.

As HCV core is a structured gene; it is mostly conserved in HCV isolates reported from different countries worldwide. However, surprisingly core sequences (genotype 3a) are evolutionarily diverse and have several amino acid substitutions. These amino acid substitutions may play a significant role in response to antiviral therapies and disease severity. For example, R70Q is linked to non-responding antiviral therapy in genotype 1b (Alestig et al., 2011). According to Alhamlan and colleagues, the variation in core region at position 70, it was statistically significant keeping in view response to antiviral therapy (Alhamlan et al., 2014). According to Kichatova and co-workers, R70Q/H and L91M in core protein of HCV subtype 1b can affect the response to IFN-based therapies (Kichatova et al., 2018). We observed several amino acid substitutions R46C, R70Q, L91C, G60E, N/S105A, P108A, N110I, S116V, G90S, A77G and G145R that could be linked with

response to antiviral therapies. As epitopes predicted from wild core protein amino acids sequence, by comparing our reported substitution with predicted epitopes sequences, we also observed these substitutions at epitopic level. However, further studies are required in this direction.

Specific T-cell responses to HCV non-structural proteins are produced during acute HCV infection, whereas during chronic infection this responses drifts to structural HCV proteins (Blackard et al., 2008). Therefore, in our study, only structural protein (core protein) was included in the multi-epitopic vaccine design to focus on the T-cell reactivity that is related to the chronic resolution of HCV. A multi-epitopic vaccine designed with such treatment could thus become a useful tool in the fight against tumours and viral infections (A. Ali et al., 2019; A. Khan, Ali, Khan, et al., 2020; A. Khan, Junaid, Kaushik, et al., 2018; A. Khan, Junaid, Li, et al., 2020; A. Khan, Khan, Saleem, Babar, et al., 2020; A. Khan, Khan, Saleem, Junaid, et al., 2020; M. Khan, Khan, Ali, et al., 2019; M. T. Khan, Ali, Wang, et al., 2020; M. T. Khan, Khan, Rehman, et al., 2019). In this study, we also reported a multi-epitope subunit vaccine designed from our own core protein amino acids sequence.

Allergenicity is among the key problems since the development of the vaccine. Evaluation of allergenicity is therefore vital at an earlier stage of developing the vaccine. Before developing the final vaccine construct, the screened epitopes were first subjected to an allergenicity assessment. In addition, the allergenicity of the vaccine's final structure was also investigated and found to be non-allergenic. Thus, the final vaccine construct was designed after a rigorous *in silico* analysis. Various trials have revealed that strong TLR3 and TLR8 agonists in HCV-positive patients reduce the level of HCV RNA (C.-H. Wang et al., 2011; N. Wang et al., 2009). To analyse the immune response of TLR-3 and TLR-8 to our vaccine construct, we performed docking assessment. This analysis indicates that the proposed vaccine may produce a particular immune response to HCV.

Conclusion

The current study reported novel mutations in the core protein of the HCV. These substitutions significantly altered the evolutionary relationship and show that emerging viral isolates are circulating in Pakistan. Among the reported mutations some mutations are drug-resistant and destabilizing the protein structure. The substitutions were also observed in predicted epitopes. Multi-epitopic vaccine was constructed using core protein HCV-3a. *In silico* analysis of proposed vaccine shows it may produce an immune response to HCV. This study presents a new paradigm for understanding the emerging HCV strains and also reported a potential vaccine candidate which could aid the therapeutic choices against it.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The study is supported by funds provided by TWAS (Ref. No. 17-278 RG/PHA/AS_C-FR3240300076). Dong-Qing Wei is supported by the grants from the Key Research Area Grant 2016YFA0501703 of the Ministry of Science and Technology of China, the National Natural Science Foundation of China (Contract no. 61832019, 61503244), the Natural Science Foundation of Henan Province (162300410060) and Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2017ZD14). The computations were partially performed at the Center for High-Performance Computing, Shanghai Jiao Tong University.

References

Abaitua, F., Rodríguez, J. R., Garzón, A., Rodríguez, D., & Esteban, M. (2006). Improving recombinant MVA immune responses: Potentiation of the immune responses to HIV-1 with MVA and DNA vectors expressing *Env* and the cytokines IL-12 and IFN-gamma. *Virus Research*, *116*(1–2), 11–20. https://doi.org/10.1016/j.virusres.2005.08.008

Afridi, S. Q., Zahid, M. N., Shabbir, M. Z., Hussain, Z., Mukhtar, N., Tipu, M. Y., Akhtar, F., & Yaqub, T. (2013). Prevalence of HCV genotypes in district Mardan. *Virology Journal*, *10*(1), 90. https://doi.org/10.1186/1743-422X-10-90

Afzal, M. S., Anjum, S., & Zaidi, N. U. S. S. (2013). Effect of functional interleukin-10 polymorphism on pegylated interferon-α plus ribavirin therapy response in chronic hepatitis C virus patients infected with 3a genotype in Pakistani population. *Hepatitis Monthly*, *13*(6), 1–3. https://doi.org/10.5812/hepatmon.10274

Afzal, M. S., Khan, M. Y., Ammar, M., Anjum, S., & Zaidi, N. U. S. S. (2014). Diagnostically untypable hepatitis C virus variants: It is time to resolve the problem. *World Journal of Gastroenterology*, *20*(46), 17690–17692. https://doi.org/10.3748/wjg.v20.i46.17690

Akuta, N., Suzuki, F., Sezaki, H., Suzuki, Y., Hosaka, T., Someya, T., Kobayashi, M., Saitoh, S., Watahiki, S., Sato, J., Matsuda, M., Kobayashi, M., Arase, Y., Ikeda, K., & Kumada, H. (2005). Association of amino acid substitution pattern in core protein of hepatitis C virus genotype 1b high viral load and non-virological response to interferon-ribavirin combination therapy. *Intervirology*, *48*(6), 372–380. https://doi.org/10.1159/000086064

Al Kanaani, Z., Mahmud, S., Kouyoumjian, S. P., & Abu-Raddad, L. J. (2018). The epidemiology of hepatitis C virus in Pakistan: Systematic review and meta-analyses. *Royal Society Open Science*, *5*(4), 180257. https://doi.org/10.1098/rsos.180257

Alestig, E., Arnholm, B., Eilard, A., Lagging, M., Nilsson, S., Norkrans, G., Wahlberg, T., Wejstål, R., Westin, J., & Lindh, M. (2011). Core mutations, IL28B polymorphisms and response to peginterferon/ribavirin treatment in Swedish patients with hepatitis C virus genotype 1 infection. *BMC Infectious Diseases*, *11*(1), 124. https://doi.org/10.1186/1471-2334-11-124

Alhamlan, F. S., Al-Ahdal, M. N., Khalaf, N. Z., Abdo, A. A., Sanai, F. M., Al-Ashgar, H. I., ElHefnawi, M., Zaid, A., & Al-Qahtani, A. A. (2014). Genetic variability of the core protein in hepatitis C virus genotype 4 in Saudi Arabian patients and its implication on pegylated interferon and ribavirin therapy. *Journal of Translational Medicine*, *12*(1), 91. https://doi.org/10.1186/1479-5876-12-91

Ali, A., Khan, A., Kaushik, A. C., Wang, Y., Ali, S. S., Junaid, M., Saleem, S., Cho, W. C. S., Mao, X., & Wei, D.-Q. (2019). Immunoinformatic and systems biology approaches to predict and validate peptide vaccines against Epstein–Barr virus (EBV). *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-018-37070-z

Ali, M., Rehman, I. U., & Idrees, M. (2014). Emergence of genetically variant hepatitis C virus population in response to increased antiviral drug pressure, Pakistan. *Virus Genes*, *48*(3), 543–549. https://doi.org/10.1007/s11262-014-1047-y

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A.,

Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., & Stockinger, H. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research*, *40*(W1), W597–W603. https://doi.org/10.1093/nar/gks400

Attaullah, S., Khan, S., & Ali, I. (2011). Hepatitis C virus genotypes in Pakistan: A systemic review. *Virology Journal*, *8*(1), 433. https://doi.org/10.1186/1743-422X-8-433

Aziz, H., Raza, A., Murtaza, S., Waheed, Y., Khalid, A., Irfan, J., Samra, Z., & Athar, M. A. (2013). Molecular epidemiology of hepatitis C virus genotypes in different geographical regions of Punjab Province in Pakistan and a phylogenetic analysis. *International Journal of Infectious Diseases*, *17*(4), e247–e253. https://doi.org/10.1016/j.ijid.2012.09.017

Aziz, S., Khanani, R., Noorulain, W., & Rajper, J. (2010). Frequency of hepatitis B and C in rural and periurban Sindh. *JPMA - Journal of the Pakistan Medical Association*, *60*(10), 853.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., & Schwede, T. (2014). SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, *42*(W1), W252–W258. https://doi.org/10.1093/nar/gku340

Blackard, J. T., Shata, M. T., Shire, N. J., & Sherman, K. E. (2008). Acute hepatitis C virus infection: A chronic problem. *Hepatology*, *47*(1), 321–331. https://doi.org/10.1002/hep.21902

Boulant, S., Vanbelle, C., Ebel, C., Penin, F., & Lavergne, J.-P. (2005). Hepatitis C virus core protein is a dimeric alpha-helical protein exhibiting membrane protein features. *Journal of Virology*, *79*(17), 11353–11365. https://doi.org/10.1128/JVI.79.17.11353-11365.2005

Bowie, J. U., Luthy, R., & Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, *253*(5016), 164–170. https://doi.org/10.1126/science.1853201

Chen, Z.-W., Li, H., Ren, H., & Hu, P. (2016). Global prevalence of pre-existing HCV variants resistant to direct-acting antiviral agents (DAAs): Mining the GenBank HCV genome data. *Scientific Reports*, *6*(1), 20310. https://doi.org/10.1038/srep20310

Choo, Q.-L., Kuo, G., Weiner, A. J., Overby, L. R., Bradley, D. W., & Houghton, M. (1989). Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science*, *244*(4902), 359–362. https://doi.org/10.1126/science.2523562

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science*, *2*(9), 1511–1519. https://doi.org/10.1002/pro.5560020916

Csete, J., Elliott, R., & Fischer, B. (2008). *"Viral Time Bomb": Health and human rights challenges in addressing hepatitis C in Canada*. Canadian HIV/AIDS Legal Network, Centre for Addictions Research of BC and Centre for Applied Research in Mental Health and Addiction.

Dhanda, S. K., Vir, P., & Raghava, G. P. (2013). Designing of interferon-gamma inducing MHC class-II binders. *Biology Direct*, *8*(1), 30. https://doi.org/10.1186/1745-6150-8-30

Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, *8*(1), 4. https://doi.org/10.1186/1471-2105-8-4

Doytchinova, I. A., & Flower, D. R. (2008). Bioinformatic approach for identifying parasite and fungal candidate subunit vaccines. *The Open Vaccine Journal*, *1*(1), 4.

Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, *7*(1), 214. https://doi.org/10.1186/1471-2148-7-214

El-Manzalawy, Y., Dobbs, D., & Honavar, V. (2008). Predicting linear B-cell epitopes using string kernels. *Journal of Molecular Recognition*, *21*(4), 243–255. https://doi.org/10.1002/jmr.893

Golden-Mason, L., & Rosen, H. R. (2006). Natural killer cells: Primary target for hepatitis C virus immune evasion strategies? *Liver Transplantation*, *12*(3), 363–372. https://doi.org/10.1002/lt.20708

Grebely, J., Page, K., Sacks-Davis, R., van der Loeff, M. S., Rice, T. M., Bruneau, J., Morris, M. D., Hajarizadeh, B., Amin, J., Cox, A. L., Kim, A. Y., McGovern, B. H., Schinkel, J., George, J., Shoukry, N. H., Lauer, G. M., Maher, L., Lloyd, A. R., Hellard, M., Dore, G. J., Prins, M., & InC3 Study Group. (2014). The effects of female sex, viral genotype, and *IL28B* genotype on spontaneous clearance of acute hepatitis C virus infection. *Hepatology*, *59*(1), 109–120. https://doi.org/10.1002/hep.26639

Green, D. S., Nunes, A. T., Annunziata, C. M., & Zoon, K. C. (2016). Monocyte and interferon based therapy for the treatment of ovarian cancer. *Cytokine & Growth Factor Reviews*, *29*, 109–115. https://doi.org/10.1016/j.cytogfr.2016.02.006

Hiramatsu, N., Oze, T., & Takehara, T. (2015). Suppression of hepatocellular carcinoma development in hepatitis C patients given interferon-based antiviral therapy. *Hepatology Research*, *45*(2), 152–161. https://doi.org/10.1111/hepr.12393

Hoofnagle, J. H. (2002). Course and outcome of hepatitis C. *Hepatology*, *36*(S1), S21–S29. https://doi.org/10.1002/hep.1840360704

Hunt, D., & Pockros, P. (2013). What are the promising new therapies in the field of chronic hepatitis C after the first-generation direct-acting antivirals? *Current Gastroenterology Reports*, *15*(1), 303. https://doi.org/10.1007/s11894-012-0303-3

Idrees, M. (2008). Development of an improved genotyping assay for the detection of hepatitis C virus genotypes and subtypes in Pakistan. *Journal of Virological Methods*, *150*(1–2), 50–56. https://doi.org/10.1016/j.jviromet.2008.03.001

Idrees, M., & Riazuddin, S. (2008). Frequency distribution of hepatitis C virus genotypes in different geographical regions of Pakistan and their possible routes of transmission. *BMC Infectious Diseases*, *8*(1), 69. https://doi.org/10.1186/1471-2334-8-69

Ikram, A., Zaheer, T., Awan, F. M., Obaid, A., Naz, A., Hanif, R., Paracha, R. Z., Ali, A., Naveed, A. K., & Janjua, H. A. (2018). Exploring NS3/4A, NS5A and NS5B proteins to design conserved subunit multi-epitope vaccine against HCV utilizing immunoinformatics approaches. *Scientific Reports*, *8*(1), 16107. https://doi.org/10.1038/s41598-018-34254-5

James, C. M., Abdad, M. Y., Mansfield, J. P., Jacobsen, H. K., Vind, A. R., Stumbles, P. A., & Bartlett, E. J. (2007). Differential activities of alpha/beta IFN subtypes against influenza virus in vivo and enhancement of specific immune responses in DNA vaccinated mice expressing haemagglutinin and nucleoprotein. *Vaccine*, *25*(10), 1856–1867. https://doi.org/10.1016/j.vaccine.2006.10.038

Jiang, W., Ren, L., & Jin, N. (2007). HIV-1 DNA vaccine efficacy is enhanced by coadministration with plasmid encoding IFN-alpha. *Journal of Virological Methods*, *146*(1–2), 266–273. https://doi.org/10.1016/j.jviromet.2007.07.016

Jolivet-Reynaud, C., Dalbon, P., Viola, F., Yvon, S., Paranhos-Baccala, G., Piga, N., Bridon, L., Trabaud, M. A., Battail, N., Sibai, G., & Jolivet, M. (1998). HCV core immunodominant region analysis using mouse monoclonal antibodies and human sera: Characterization of major epitopes useful for antigen detection. *Journal of Medical Virology*, *56*(4), 300–309. https://doi.org/10.1002/(SICI)1096-9071(199812)56:4<300::AID-JMV3>3.0.CO;2-8

Khan, A., Ali, S. S., Khan, M. T., Saleem, S., Ali, A., Suleman, M., Babar, Z., Shafiq, A., Khan, M., & Wei, D.-Q. (2020). Combined drug repurposing and virtual screening strategies with molecular dynamics simulation identified potent inhibitors for SARS-CoV-2 main protease (3CLpro). *Journal of Biomolecular Structure and Dynamics*, *38*, 1–12.

Khan, A., Junaid, M., Kaushik, A. C., Ali, A., Ali, S. S., Mehmood, A., & Wei, D.-Q. (2018). Computational identification, characterization and validation of potential antigenic peptide vaccines from hrHPVs E6 proteins using immunoinformatics and computational systems biology approaches. *PLoS One*, *13*(5), e0196484. https://doi.org/10.1371/journal.pone.0196484

Khan, A., Junaid, M., Li, C.-D., Saleem, S., Humayun, F., Shamas, S., Ali, S. S., Babar, Z., & Wei, D.-Q. (2020). Dynamics insights into the gain of flexibility by Helix-12 in ESR1 as a mechanism of resistance to drugs in breast cancer cell lines. *Frontiers in Molecular Biosciences*, *6*, 159. https://doi.org/10.3389/fmolb.2019.00159

Khan, A., Khan, M., Saleem, S., Babar, Z., Ali, A., Khan, A. A., Sardar, Z., Hamayun, F., Ali, S. S., & Wei, D.-Q. (2020). Phylogenetic analysis and structural perspectives of RNA-dependent RNA-polymerase inhibition from SARs-CoV-2 with natural products. *Interdisciplinary Sciences: Computational Life Sciences*, *12*(3), 335–314. https://doi.org/10.1007/s12539-020-00381-9

Khan, A., Khan, M. T., Saleem, S., Junaid, M., Ali, A., Shujait Ali, S., Khan, M., & Wei, D.-Q. (2020). Structural Insights into the mechanism of RNA recognition by the N-terminal RNA-binding domain of the SARS-CoV-2 nucleocapsid phosphoprotein. *Computational and Structural Biotechnology Journal*, *18*, 2174–2184. https://doi.org/10.1016/j.csbj.2020.08.006

Khan, M., Khan, S., Ali, A., Akbar, H., Sayaf, A. M., Khan, A., & Wei, D.-Q. (2019). Immunoinformatics approaches to explore *Helicobacter pylori* proteome (virulence factors) to design B and T cell multi-epitope subunit vaccine. *Scientific Reports*, *9*(1), 1–13. https://doi.org/10.1038/s41598-019-49354-z

Khan, M. T., Ali, A., Wang, Q., Irfan, M., Khan, A., Zeb, M. T., Zhang, Y. J., Chinnasamy, S., & Wei, D.-Q. (2020). Marine natural compounds as potents inhibitors against the main protease of SARS-CoV-2. A molecular dynamic study. *Journal of Biomolecular Structure and Dynamics, 38*, 1–14.

Khan, M. T., Khan, A., Rehman, A. U., Wang, Y., Akhtar, K., Malik, S. I., & Wei, D.-Q. (2019). Structural and free energy landscape of novel mutations in ribosomal protein S1 (rpsA) associated with pyrazinamide resistance. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-44013-9

Khan, S., Attaullah, S., Ayaz, S., Niaz Khan, S., Shams, S., Ali, I., Bilal, M., & Siraj, S. (2011). Molecular epidemiology of HCV among health care workers of Khyber Pakhtunkhwa. *Virology Journal*, *8*(1), 105. https://doi.org/10.1186/1743-422X-8-105

Khan, S. T., Karges, W., Cooper, C. L., & Crawley, A. M. (2018). Hepatitis C virus core protein reduces $CD8^+$ T-cell proliferation, perforin production and degranulation but increases STAT5 activation. *Immunology*, *154*(1), 156–165. https://doi.org/10.1111/imm.12882

Kichatova, V., Kyuregyan, K., Soboleva, N., Karlsen, A., Isaeva, O., Isaguliants, M., & Mikhailov, M. (2018). Frequency of interferon-resistance conferring substitutions in amino acid positions 70 and 91 of core protein of the Russian HCV 1b isolates analyzed in the T-cell epitopic context. *Journal of Immunology Research*, *2018*, 1–13. https://doi.org/10.1155/2018/7685371

Kim, D. E., Chivian, D., & Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research*, *32*(suppl_2), W526–W531. https://doi.org/10.1093/nar/gkh468

Kühl, U., Lassner, D., von Schlippenbach, J., Poller, W., & Schultheiss, H.-P. (2012). Interferon-Beta improves survival in enterovirus-associated cardiomyopathy. *Journal of the American College of Cardiology*, *60*(14), 1295–1296. https://doi.org/10.1016/j.jacc.2012.06.026

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Larsen, M. V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., & Nielsen, M. (2007). Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics*, *8*(1), 424. https://doi.org/10.1186/1471-2105-8-424

Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., & Thornton, J. M. (2018). PDBsum: Structural summaries of PDB entries. *Protein Science*, *27*(1), 129–134. https://doi.org/10.1002/pro.3289

Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR*, *8*(4), 477–486. https://doi.org/10.1007/BF00228148

Liang, T. J., Rehermann, B., Seeff, L. B., & Hoofnagle, J. H. (2000). Pathogenesis, natural history, treatment, and prevention of hepatitis C. *Annals of Internal Medicine*, *132*(4), 296–305. https://doi.org/10.7326/0003-4819-132-4-200002150-00008

Lin, Z., Zhang, J., Ma, X., Yang, S., Tian, N., Lin, X., Zhou, S., Liu, L., & Gao, Y. (2016). The role of interferon lambda 3 genetic polymorphisms in response to interferon therapy in chronic hepatitis B patients: An updated meta-analysis. *Hepatitis Monthly*, *16*(7), 1–8. https://doi.org/10.5812/hepatmon.37534

Manns, M. P., McHutchison, J. G., Gordon, S. C., Rustgi, V. K., Shiffman, M., Reindollar, R., Goodman, Z. D., Koury, K., Ling, M.-H., & Albrecht, J. K. (2001). Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: A randomised trial. *The Lancet*, *358*(9286), 958–965. https://doi.org/10.1016/S0140-6736(01)06102-5

Mashiach, E., Schneidman-Duhovny, D., Andrusier, N., Nussinov, R., & Wolfson, H. J. (2008). FireDock: A web server for fast interaction refinement in molecular docking. *Nucleic Acids Research*, *36*(suppl_2), W229–W232. https://doi.org/10.1093/nar/gkn186

McLauchlan, J. (2000). Properties of the hepatitis C virus core protein: A structural protein that modulates cellular processes. *Journal of Viral Hepatitis*, *7*(1), 2–14. https://doi.org/10.1046/j.1365-2893.2000.00201.x

Memarnejadian, A., Roohvand, F., Arashkia, A., Rafati, S., & Shokrgozar, M. A. (2009). Polytope DNA vaccine development against hepatitis C virus: A streamlined approach from in silico design to in vitro and primary in vivo analyses in BALB/c mice. *Protein and Peptide Letters*, *16*(7), 842–850. https://doi.org/10.2174/092986609788681788

Ménez, R., Bossus, M., Muller, B. H., Sibaï, G., Dalbon, P., Ducancel, F., Jolivet-Reynaud, C., & Stura, E. A. (2003). Crystal structure of a hydrophobic immunodominant antigenic site on hepatitis C virus core protein complexed to monoclonal antibody 19D9D6. *Journal of Immunology*, *170*(4), 1917–1924. https://doi.org/10.4049/jimmunol.170.4.1917

Moradpour, D., Penin, F., & Rice, C. M. (2007). Replication of hepatitis C virus. *Nature Reviews. Microbiology*, *5*(6), 453–463. https://doi.org/10.1038/nrmicro1645

Öhlschläger, P., Quetting, M., Alvarez, G., Dürst, M., Gissmann, L., & Kaufmann, A. M. (2009). Enhancement of immunogenicity of a therapeutic cervical cancer DNA-based vaccine by co-application of sequence-optimized genetic adjuvants. *International Journal of Cancer*, *125*(1), 189–198. https://doi.org/10.1002/ijc.24333

Ohno, O., Mizokami, M., Wu, R. R., Saleh, M. G., Ohba, K., Orito, E., Mukaide, M., Williams, R., & Lau, J. Y. (1997). New hepatitis C virus (HCV) genotyping system that allows for identification of HCV genotypes 1a, 1b, 2a, 2b, 3a, 3b, 4, 5a, and 6a. *Journal of Clinical Microbiology*, *35*(1), 201–207. https://doi.org/10.1128/JCM.35.1.201-207.1997

Okamoto, K., Mori, Y., Komoda, Y., Okamoto, T., Okochi, M., Takeda, M., Suzuki, T., Moriishi, K., & Matsuura, Y. (2008). Intramembrane processing by signal peptide peptidase regulates the membrane localization of hepatitis C virus core protein and viral propagation. *Journal of Virology*, *82*(17), 8349–8361. https://doi.org/10.1128/JVI.00306-08

Okazaki, K., Nishiyama, Y., Saitou, T., Shibata, N., Yamamoto, C., Oosaga, J., Tanaka, M., & Ichiyama, S. (2008). Fundamental evaluation of HCV core antigen method comparison with Cobas Amplicor HCV monitor v2. 0 (high range method). *Rinsho Byori. The Japanese Journal of Clinical Pathology*, *56*(2), 95–100.

Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., DeBolt, S., Ferguson, D., Seibel, G., & Kollman, P. (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications*, *91*(1–3), 1–41. https://doi.org/10.1016/0010-4655(95)00041-D

Perz, J. F., Armstrong, G. L., Farrington, L. A., Hutin, Y. J., & Bell, B. P. (2006). The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. *Journal of Hepatology*, *45*(4), 529–538. https://doi.org/10.1016/j.jhep.2006.05.013

Pires, D. E., Ascher, D. B., & Blundell, T. L. (2014). mCSM: Predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics*, *30*(3), 335–342. https://doi.org/10.1093/bioinformatics/btt691

Reed, K., & Rice, C. (2000). Overview of hepatitis C virus genome structure, polyprotein processing, and protein properties. In C. H. Hagedorn & C. M. Rice (Eds.), *The hepatitis C viruses. Current topics in microbiology and immunology* (Vol. 242, pp. 55–84). Springer.

Sabet, L. P., Taheri, T., Memarnejadian, A., Mokhtari Azad, T., Asgari, F., Rahimnia, R., Alavian, S. M., Rafati, S., & Samimi Rad, K. (2014). Immunogenicity of multi-epitope DNA and peptide vaccine candidates based on core, E2, NS3 and NS5B HCV epitopes in BALB/c mice. *Hepatitis Monthly*, *14*(10), 1–9. https://doi.org/10.5812/hepatmon.22215

Sabourin, M., Tuzon, C. T., Fisher, T. S., & Zakian, V. A. (2007). A flexible protein linker improves the function of epitope-tagged proteins in *Saccharomyces cerevisiae*. *Yeast*, *24*(1), 39–45. https://doi.org/10.1002/yea.1431

Saha, S., & Raghava, G. (2006). Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins*, *65*(1), 40–48. https://doi.org/10.1002/prot.21078

Sakamoto, M., Akahane, Y., Tsuda, F., Tanaka, T., Woodfield, D. G., & Okamoto, H. (1994). Entire nucleotide sequence and characterization of a hepatitis C virus of genotype V/3a. *Journal of General Virology*, *75*(7), 1761–1768. https://doi.org/10.1099/0022-1317-75-7-1761

Salomon-Ferrer, R., Case, D. A., & Walker, R. C. (2013). An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *3*(2), 198–210. https://doi.org/10.1002/wcms.1121

Scarselli, E., Ansuini, H., Cerino, R., Roccasecca, R. M., Acali, S., Filocamo, G., Traboni, C., Nicosia, A., Cortese, R., & Vitelli, A. (2002). The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. *The EMBO Journal*, *21*(19), 5017–5025. https://doi.org/10.1093/emboj/cdf529

Scheel, T. K., & Rice, C. M. (2013). Understanding the hepatitis C virus life cycle paves the way for highly effective therapies. *Nature Medicine*, *19*(7), 837–849. https://doi.org/10.1038/nm.3248

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2005). PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Research*, *33*(suppl_2), W363–W367. https://doi.org/10.1093/nar/gki481

Shavinskaya, A., Boulant, S., Penin, F., McLauchlan, J., & Bartenschlager, R. (2007). The lipid droplet binding domain of hepatitis C virus core protein is a major determinant for efficient virus assembly. *The Journal of Biological Chemistry*, *282*(51), 37158–37169. https://doi.org/10.1074/jbc.M707329200

Spengler, U. (2018). Direct antiviral agents (DAAs) – A new age in the treatment of hepatitis C virus infection. *Pharmacology & Therapeutics*, *183*, 118–126. https://doi.org/10.1016/j.pharmthera.2017.10.009

Toporovski, R., Morrow, M. P., & Weiner, D. B. (2010). Interferons as potential adjuvants in prophylactic vaccines. *Expert Opinion on Biological Therapy*, *10*(10), 1489–1500. https://doi.org/10.1517/14712598.2010.521495

Umer, M., & Iqbal, M. (2016). Hepatitis C virus prevalence and genotype distribution in Pakistan: Comprehensive review of recent data. *World Journal of Gastroenterology*, *22*(4), 1684–1700. https://doi.org/10.3748/wjg.v22.i4.1684

Vita, R., Overton, J. A., Greenbaum, J. A., Ponomarenko, J., Clark, J. D., Cantrell, J. R., Wheeler, D. K., Gabbard, J. L., Hix, D., Sette, A., & Peters, B. (2015). The immune epitope database (IEDB) 3.0. *Nucleic Acids Research*, *43*(D1), D405–D412. https://doi.org/10.1093/nar/gku938

Wang, C.-H., Eng, H.-L., Lin, K.-H., Chang, C.-H., Hsieh, C.-A., Lin, Y.-L., & Lin, T.-M. (2011). TLR7 and TLR8 gene variations and susceptibility to hepatitis C virus infection. *PLoS One*, *6*(10), e26235. https://doi.org/10.1371/journal.pone.0026235

Wang, N., Liang, Y., Devaraj, S., Wang, J., Lemon, S. M., & Li, K. (2009). Toll-like receptor 3 mediates establishment of an antiviral state against hepatitis C virus in hepatoma cells. *Journal of Virology*, *83*(19), 9824–9834. https://doi.org/10.1128/JVI.01125-09

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, *35*(suppl_2), W407–W410. https://doi.org/10.1093/nar/gkm290

World Health Organization. (2017). *Global hepatitis report 2017*. World Health Organization.

Xu, D., & Zhang, Y. (2013). Toward optimal fragment generations for ab initio protein structure assembly. *Proteins*, *81*(2), 229–239. https://doi.org/10.1002/prot.24179

Yasui, K., Wakita, T., Tsukiyama-Kohara, K., Funahashi, S. I., Ichikawa, M., Kajita, T., Moradpour, D., Wands, J. R., & Kohara, M. (1998). The native form and maturation process of hepatitis C virus core protein. *Journal of Virology*, *72*(7), 6048–6055. https://doi.org/10.1128/JVI.72.7.6048-6055.1998

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, *9*(1), 40. https://doi.org/10.1186/1471-2105-9-40