Research Article

Phylogenetic and 2D/3D Analysis of HCV 1a NS4A Gene/Protein in Pakistani Isolates

Abrar Hussain [1,2]; Muhammad Idrees [2,*]; Muhammad Asif [1]; Liaqat Ali [3,4]; Mahmood Rasool [5]

[1]Department of Biotechnology and Informtics, BUITEMS, Quetta, Pakistan
[2]National Centre of Excellence in Molecular Biology, University of the Punjab, Lahore, Pakistan
[3]Division of Infectious Diseases, Department of Internal Medicine II, University Hospital Freiburg, Freiburg, Germany
[4]Faculty of Biology, Albert Ludwigs University of Freiburg, Freiburg, Germany
[5]Center of Excellence in Genomic Medicine Research, (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia

Corresponding Author: Muhammad Idrees, National Centre of Excellence in Molecular Biology, University of the Punjab, Lahore, Pakistan. Tel: +92-3214769212, E-mail: idrees.khan96@yahoo.com

Background: The nonstructural protein NS4A of hepatitis C virus is composed of 54 amino acids. This small size protein has vital role in many cellular functions. The most important reported function is being a cofactor of viral enzymes serine protease and helicase.

Objectives: The objective of this study was to analyze the phylogenetic variation, its impact in terms of translation and any functional change in protein structure at primary 2D/3D structure using computational tools from Pakistani patients isolates.

Materials and Methods: Patient sera infected with Hepatitis C virus, genotype 1A, were obtained from Molecular Diagnostics lab, CEMB, University of the Punjab Lahore by using BD Vacutainer collection tubes (Becton Dickenson).

Results: Phylogenetic analysis of the gene revealed that Pakistani 1a HCV strains are in the start of third cluster and there is a difference between inter Pakistani isolates at primary, secondary and tertiary levels.

Conclusions: Mutations were present in the central domain of NS4A (amino acids 21 - 34).

Keywords: Hepatitis C Virus; Reverse Transcriptase Polymerase Chain Reaction; RNA

1. Background

Hepatitis C virus infection affected approximately 170 million population worldwide, which 60% to 80% of them are chronically infected. HCV has been classified into 6 different major genotypes and 11 different subtypes (1). The genome of HCV consists of 9.6 kb positive stranded RNA, which is translated into a polyprotein precursor of 3000 amino acids. The polyprotein encodes functional structural proteins, core, envelope (E1 and E2), P7 and nonstructural proteins like NS2, NS3, NS4A, NS4B, NS5A and NS5B (2). NS4A is a nonstructural protein compromising of 54 amino acids, 7-KDa protein composed of three domains: the N-terminal membrane anchor, the central domain that functions as cofactor and C-terminal domain and highly hydrophobic amino acid residues (3, 4). It acts as enhancer viral serine protease, helicase activities and NS3 cleavage. The central domain region of NS4A, amino acid 21 - 34, suffices to mimic the effects of the full protein. NS4A rearranges the secondary structure of both the N-terminus and catalytic site of the NS3 protease, reduces the mobility of the global structure of the NS3 protease, especially the catalytic site, and provides a rigid and tight structure (5). The hydrophobic N-terminal region is indicated to form transmembrane alpha helix and collectively NS3-4A complex has role in membrane anchorage (6). The complex of NS3 protease with NS4A cofactor plays a key role in replication and maturation of HCV (7). The C terminal of NS4A role has been verified in the regulation of NS5A hyperphosphorylation and replication (8). Among its described functions, NS4A transactivator of interleukin 8 promoter, enhances production of IL-8 protein. NS4A (amino acids 21 - 34) is also involved in diminishing both cap-dependent and HCV IRES-mediated translation through interaction with eukaryotic elongation factor 1A (eEF1A). As already mentioned, NS4A is a cofactor of NS3 serine protease and facilitates the NS4A-dependent cleavage at the NS3 - NS4A and NS4B - NS5A junctions. NS4A has also been reported to enhance the phosphorylation of NS5A by cellular kinase(s). The structure of the NS3 protease without NS4A is not suitable for binding and hydrolysis of substrates (5).

2. Objectives

The objective of study was to analyze the phylogenetic variation, its impact in terms of translation and any functional change in protein structure at primary 2D/3D structure using computational tools of Pakistani patients isolates.

3. Materials and Methods

3.1. Sample Collection and RNA Isolation

Patient sera infected with Hepatitis C virus, genotype 1A, were obtained from Molecular Diagnostics lab, CEMB, University of the Punjab Lahore using BD Vacutainer collection tubes (Becton Dickenson). For isolation of serum, serum separation tubes (SST) were used. The serum was recovered after centrifugation at 2000 g for 10 minutes. RNA was isolated from 100 μL of serum, using commercially available Nucleospin RNA isolation kit, according to the kit protocol with little modification (Cat.No.740956.250).

3.2. Complimentary DNA (cDNA) Synthesis

Reverse transcription (RT) was performed by incubating the 10 μL template RNA in a 20-μL reaction mixture containing 1 μL of 10 pmole of specific antisense primer NS4AOAS (5'- 3'), 1 μL of 100 units of MMLV reverse transcriptase (Invitrogen), 0.2 μL of 10 units Ribolock RNase Inhibitor (Fermentas), 4 μL of 5X FSB i.e., First Strand Buffer (Invitrogen), and 2 μL of 10 mM deoxynucleotide triphosphates (dNTPs) at 37°C for 60 minutes, 42°C for 10 minutes, 95°C for 3 minutes and 20°C for 5 minutes.

3.3. PCR Amplification

PCR Amplification of NS4A region of HCV was performed using 4 μL of synthesized cDNA as template with outer forward NS4A-OS (5' - 3') and reverse NS4A-OAS (5' -GTATTACGAGGTTCTCCAAAGC -3') primers. Using 2 μL of first round product as template, re-amplified by performing nested PCR with internal forward NS4AIS (5' - 3') and reverse NS4A-IAS (5' - 3') primers. The PCR product amplified was run on 1.2% agarose gel stained with ethidium bromide and visualized under Ultra Violet transilluminator. The product was cut from the gel and purified using Bead DNA Gel Extraction Kit (Fermentas CAT# K0513) and subjected to sequencing PCR using ABI PRISM 3100 Genetic Analyzer (Applied Biosystem Inc., Foster City, CA, USA) in both directions.

3.4. Sequencing PCR

The PCR amplified product was further verified by sequencing ABI PRISM 3100 Genetic Analyzer (Applied Biosystem Inc., Foster City, CA, USA) in both directions. The sequenced searched for homology with other sequences in Gene Bank and submitted to NCBI Gene bank data base. The assigned accession numbers for the local NS4A genes sequence are JQ679096, JQ679097, JQ679098 and JQ679099.

3.5. Phylogenetic Analysis

3.5.1. Wide-Ranging Phylogenetic Analysis

A reasonable number (219) of sequences were extracted from hepatitis C virus data base present at Gene Bank (http://www.ncbi.nlm.nih.gov/) of HCV genotype 1a to study the phylogenetic analysis of Pakistani isolates. The sequences were extracted using MEGA Web option of MEGA 6.0. They alignment of sequences were performed by ClustalW using progressive alignment methods (9). The Phylogram was generated by Waterhouse et al. (10) and visualized by Archaeopteryx version 0.968 beta BG (10-12). HCV NS4A sequences were used in the analysis of phylogenetic tree.

3.6. Phylogenetic of NS4A by Origin

Based on the country of origin, a brief phylogenetic tree was constructed by sorting NS4A gene present at Gene bank. The sequences included were Australia-FJ931848, Brazil-GU126600, France AY588752, AY588828, AY588864, AY588880, HQ149081, HQ623264, HQ623266, Germany-EU862824, EU862832, EU687194, EU482873, Iran-JF6950 12, Pakistan-JQ679096, JQ679097, JQ679098, JQ679099, Japan-AB520610, Spain-HQ892145, Switzerland-EU256070, EU256071, EU155345, United Kingdom-GU945454, USA, EU155338, EU255963, EU255995, EU622930, EU781777, EU781803, GQ149768, HM568422, HM568426 and Vietnam-FJ768831.

3.7. Prediction of RNA Folding/ Secondary Structures (2D)

For each RNA sequence, 24 structures were predicted through web based server in the form of single strand frequency plots, energy dot plot. Among these, best sequences were selected based on thermodynamic energy $\Delta G = kcal/mol$. Dot plot folding comparisons were performed for 24 structures, each sequences. The portal for the Mfold web server is (http://mfold.rna.albany.edu/?q = mfold/RNA-Folding-Form) (13).

3.8. Tertiary structures (3D)

The RNA sequences were translated into the secondary structure of amino acids. These amino acid sequences were submitted to I-TASSER for generating 3D model of tertiary structures of each sequence. Among five models created for each sequence, one lowest C-score was selected (14, 15). Further analysis of 3D model was performed using PyMol (The PyMol Molecular Graphics System, Version 1.2r3pre, Schrodinger, LLC) (16) and UCSF Chimera (17).

4. Results

4.1. PCR Amplification

The cDNA was synthesized using gene specific outer antisense primer NS4A-OAS (10 pm). Then nested PCR was performed using two sets of primer outer set NS4A-OAS and NS4AOS and inner set NS4A-IAS and NS4A-IS of primers. The PCR products obtained after amplification were run on 1.5% agarose gel and stained with ethidium bromide (0.5 mg/mL) and visualized by UV transillumination.

4.2. Sequencing PCR

Identification of amplified product was confirmed by sequencing PCR in replicate. Sequences of NS4A gene of HCV genotype 1A of Pakistani isolate were searched for homology with other sequences in Gene Bank and submitted to NCBI GenBank data base. The assigned accession numbers for the local NS4A gene sequences are JQ679096, JQ679097, JQ679098 and JQ679099.

4.3. Computational and Phylogenetic Analysis

The computational and phylogenetic analysis of HCV NS4 genotype 1a of Pakistani isolates and sequences from different countries were performed. Various software and web based severs were used in the phylogenetic and computational studies. The sequence from each sample of nonstructural protein 4a was amplified and submitted to GenBank. The accession numbers of the sequences are JQ679096, JQ679097, JQ679098 and JQ679099.

4.4. Evolutionary Pattern Analysis

The pattern of evolutionary analysis was performed by constructing phylogenetic tree. The nucleotide data of 219 sequences was extracted from the National Center for Biotechnology Information database with the help of MEGA Web option of MEGA 6.0. The Multiple sequence alignment was performed by MAFFT: a multiple sequence alignment program described earlier (18). The phylogenetic tree was constructed using The Neighbor-Joining (NJ) method and visualization, analysis, editing was performed using Archaeopteryx software (http://mafft.cbrc.jp/alignment/server/cgi-bin/mafft4.cgi) (10, 19, 20). Two hundred and nineteen different sequences of NS4A protein of HCV genotype 1a including Pakistani isolates investigation revealed that there are basically three clusters as shown in Figure 1.
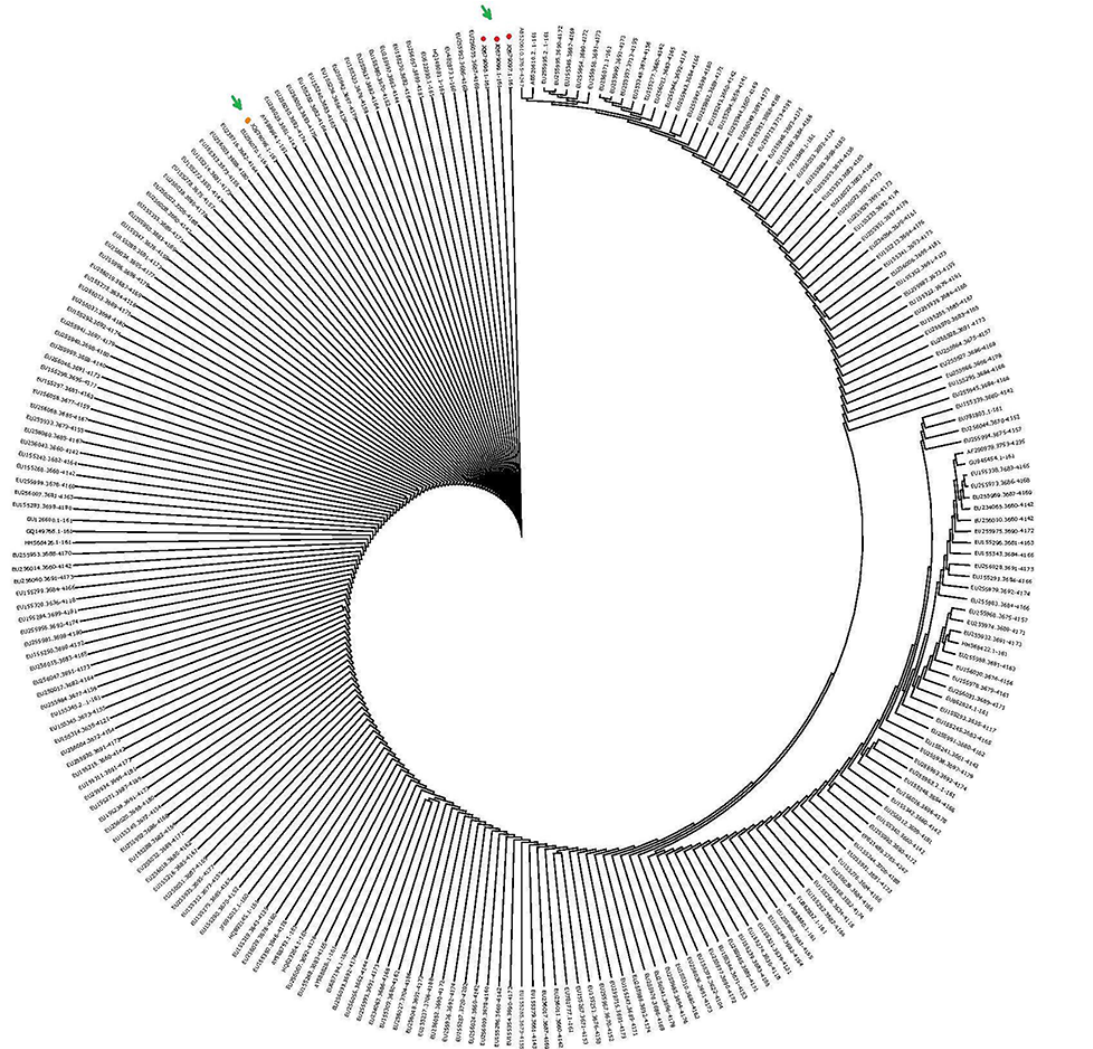

Figure 1. Phylogenetic Analysis 219 Sequence Present at Gene Bank

To further refine the analysis, countries wise phylogenetic tree was constructed by selected sequences of different origins using MEGA 6.0. The phylogenetic analysis revealed that the conserved region was the same among all sequences of different countries. The cluster A represents Germany, Iran and the USA; whereas, the second cluster contains Vietnam, France, Spain, Brazil and the third cluster represents countries including Pakistan and Australia (Figure 2). Based on the phylogenetic analysis the sequences, Australia-FJ931848 (Reference), Pakistan-JQ679096, JQ679097, JQ679098, JQ679099 were selected for Mfold, 2D and 3D for further analysis.
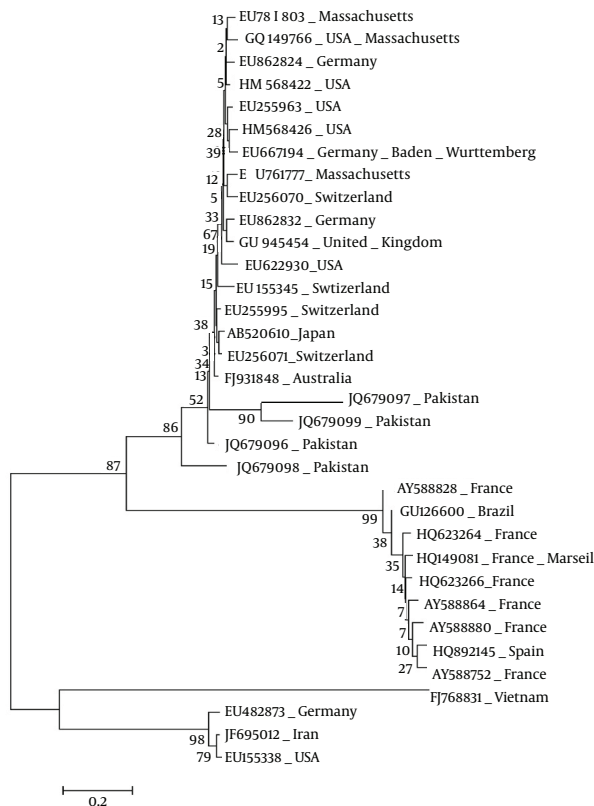

Figure 2. Countries Wise Phylogenetic Tree of NS4A and Pakistani Isolates (Two Separate Clusters)

4.5. Mfold Analysis

The Mfold nucleotide based analysis Australia-FJ931848 (Reference), Pakistan-JQ679096; JQ679097 elaborated the variation in the structure. Multiples structures were generated for a single RNA sequence. The best one is selected on bases of ΔG. A slight change in primary structure of RNA is visible at this level in the structure. Mutational difference between Pakistan-JQ679096/

Australia-FJ931848 at 24 (G/T), 27 (C/T), 78 (A/G), 79 (A/G), 102 (A/G), 106 (A/G), 152 (A/T) and 156 (A/G) substantially alters stem loops of the reference sequence structure (Australia-FJ931848) by converting loop number three and four into a single loop in the sequence (Pakistan-JQ679096) change the RNA structure from Y shape. Whereas, the other isolates are dramatically altered because of the number of mutations (Figure 3). For secondary structure analysis sequences JQ679096 and JQ679097 were selected (13).

4.6. 2D Model/ Secondary Structure Predication of HCV NS4A RNA

The secondary structures of sequence JQ679096 and FJ931848 were similar to each amino acid except positions 27 (S/R), 36 (T/A) and 46 (Q/R). Based on this observation, sequences JQ679096 and JQ679097 were selected. The sequence JQ679096 has three alpha helixes, one beta strand and JQ679097 three alpha helix stands (Figure 4).

4.7. Tertiary Structure of HCV NS4A (3D)

The mutations in RNA are translated into the protein. The changes in primary structure of amino acid are reflected in secondary and subsequently appear into tertiary or 3D structure of the protein (1). The tertiary structure of JQ679096 and JQ679097, produced by I-TASSER server (14, 15), and visualized by PyMol (16). The sequence mutations in primary sequences were reflected in secondary structure, but to lesser extent in the tertiary structure of HCV NS4A. Only three tertiary are presented in Figure 5.

5. Discussion

HCV is a member of Flaviviridae family and genus hepacivirus reported back in 1989. It has somewhat sequence homology with flaviviruses and pestiviruses. HCV genome consists of positive sense ssRNA approximately 9.6 kb of a single open reading frame, translatable into 3000 amino acids (21). ORF (Open Reading Frame) is edged un-translated regions that are 5' - 3'. The polyprotein after translational modification converted into 10 different proteins, three structural and seven nonstructural proteins. The structural proteins are core; E1, E2 and nonstructural are P7, NS2, NS3, NS4A, NS4B, NS5A and NS5B. On average, 1012 virions are produced in chronically infected patients with HCV per day. HCV protease is a multifunctional protein with C- terminal domain with helicase and N-terminal with protease activity (22). NS4A which is 54 residues in length amplifies the activity of protease by contributing one of its beta strands located at amino terminal domain in coordination with metal ion at active site. NS4A intercalates within beta sheet of enzyme core stabilizing the protease from proteolytic degradation (5).
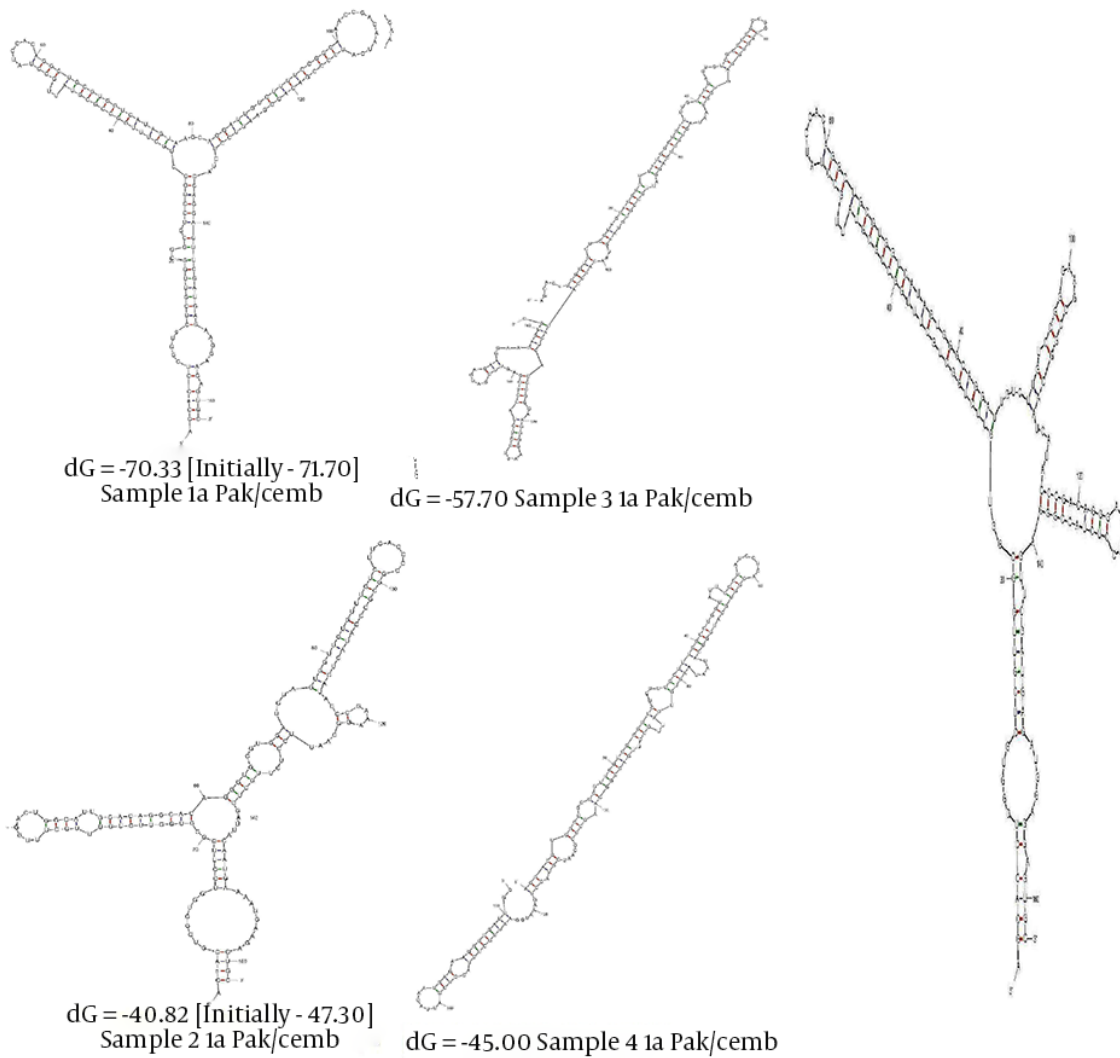
dG = -70.33 [Initially - 71.70]
Sample 1a Pak/cemb

dG = -57.70 Sample 3 1a Pak/cemb

dG = -40.82 [Initially - 47.30]
Sample 2 1a Pak/cemb

dG = -45.00 Sample 4 1a Pak/cemb

Figure 3. Mfold Nucleotide Based Analysis Pakistan-JQ679096, JQ679097 and Australia-FJ931848
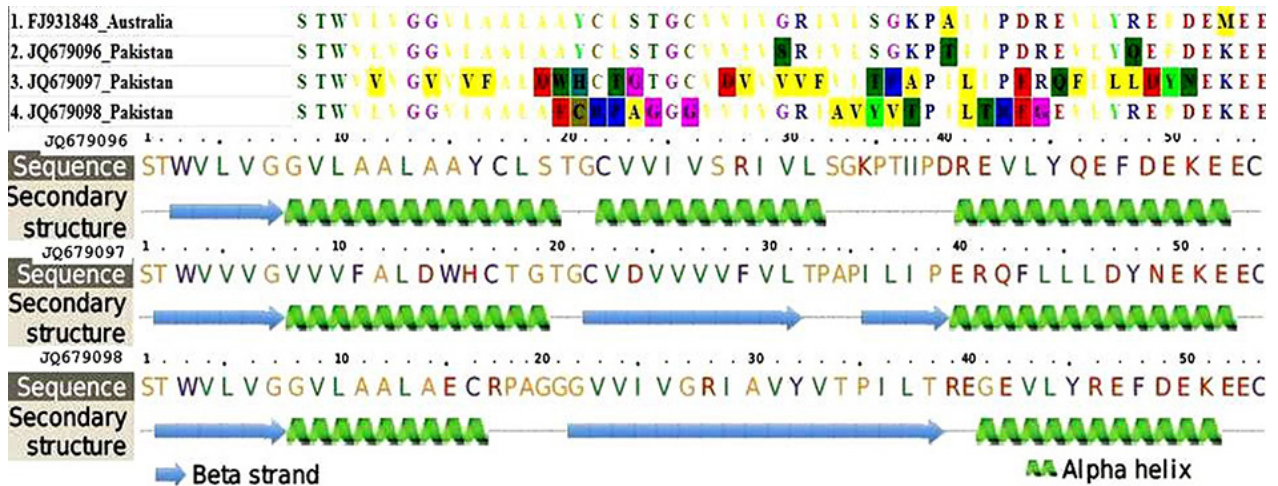

Figure 4. Secondary Structure of NS4A, Comparisons of Mutations in Secondary Structure of Reference Sequences and Pakistani Isolates
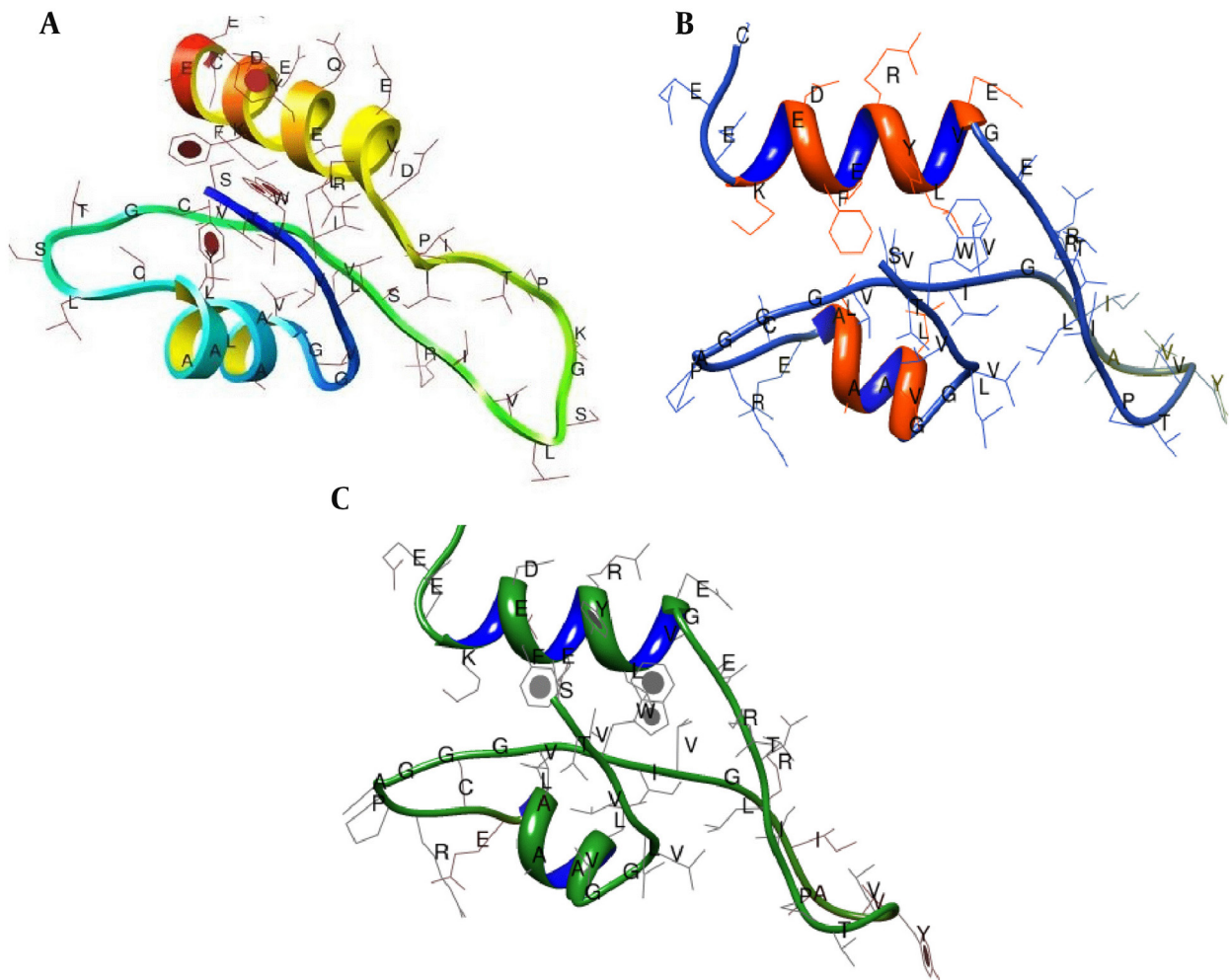
Figure 5. Tertiary Structure of HCV NS4A, Blue Color Presents Central Domain, Red Color Represents Mutations, Labeled and Numbered

The phylogenetic analysis of NS4A presents at NCBI discloses basically three clusters. Furthermore, phylogenetic analysis based on country of origin endorses the three clusters observation. The observation places Pakistani isolate in the third cluster the same as Australian FJ931848 origin sequence hence adopted as reference sequence. Only three mutations (JQ679096/ FJ931848) at positions 27 (S/R), 36 (T/A) and 46 (Q/R) had no significant difference in tertiary structure of the protein function. To study the variation among Pakistani isolates, JQ679096 was selected as the reference molecule. Based on this observation, sequences JQ679096 and JQ679097 were selected. RNA secondary structure has a key role in the biological processes and accurate structure predication can give vital directions for lab work. Many computational methods are present for predicating RNA secondary and tertiary structures. The secondary structure of NS4A of sequence JQ679096 has three alpha helix, one beta strand, JQ679097 three alpha helix stands (Figure 4). The central region of NS4A (21 - 34 amino acid) is sufficient to imitate

the effects of functional protein and has an association with membrane of endoplasmic reticulum by first N-terminal 20 residues in trans-membrane hydrophobic helix manner (23). The NS4A peptide forms a beta strand and is completely buried between two beta strands of N-terminal domain of NS3 (N-terminal /4 - 9 and 32 - 38/C-terminal amino acid) in an anti-parallel confirmation (24). This association makes NS3 more stable. The substitution Val23, Gly27, Arg28 and particularly, Ile25 in the central domain of NS4A are important for its function (24).

JQ679096 mutation at positions 27, 36 and 46 related to Ser, Thr and Gln did not produce any impact on the function of protein. Whereas, JQ679096 mutations at positions 24, 25, 27, 28, 29, 33, 34 and 37 altered the three dimensional structure of NS4A but still remained functional. The functional impact is of theses mutations is not translated in secondary structure of the protein. These mutations were present in central domain, which interact with the NS3 and augment its protease activity. To verify the hypothesis regarding the result of experi-

mental and computational alignment high number of samples, data analysis would be more effective and recommended.

Authors' Contributions

The study was designed by Abrar Hussain. Analysis and data interpretation was performed by Abrar Hussain, Muhammad Asif and Liaqat Ali. Critical revision of the manuscript was contributed by Mahmood Rasool and Abrar Hussain. Muhammad Idrees supervised the entire work, wet lab, dry lab and manuscript writing, correction and submission.

References

1. Sarwar MT, Kausar H, Ijaz B, Ahmad W, Ansar M, Sumrin A, et al. NS4A protein as a marker of HCV history suggests that different HCV genotypes originally evolved from genotype 1b. *Virol J.* 2011;**8**:317.
2. Curcio F, Villano G, Masucci S, Plenzik M, Veneruso C, De Rosa G. Epidemiological survey of hepatitis C virus infection in a cohort of patients from a ser.T in naples, Italy. *J Addict Med.* 2011;**5**(1):43–9.
3. Lindenbach BD, Pragai BM, Montserret R, Beran RK, Pyle AM, Penin F, et al. The C terminus of hepatitis C virus NS4A encodes an electrostatic switch that regulates NS5A hyperphosphorylation and viral replication. *J Virol.* 2007;**81**(17):8905–18.
4. Back SH, Kim JE, Rho J, Hahm B, Lee TG, Kim EE, et al. Expression and purification of an active, full-length hepatitis C viral NS4A. *Protein Expr Purif.* 2000;**20**(2):196–206.
5. Zhu H, Briggs JM. Mechanistic role of NS4A and substrate in the activation of HCV NS3 protease. *Proteins.* 2011;**79**(8):2428–43.
6. Hara H, Aizaki H, Matsuda M, Shinkai-Ouchi F, Inoue Y, Murakami K, et al. Involvement of creatine kinase B in hepatitis C virus genome replication through interaction with the viral NS4A protein. *J Virol.* 2009;**83**(10):5137–47.
7. Shiryaev SA, Chernov AV, Shiryaeva TN, Aleshin AE, Strongin AY. The acidic sequence of the NS4A cofactor regulates ATP hydrolysis by the HCV NS3 helicase. *Arch Virol.* 2011;**156**(2):313–8.
8. Yan Y, Li Y, Munshi S, Sardana V, Cole JL, Sardana M, et al. Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: a 2.2 A resolution structure in a hexagonal crystal form. *Protein Sci.* 1998;**7**(4):837–47.
9. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994;**22**(22):4673–80.
10. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 2009;**25**(9):1189–91.
11. Katoh K, Toh H. Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics.* 2010;**26**(15):1899–900.
12. Han MV, Zmasek CM. phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics.* 2009;**10**:356.
13. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 2003;**31**(13):3406–15.
14. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc.* 2010;**5**(4):725–38.
15. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics.* 2008;**9**:40.
16. Lua RC, Lichtarge O. PyETV: a PyMOL evolutionary trace viewer to analyze functional site predictions in protein complexes. *Bioinformatics.* 2010;**26**(23):2981–2.
17. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem.* 2004;**25**(13):1605–12.
18. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002;**30**(14):3059–66.
19. Katoh K, Toh H. Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. *BMC Bioinformatics.* 2008;**9**:212.
20. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 2011;**28**(10):2731–9.
21. Hussain A, Idrees M. The first complete genome sequence of HCV-1a from Pakistan and a phylogenetic analysis with complete genomes from the rest of the world. *Virol J.* 2013;**10**:211.
22. Kim JL, Morgenstern KA, Lin C, Fox T, Dwyer MD, Landro JA, et al. Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell.* 1996;**87**(2):343–55.
23. Phan T, Kohlway A, Dimberu P, Pyle AM, Lindenbach BD. The acidic domain of hepatitis C virus NS4A contributes to RNA replication and virus particle assembly. *J Virol.* 2011;**85**(3):1193–204.
24. Shimizu Y, Yamaji K, Masuho Y, Yokota T, Inoue H, Sudo K, et al. Identification of the sequence on NS4A required for enhanced cleavage of the NS5A/5B site by hepatitis C virus NS3 protease. *J Virol.* 1996;**70**(1):127–32.