Original Article

MUTATIONAL ANALYSIS OF HCV GENE ENCODING E1 GLYCOPROTEIN

Muhammad Saad Janjua, Rehman Shahzad, Ghazala Jaffery, Faria Malik , Aneeq Waqar and Hina Bukhari

Objective: To evaluate the variations in HCV glycoprotein E1 gene and to map epitopes in the variable E1 regions informing the development of an effective vaccine against HCV.
Methods: To isolate the E1 gene, RNA extraction was done by using the kit method and then it was converted to the cDNA. Confirmation of HCV presence in the collected samples was done through highly conserved core primers. This was then followed by PCR amplificationfor E1 gene. The sequenced E1 genes were translated in silico into proteinsequences.
Results: A These proteins sequences were then analyzed for the presence of B-cell and T-cell epitopes; two B-cell epitopes (CSLYPGHLSGHRMAWD, TASIRSHVDLLVGAAT) and one T-cell epitope (QAFTFRPRR) were found useful. These could be helpful in the formation of a proper vaccine against HCV.
Conclusions: We found 2 B-cell epitopes and 1 T-cell epitope conserved in 3a genotype that may help in vaccine development.
Keywords: Hepatitis C Virus, PCR, E1 Glycoprotein, B-cell Epitopes.

Introduction

The most indigenous pathological agent of non-A and non-B hepatitis is Hepatitis C virus. It was discovered by Choo and Chiron co-operation group in 1989.[1] HCV causes acute and chronic liver disease, cirrhosis and hepatocellular carcinoma.

Approximately 130 to 170 million population of the world are suffering from hepatitis C virus infection. The epidemiology of transmission of this infection has changed from primarily blood route to currently injection abuse commonly in young adults. Other routes are piercing, tattooing, mother to child and sexual transmission.[2]

Hepatitis C virus transmission progresses slowly and asymptomatically in the acute phase of infection. Of the total population infected only 20 percent is capable of clearing the infection and 80 percent progress to chronic liver disease.[3] Hepatitis C virus belongs to family flaviviridae, genus hepacivirus that has six different phylogenetic groups that further have numerous clades. The most frequently found genus in Pakistan is 3a.[4]

By using cell culture, electron microscopy and other scientific procedures we can study the ultrastructure of Hepatitis C virus. The negative stain electron microscopic study of the virus has shown spherical particles of 45 to 70nm in diameter with a bilayer membrane called envelope consisting of E1 and E2 proteins, viral proteases, nonstructural proteins and apolipoprotein E. They also have internal structure presumably the capsid, single positive stranded RNA of 9500 bases in genome.[5] **(Fig 1).**

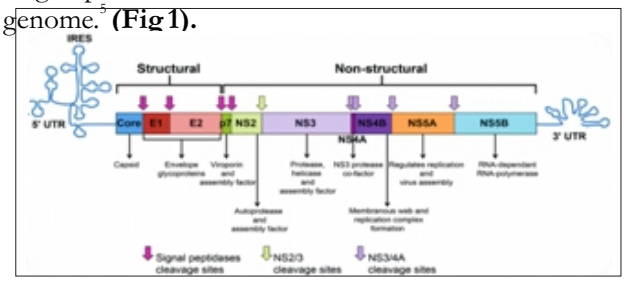

Fig-1: Structure of HCV genome.

!The nature of immune response required to clear the hepatitis C infection is still under consideration, but its protein evokes both humoral and cellular immunity.[6] RNA binding protein is a core protein so it is supposed to form the viral nucleocapsid. This protein receives signals from the host and converts them to the mature form that acts as the membrane protein. This core protein is involved in influencing multiple host cell functions, apoptosis, cell signaling, carcinogenesis, and lipid metabolism. This protein has two domains D1 (at N-terminal consists of two-third hydrophilic domain) and D2 (at C-terminal consists of one-third hydrophobic domain). D2 domain is required for membrane characteristics of core.[7] The E1 and E2 envelope proteins play a key role in the entry of virus into host cells. E1 is a transmembrane glycoprotein, with C-terminal liable for association and membrane permeability. It also

proper folding of HCV glycoprotein and in virus entry[8] E2 also plays a key role in viral breakout from the immune system, as Hyper Variable Regions (HVRs) are found in E2 protein. Mainly two HVRs are responsible in escape of virus. HVR1 plays role in modulating virus entry and HVR2 modulates receptor binding of E2. These envelope proteins E1 and E2 form the heterodimer which appears at the surface of HCV and is surely a contender ligand for cellular receptors like CD81, tetraspanin, heparin sulphate, mannose binding lectins, scavenger receptor class B type I and others. When E2 binds to CD81, natural killer cell surface, it inhibits the cytotoxicity and cytokine production by these cells. It is therefore necessary to kill envelope proteins to inhibit the viral entry. Junction of p7 protein, is present between structural and non-structural proteins, it is small polytypic membrane protein, whose proper function is unknown. Studies have shown that the heterodimerization of glycoproteins E1 and E2 are liable for the virus entry and it is believed that glycans on the heavy glycosylated E1 and E2 are involved in the proper folding of heterodimerization separately; they are non-infectious.[9] Other regions of E2 like HVR1 are responsible for virus interaction with receptors and evasion of immune system due to high rate of variations. However, deletion of HVR1 does not decrease the infectivity of virus. The conserved glycosylation sites specially 4 in E1 and 11 in E2 have been shown to play a role in HCV life cycle, because the deletion of these sites lessen the infectivity of HCV.

After attachment to receptors, HCV interacts with occluding or CLDN-1, which facilitates its cellular uptake. HCV entry is mediated by clathrin-dependent endocytosis. Due to acidification, viral envelope fuses with the early endo some membrane which releases the viral nucleocapsid into the cytoplasm. During virion assembly and replication HCV shows unique character, both these processes depend upon the fatty acid pathway and cholesterol metabolism in host cell. Apolipoproteins are also effective regulators of HCV infectivity. Lipoproteins also play imperative role for the HCV entry and initiation of infection. In other words infection starts when virus particles associated with lipoprotein interact with lipoprotein receptors.[10-12]

According to epidemiological studies related to environment and geographical conditions HCV has 11 major genotypes and over 100 subtypes. Six of these genotypes are the major types which are further classified into many subtypes. 3a genotype is the most common in Pakistan and infects 49.5%HCV patients whereas 3b infects 17%, 1a infects 8.35%, 2a is 7.52% and other genotypes are very rare.[13]

These changes in genomes, within the single host, are responsible for evasion of HCV from the immune system.[14]

Treatment of the chronic HCV has improved during the past few years. In HCV infected patients, it persists in about sixty to eighty percent. Peg-interferon (alfa 2a, 2b) with Ribavirin and interferon (alfa 2a, 2b) have been used widely. Currently several other FDA approved treatments are available. For chronic HCV infection, cirrhosis and liver dysfunction the last option is liver transplantation. However all these treatment options are expensive, therefore there is a need of development of a vaccine or a cheaper treatment. Development of HCV vaccine is challenging as the virus exists in the form of Quasispecies. Reason behind this is too much diversification of HCV due to the lack of proofreading mechanism in RNA polymerase of virus.[14] The current study was undertaken to conduct mutational analysis of HCV gene encoding E1 glycoprotein that could be a candidate for HCV vaccine.

Methods

Research work of the study was conducted in the Molecular Biology and Genomics Laboratory of the Institute of Biochemistry & Biotechnology, University of Veterinary and Animal Sciences, Lahore. Sample Collection: Blood samples from 120 HCV PCR positive patients with 3a genotype were collected from Shalimar Hospital and Nawaz Sharif Hospital, Lahore. Five ml of blood sample was collected aseptically from each patient after taking informed consent. Serum was separated by centrifugation and stored at -20 °C for further processing.

RNA Extraction: RNA was extracted from frozen serum samples after thawing them using AJ Roboscreen Innuprep kit (AJ Roboscreen GmBH, Leipzig, Germany). Extraction method was followed according to the manufacturer's instructions.

RNA Quantification: The extracted RNA was quantified using Nano Drop (Thermo Scientific, Dubuque, USA). The values were in the range of 5 ng/µl to 50 ng/µl.

Primer designing: About 53 sequences of HCV 3a genotype from NCBI (http://www.ncbi.nlm.nih.gov/) were retrieved. Sequences were aligned by

regions.

Primers were chosen by keeping in mind that for good primers GC content should be 40% - 60%, and temperature should be 50˚C to 60˚C. After selecting the primer sequences **Table 1**, these primers were checked through online software Oligo Tm calculator (http://www.basic.northwestern. Edu/biotools/oligocalc.html). Complementarity, hairpin formation, self-annealing sites were checked and synthesized from Macrogen Company.

CDNA Synthesis:

From RNA, cDNA was synthesized by using Thermo Scientific cDNA preparation kit (Thermo Scientific, Dubuque, USA) following manufacturer's instructions.

Quantification of cDNA: After the formation of cDNA, quantification was done using Nano Drop (Thermo Scientific, Dubuque, USA). As cDNA is single-stranded, so the option for ssDNA was selected on Nano Drop. Quantitative values were in the range of 1300 to 1900 ng/µl.

Confirmation of HCV in collected samples:

To confirm the cDNA of HCV virus, we used the primers of core region and 5' UTR region of HCV virus (Primers sequences are given in **Table 2**).

Table-1: Sequences of primers used in the study.

Primers	Sized bp	Sequence	Tm (˚C)
E1P1-F	27	GGGAGGTCTCGTAGACCGTGCACCATG	57
E1P1-R	31	GAG(AC)GG(GT)AT(AG)TACCCCATGAG(AG)TCGGC	57
E1P1-F	27	AGACCGTGCACCATGAGCAC	57
E1P1-R	20	TACGCCGGGGGTCA(TG)T(GA)GGGCCCCA	57

Table-2: Core primers sequences for HCV detection.

Primers	Sized bp	Sequence	Tm (˚C)
P1-F	27	GGGAGGTCTCGTAGACCGTGCACCATG	57
P1-R	31	GAG(AC)GG(GT)AT(AG)TACCCCATGAG(AG)TCGGC	57
P1-F	20	AGACCGTGCACCATGAGCAC	57
P1-R	27	TACGCCGGGGGTCA(TG)T(GA)GGGCCCCA	57

For this nested PCR was done in which two rounds were performed. Reaction mixture of 20 µl was prepared.

Agarose Gel Electrophoresis: After completing the two rounds of the PCR, amplicons were checked by running on the 3% agarose gel. Presence of sharp bands concluded that the samples used were HCV positive.

Amplification of HCV E1 gene: As double round PCR confirmed the presence of HCV in samples, different recipes and PCR conditions were applied to amplify 5 out of 120 HCV positive samples for E1 gene.

Agarose Gel Electrophoresis: After PCR from new primers (Table 1), PCR product was run on 1.2% agarose gelalong with a 1 kb ladder. Amplicon size was around 576 BP as it is the size of the required gene E1.

PCR product purification: PCR product was then purified for further processing. The purified product was run on 1.2% gel to check the presence of required gene. After confirmation of presence of bands, it was sent for sequencing.

Sequencing: The PCR purified product was sent to Centre of Excellence in Molecular Biology (CAMB), Punjab University, Lahore, for sequencing.

Bioinformatics Analysis: After sequencing, bioinformatics analysis was done on the sequenced samples.

Conservancy Analysis: This was done by aligning our sequences with other sequences downloaded from NCBI by using BLAST and Immune Epitope Database (IEDB) analysis tool (http://tools.Immuneepitope.org/main/).

Phylogenetic Analysis: Phylogenetic analysis was done by Mega 6 or CLC sequence analyzer to study the relation of the genomic sequence with the others.Protein Modeling: Protein modeling was performed with I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) online server and quality of the model was assessed with Ramchandran plot and Z-score.

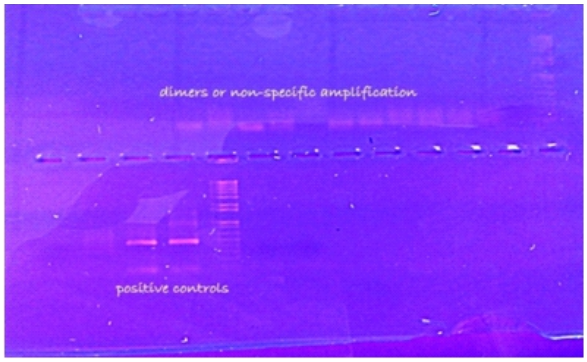

Fig-2: Amplification of cDNA through Core primers

Epitope Mapping Epitope mapping and conservancy analysis was done by Immune Epitope Database (IEDB) analysis tool (http://tools.immuneepitope.org/main/).

Results

Confirmation of HCV samples:

To confirm the presence of hepatitis C virus, RNA was extracted from collected serum samples and amplified through nested PCR by using the Core region and 5' UTR region primers. The product of PCR was then run on 3% agarose gel. Bands were present showing the presence of HCV in the samples as seen in Figure 2.As double round PCR confirmed the presence of HCV in samples, different recipes and PCR conditions were applied to amplify E1 gene as seen in **Fig 3.**

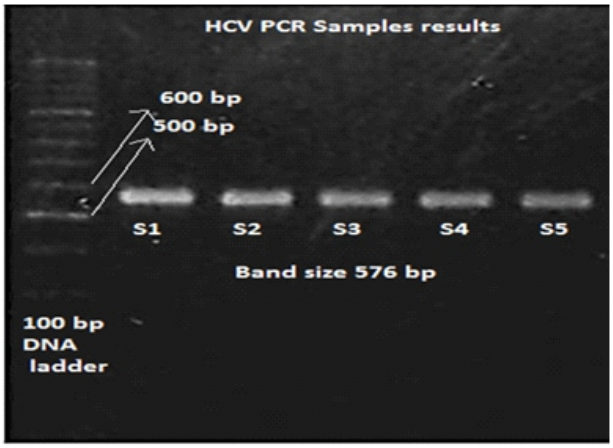

Fig-3: Amplification of HCV E1genes.

After confirming the presence of HCV, the samples were amplified using the E1 primers designed from the Pakistani genomic sequence. PCR were run multiple times, with different recipes and different PCR conditions.

Discussion

HCV is a global health problem and is difficult to treat because of high variations in its genome. HCV is spreading rapidly and it is alarming that there is no vaccine available for its prevention. About 170 million people are affected by HCV worldwide and in Pakistan 3% to 6% population has HCV infection.[15] HCV causes several problems that eventually lead to chronic liver disease, cirrhosis, and hepatocellular

Table-3: Comparison of sequences with standard sequence.

Sr. No	Sequences	Similarity	Gaps	E-value	Score
1	Sequence 1	93%	3%	0.0	837 bits
2	Sequence 2	88%	3%	1e-175	599 bits
3	Sequence 3	91%	2%	0.0	658 bits
4	Sequence 4	90%	2%	1e-180	616 bits
5	Sequence 5	90%	3%	3e-176	610 bits

Table-4: B-Cell epitopes conservancy analysis.

Epitope Name	Epitope Sequence	Epitope lenght	% of protein sequence matches at identity =100%
33	CSLYPGHLSGHRMAWD	16	35.00% (14/40)
31	TASHIRSHVDLLVGAAT	16	32.50% (13/40)

Table-5: T-cell epitopes conservancy analysis

Epitope Name	Epitope Sequence	Epitope lenght	% of protein sequence matches at identity =100%
8	QAFTFRPRR	9	46.15% (18/39)

Carcinoma. Treatment is very complicated due to high mutation rate in HCV virus. There are multiple genotypes in which there are 6 major genotypes. Some genotypes are common in specific regions. 3a genotype is the most common genotype in Pakistan. A serious step is needed to be taken to control its spread. Numerous researches have been conducted on different genes of HCV to control it and to understand how it evades the human immune system. Genes responsible for escaping from immune system are considered to be very important in treatment. E1 is one of these important genes which play a role in the escape of HCV from human immune system. E1 is an enveloped glycoprotein involved in epitope formation. Hyper variable regions are present in E1 gene; these regions help in the escape of HCV from human immune response.[8]

To study variations in E1 gene, serum samples from HCV positive patients were taken and RNA was extracted from them and converted to cDNA which was followed by amplification of the target gene. Amplified PCR products were sequenced and then protein analysis was performed on these sequences.

When nucleotide sequences were obtained, they were converted to protein using ExPaSy translation tool. Primary structure analysis was performed using ProtParam online server and to compare the results, Pakistani HCV E1 gene sequence was used as standard, but the primary structure of all the sequences was almost the same, means no structural changes were detected at the primary level.

To study more about the structures of sample Sequences, secondary structures were made using Phyre 2 online server, changes were noticed in the alpha helices and beta sheets, but they were very minor changes at this level. To check the transmembrane activity of the sequences the TMHMM online server was used and it was confirmed that they are present outside the membrane which means that these genes are involved in the epitope formation or may act as an epitope.

To find out the B-cell and T-cell epitopes, ABCpred and Epijen online servers were used respectively. The maximum epitopes were predicted which were then checked for the antigenicity through Vexijen, and antigenic epitopes were analyzed against 40 E1 gene sequences obtained from NCBI, using IEDB database. So in the results, two B-cell epitopes(CSLYPGHLSGHRMAWD, TASIRSHVDLLVGAAT) and one T-cell epitopes (QAFTFRPRR) were found conserved as they are presentin mostly half of the sequences, which means these epitopes could be of lot of use for the purpose of development of vaccine as these two B-cell epitopes could be used to target the E1 gene of HCV. Development of an effective vaccine against the hepatitis C virus (HCV) has long been defined as a difficult challenge due to remarkable variability of this RNA virus and it has been observed that humans and chimpanzees could be re-infected after re-exposure. On the other hand, progress in the understanding of antiviral immune responses in patients with viral clearance has elucidated key mechanisms playing a role in the control of viral infection. Studies investigating prophylactic vaccine approaches in chimpanzees have confirmed that the induction and maintenance of strong helper and cytotoxic T-cell immune responses against multiple viral epitopes is necessary for protection against viral clearance and chronic infection. A multi-specific B-cell response, resulting in rapid induction of cross-neutralizing antibodies may assist cellular responses. Therapeutic vaccine formulations currently being evaluated in clinical trials are facing the fact that the immune system of chronic carriers is impaired and needs the restoration of T-cell functions to enhance their efficacy.[16] The result of current study would help in identifying the candidate protein of 3a genotypes for HCV vaccine development.

Conclusion

We have found two B-cell epitopes and one T-cell epitope conserved in 3a genotype. CSLYPGHLSGHRMAWD, a B-cell epitope was 35% similar in 40 sequences of 3a genotypes. TASIRSHVDLLVGAAT, a B-cell epitope had 32.50% similarity in 40 sequences of 3a genotype QAFTFRPRR, a T-cell epitope had 46.15% similarity in 39 sequences of 3a genotype.

References

1. Khan UH, Ali A, Akram M, Khattak AM. SEROPREVALENCE OF HEPATITIS C; AMONG HEALTHY BLOOD DONORS IN BLOOD BANKS OF KHYBER PAKHTUNKHWA. Professional Medical Journal. 2014;21(4).
2. Baldo V, Baldovin T, Trivello R, Floreani A. Epidemiology of HCV infection. Current pharmaceutical design. 2008;14(17):1646-54.
3. Hajarizadeh B, Grebely J, Dore GJ. Epidemiology and natural history of HCV infection. Nature Reviews Gastroenterology and Hepatology. 2013;10(9):553-62.
4. Catanese MT, Uryu K, Kopp M, Edwards TJ, Andrus L, Rice WJ, et al. Ultrastructural analysis of hepatitis C virus particles. Proceed- ings of the National Academy of Sciences. 2013;110(23):9505-10.
5. Gastaminza P, Dryden KA, Boyd B, Wood MR, Law M, Yeager M, et al. Ultrastructural and biophysical characterization of hepatitis C virus particles produced in cell culture. Journal of virology. 2010;84(21):10999-1009.
6. Grollo L, Torresi J, Drummer H, Zeng W, Williamson N, Jackson DC. Exploiting information inherent in binding sites of virus-specific antibodies: design of an HCV vaccine candidate cross-reactive with multiple genotypes. Antiviral therapy. 2006;11(8):1005.
7. Dubuisson J. Hepatitis C virus proteins. World Journal of Gastroenterology. 2007;13(17):2406.
8. Goffard A, Callens N, Bartosch B, Wychowski C, Cosset F-L, Montpellier C, et al. Role of N-linked glycans in the functions of hepatitis C virus envelope glycoproteins. Journal of virology. 2005;79(13):8400-9.
9. Owsianka AM, Timms JM, Tarr AW, Brown RJ, Hickling TP, Szwejk A, et al. Identification of conserved residues in the E2 envelope glycoprotein of the hepatitis C virus that are critical for CD81 binding. Journal of virology. 2006;80(17):8695-704.
10. Burlone ME, Budkowska A. Hepatitis C virus cell entry: role of lipoproteins and cellular receptors. Journal of General Virology. 2009;90(5):1055-70.
11. von Hahn T, Rice CM. Hepatitis C virus entry. Journal of Biological Chemistry. 2008;283(7):3689-93.
12. Zeisel MB, Felmlee DJ, Baumert TF. Hepatitis C virus entry. Hepatitis C Virus: From Molecular Virology to Antiviral Therapy: Springer; 2013. p. 87-112.
13. Attaullah S, Khan S, Ali I. Hepatitis C virus genotypes in Pakistan: a systemic review. Virol J. 2011;8(1):433.
14. Keck Z-y, Xia J, Wang Y, Wang W, Krey T, Prentoe J, et al. Human monoclonal antibodies to a novel cluster of conformational epitopes on HCV E2 with resistance to neutralization escape in a genotype 2a isolate. PLoS pathogens. 2012;8(4):e1002653.
15. Umer M, Iqbal M. Hepatitis C virus prevalence and genotype distribution in Pakistan: Comprehensive review of recent data. World J Gastroenterol. 2016;22(4):1684-700.
16. Stoll-Keller F, Barth H, Fafi-Kremer S, Zeisel MB, Baumert TF. Development of hepatitis C virus vaccines: challenges and progress. 2009.