Sequence comparison and phylogenetic analysis of Hepatitis C virus genotype 3 polymerase

YASIR WAHEED, ATTYA BHATTI, SADIA ANJUM and MUHAMMAD ASHRAF

Atta-ur-Rahman School of Applied BioSciences, National University of Sciences and Technology, Islamabad 44000, Pakistan

Received February 6, 2013; Accepted July 4, 2013

DOI: 10.3892/mmr.2014.1926

Abstract. Hepatitis C virus (HCV) is a worldwide health problem with high morbidity and mortality. HCV polymerase is an attractive target for the development of antiviral strategies. The aim of the present study was to report the sequence variation in the HCV NS5B gene from genotype 3 patient samples. The gene was amplified, cloned and sequenced. A nucleotide and amino acid sequence comparison of conserved motifs of HCV NS5B from the current reported sequences and previously reported genotype 3 sequences was performed. The sequence comparison indicated that the motifs A, B, C and F and β loop sequences are conserved in the reported sequences, while sequence variation was observed in motifs D and E. Amino acids E18, Y191, C274, Y276 and H502, which are involved in the interaction between template and primer, are highly conserved in the reported sequences. R48, R158, D225, S367, R386 and R394 amino acids interact with initiating GTP, and are also highly conserved in the reported sequences. A phylogenetic tree revealed that the sequences are clustered with sequences from India. HCV polymerase lacks proofreading ability and has high error rates. The present study revealed that the residues that form the important motifs of HCV NS5B remain conserved. However, it was observed that numerous place changes in the nucleotide sequences did not affect the amino acid sequences of HCV NS5B. The conserved motifs are strong targets for the development of peptide vaccines against HCV.

Introduction

Hepatitis C virus (HCV) is a global health issue and is a causative agent of chronic liver diseases leading to liver cirrhosis followed by complications, including portal hypertension and hepatocellular carcinoma. The infection becomes chronic in 50-80% of individuals. The virus has evolved strategies to regulate the inflammatory response, to control host antiviral defense, to prevent virally infected cells from apoptosis and to use the host cell infrastructure without causing major cytopathogenicity (1,2). The hepatitis C virus is a positive strand RNA virus, classified in the genus Hepacivirus of the family Flaviviridae (3). Approximately 200 million individuals suffer from HCV, an estimated 3.3% of the world's population (4).

The HCV genome contains a single RNA molecule of 9600 nucleotides, carrying a single open reading frame (ORF) flanked by non-translated regions (NTRs). The 5' NTR contains an internal ribosomal entry site and is essential for the translation of the ORF. The 3' NTR together with 5' NTR is essential for viral replication. The virus encodes a single polypeptide that is cleaved by cellular and viral proteases into 10 different proteins. The structural proteins are core, E1 and E2, followed by a p7 protein which is derived from an ion channel. The core protein forms the viral capsid while E1 and E2 are envelope glycoproteins. The non-structural proteins are NS2, NS3, NS4A, NS4B, NS5A and NS5B (5). NS2, along with the amino terminus of NS3, forms the viral protease (6). NS3 acts as a helicase and NTPase. NS4A is cofactor for viral protease (7). NS4B forms the membranous web (8). NS5A is important in HCV replication and regulation of cellular pathways (9), NS5B forms the RNA-dependent RNA polymerase (RdRp) (10).

The NS5B possess the structure of a typical human right hand with fingers, palm and thumb domains. The palm domain has five highly conserved motifs named A-E. Motif A forms the catalytic pocket, motif B plays a role in sugar selection via RdRp, motif C is involved in binding with divalent cations, motif D forms the core structure of the palm and motif E is involved in the interaction between the palm and thumb (11-13). The finger tips connect the thumb with fingers and completely encircle the active site of the enzyme (14). A short β hairpin loop protrudes from the active site of the enzyme and interferes with double-stranded RNA (15). The final 21 amino acids form the membrane anchor domain and are important in HCV replication in cell lines (16). In the present study, an HCV polymerase gene was cloned from five different patient samples. The nucleotides and amino acid sequences of important motifs were compared and a phylogenetic tree of HCV NS5B gene was constructed.

Materials and methods

Patient selection and RNA extraction. Hepatitis C virus positive patients who attended the diagnostic lab in the National University of Sciences and Technology (NUST) Center of
Virology and Immunology (Islamabad, Pakistan) for HCV genotyping, were included in this study. All the patients were >18 years of age and had previously been diagnosed with HCV RNA. Informed verbal consent was obtained from all the patients that participated in the current study and the study was approved by ethical review committee of Atta-ur-Rahman School of Applied Biosciences, NUST, Islamabad, Pakistan. Blood (1500 µl) was collected from patients in an ethylenediaminetetraacetic acid (EDTA) tube and centrifuged at 16,000 x g for 90 sec to separate the serum. Serum was aliquoted into different tubes. One aliquot was subjected to RNA extraction by Qiagen RNA extraction kit (Qiagen, Hilden, Germany) according to the manufacturer’s instructions, while the remaining aliquots were stored at -80°C.

**Primer design and HCV NS5B gene amplification.** NS5B specific primers were designed by a sequence comparison of AM423016, D17763, D28917, EU660386, NC009827, AF238483 and D49374 HCV isolates from NCBI. The primers used were: 5'-AAAGAATTCGTCATGTCGTA CTCTTTGGACCCGGCCG-3' and 5'-AAACTCGAAGTCGGA GCTGGCGAGGAAAGATGC-3'. These primers amplified a fragment of 1,773 base pairs from HCV-positive samples.

Extracted RNA was used as a template for complementary strand (cDNA) synthesis. The reaction mixture for cDNA contained 13 µl of RNA as the template, 1 µl of antisense primer and 2 µl of 10 mM dNTPs and the mix was incubated at 65°C for 5 min. Then, 20 units Molony Murine Leukemia Virus (M.Mulv) reverse transcriptase enzyme (Fermentas, Vilnius, Lithuania), 4 µl M.Mulv buffer, 0.5 µl RNAse inhibitor (Fermentas), 0.6 µl 0.1 M DTT (Fermentas) were added and the mix was incubated at 42°C for 60 min followed by 70°C for 10 min.

The cDNA synthesized was used as a template for PCR. The PCR mix contained 10 µl of cDNA, 5 µl of dNTPs (2 mmol), 1 µl of each primer, 5 µl of DreamTaq buffer (Fermentas), 2.5 units of DreamTaq Enzyme and 27.5 µl of nuclease-free water. The cycle conditions for PCR were as follows: 95°C for 2 min 30 sec followed by 35 cycles of 95°C for 40 sec, 62°C for 30 sec and 68°C for 1 min 50 sec and a final extension at 70°C for 7 min. The reaction was held at 4°C. The PCR products were resolved in a 1% agarose gel.

**Cloning and sequencing of HCV NS5B gene.** NS5B of HCV was amplified in bulk (200 µl), run on a 1% TAE gel and purified using a gel extraction kit (Invitrogen Life Technologies, Carlsbad, CA, USA) according to the manufacturer’s instructions. The purified PCR product was ligated using a TA cloning kit (Invitrogen Life Technologies), according to the manufacturer’s instructions. The ligate was transformed in BL10 competent cells (from the laboratory stock of Atta-ur-Rahman School of Applied Biosciences) by heat shock method. Then, 50 µl of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside and 50 µl of isopropyl β-D-1-thiogalactopyranoside were spread on agar plate containing 1% ampicillin, the transformed cells were spread and incubated at 37°C overnight. Clones were selected via blue/white colony selection and clone confirmation was performed by a plasmid prep followed by restriction digestion.

Positive clones from five different patients were subjected to sequencing by dideoxy chain-termination method (17) using a Beckman Coulter CEQ 8000 (Beckman Coulter, Inc., Brea, CA, USA). The sequencing reaction contained 1 µl of template DNA, 1 µl of primer, 10 µl of water and 8 µl of dye terminator cycle sequencing mix. The thermo cycler conditions for sequencing were as follows: 95°C for 20 sec, 50°C for 20 sec, 60°C for 4 min followed by final hold at 4°C. Stopping solution (5 µl) containing 2 µl of 100 mM disodium EDTA, 2 µl of 3 M sodium acetate and 1 µl of 20 mg/ml of glycogen was added to each tube. The mix was washed with 100% ethanol followed by 70% ethanol and vacuum dried. The pellet was re-suspended in 40 µl of sample loading solution, transferred to the wells of the sample plate and placed in a sequencer.

**Sequence comparison and phylogenetic analysis.** The sequence comparison of conserved motifs of the HCV NS5B gene was performed by using CLC work bench software. (CLCbio, Aarhus, Denmark). The phylogenetic tree of our reported sequences was constructed using 13 different HCV genotypes 3a sequences obtained from NCBI by UPJMA method using CLC work bench software.

**Results**

**General.** The NS5B gene of HCV was amplified using specific sense and antisense primers. A single band of 1,773 bp was obtained and cloned into a TA vector. Positive clones were selected via blue/white color selection. Clones were sequenced from the two directions and a consensus sequence was generated by sequence alignment in CLC work bench software. The consensus sequences were submitted to NCBI for confirmation.

It was observed that HCV NS5B possess five conserved motifs designated as A-E in the palm domain, a motif F in the finger domain and a β loop protruding from the active site of the enzyme. Nucleotides and amino acid sequence comparison of conserved motifs and β loop of the five newly reported HCV NS5B genotype 3a sequences from a Pakistani population together with two other HCV NS5B genotype 3a sequences was performed using the CLC work bench software.

**Motifs.** The motif A contains 212 to 234 amino acids of HCV NS5B. The motif possesses D-X4-D region, while aspartates in the motif are highly conserved. The motif is involved in binding with divalent actions. Nucleotides and amino acid sequence comparisons of motif A are shown in Fig. 1.

Motif B contains 283 to 291 amino acids of HCV NS5B. The motif is important in sugar selection. Nucleotides and amino acids sequence comparisons of motif B are shown in Fig. 2.

Motif C contains 317 to 319 amino acids of HCV NS5B. The motif forms the active site of the enzyme, and aspartate in the motif interacts with the divalent cations. Nucleotides and amino acid sequence comparisons of motif C are shown in Fig. 3.

Motif D contains 326 to 347 amino acids of HCV NS5B. The motif forms the palm core region. Nucleotides and amino acid sequence comparisons of motif D are shown in Fig. 4.

Motif E contains 360 to 376 amino acids of HCV NS5B. The motif is involved in the interaction between palm and thumb domains. Nucleotides and amino acid sequence comparisons of motif E are shown in Fig. 5.

Motif F contains 161 to 169 amino acids of HCV NS5B. The motif forms the interconnecting loops. Nucleotides and amino acid sequence comparisons of motif F are shown in Fig. 6.
The β loop contains 442 to 456 amino acids of HCV NS5B. The β loop interferes with binding to double-stranded RNA caused by steric hindrance. Nucleotides and amino acid sequence comparison of the β loop are shown in Fig. 7.

The phylogenetic tree of the reported HCV NS5B sequences together with 13 other HCV genotype 3a sequences was performed by the UPJMA method. The phylogenetic tree is shown in Fig. 8.

Discussion

In the present study, HCV NS5B from Pakistani patient samples was cloned and sequenced and a sequence comparison
of highly conserved motifs was performed. Nucleotide and amino acid sequence comparison of motif A indicates that the change in nucleotide sequences did not affect the amino acids serine 218 (TCA/TCG), threonine 221 (ACC/ACT) and serine 226 (TCA/TCG). Motif A remains conserved in the reported sequences except at amino acid position 234 where arginine is changed to methionine in clone 1 and to threonine in clone 4.

As reported in our previous study, the D220-X4-D225 motif, which plays a role in binding with the first divalent cation in highly conserved in reference and reported sequences (10). Nucleotide and amino acid sequence comparison of motif B indicated a variation in nucleotides at different positions but the sequence of amino acids remains highly conserved. Alterations in the nucleotide sequences did not affect leucine 285 (CTG/TTA), proline 286 (CCT/CCC) or phenylalanine 289 (TTG/TTC) amino acids. It is reported that G283, T286, T287 and N291 participate in sugar selection by RdRp (18).

Motif A remains conserved in the reported sequences except at amino acid position 234 where arginine is changed to methionine in clone 1 and to threonine in clone 4. As reported in our previous study, the D220-X4-D225 motif, which plays a role in binding with the first divalent cation in highly conserved in reference and reported sequences (10).

Nucleotides and amino acid sequence comparison of motif B indicated a variation in nucleotides at different positions but the sequence of amino acids remains highly conserved. Alterations in the nucleotide sequences did not affect leucine 285 (CTG/TTA), proline 286 (CCT/CCC) or phenylalanine 289 (TTG/TTC) amino acids. It is reported that G283, T286, T287 and N291 participate in sugar selection by RdRp (18). Additionally, the mutation in G283 and T287 completely eliminates RdRp activity (12,18). Sequence comparison indicates that the G283, T287 and N291 are highly conserved in reference and reported sequences, while T286 is mutated to proline in all the sequences.

Motif C forms the active site of the polymerase with the mutation in this motif completely eliminating the polymerase.
function (18). Nucleotides and amino acid sequences of motif C are highly conserved in reference and reported sequences.

Variation in nucleotides and amino acid sequences is observed in reported and reference sequences of motif D. At amino acid position 330, aspartic acid is replaced by glutamic acid in all the reported sequences. At amino acid position 336, leucine is replaced by methionine in clone 1, 2 and 5. Arginine 345 is highly conserved in reference and reported sequences and it is reported that the mutation of arginine 345 to lysine increases RdRp activity to 152% (12,18).

High nucleotide and amino acid sequence variation was observed in the reported and reference sequences of motif E. At amino acid position 362, leucine is replaced by proline, while at 363 isoleucine is replaced by valine, at 374 arginine is replaced by leucine, at 376 aspartic acid is replaced by glycine and at 379 arginine is replaced by lysine amino acids.

Motif F is highly conserved in reference and reported sequences. A single nucleotide change at position 161, did not affect the valine (GTG/GTT) amino acid. The nucleotide sequence comparison of the β loop reveals variation at five different positions but the amino acid sequences are highly conserved in reported and reference sequences.

It has previously been reported that amino acids E18, Y191 and C274, Y276 and H502 are involved in the genome. The diversity is caused by various sequence variations and mutation patterns. It has previously been reported that amino acids E18, Y191, and C274, Y276 and H502 are involved in the genome. The diversity is caused by various sequence variations and mutation patterns.

The hepatitis C virus possesses high genetic diversity in the genome. The diversity is associated with two parameters: i) the estimated half-life of HCV is extremely short with production and clearance of 10⁹–10⁴ virus particles/day in infected patients (20,21) and ii) HCV polymerase lacks the proofreading ability with 10⁻³ to 10⁻⁴ mutations per genomic replication (22,23). Mutations in the HCV genome may also affect the response rate to therapy and protein-protein interactions (24). Despite these factors associated with variation in HCV NS5B, the residues that are involved in important activities by the polymerase remain highly conserved. It was observed that at a number of points in HCV NS5B, the change in nucleotides did not affect the amino acid sequence. Thus, ongoing studies have focused on investigating the immunogenicity of highly conserved motifs in order to design peptide vaccines against HCV.

References


