Research Article

Comparison of Structural Architecture of HCV NS3 Genotype 1 versus Pakistani Genotype 3a

Kaneez Fatima,[1] Esam Azhar,[2] Shilu Mathew,[2,3] Ghazi Damanhouri,[2] and Ishtiaq Qadri[2]

[1] *IQ Institute of Infection and Immunity, P.O. Box 54000, Lahore, Punjab, Pakistan*
[2] *King Fahd Medical Research Center, King Abdul Aziz University, P.O. Box 80216, Jeddah 21589, Saudi Arabia*
[3] *Center of Excellence in Genomic Medicine, King Abdul Aziz University, P.O. Box 80216, Jeddah 21589, Saudi Arabia*

Correspondence should be addressed to Ishtiaq Qadri; ishtiaq80262@yahoo.com

This study described the structural characterization of Pakistani HCV NS3 GT3a in parallel with genotypes 1a and 1b NS3. We investigated the role of amino acids and their interaction patterns in different HCV genotypes by crystallographic modeling. Different softwares were used to study the interaction pattern, for example, CLCBIO sequence viewer, MODELLER, NMRCLUST, ERRAT score, and MODELLER. Sixty models were produced and clustered into groups and the best model of PK-NCVI/Pk3a NS3 was selected and studied further to check the variability with other HCV NS3 genotypes. This study will help in future to understand the structural architecture of HCV genome variability and to further define the conserved targets for antiviral agents.

1. Background

The hepatitis C virus (HCV) is a positive-sense, approximately 9.6 kilo base (kb) single-stranded RNA (ssRNA) which is translated into a polypeptide of over 3,000 amino acids long from a single open reading frame. HCV belongs to genus *Hepacivirus* of *Flaviviridae* family [1] and sole member of its family. Both the host and viral proteases process the HCV polyprotein into four structural proteins (core, E1, E2, and p7) and six nonstructural (NS) proteins (NS2, NS3, NS4A, NS4B, NS5A, and NS5B) [2].

HCV existed as quasispecies due to its better evolutionarily stable strategy. To date six genotypes [1–7] and more than 100 subtypes have been described [3]. These subtypes can exist as separate or in combination in an individual [4]. In Pakistan more than 10 million people are living with HCV, with high morbidity and mortality and most prevalent genotype of HCV was 3a [5]. HCV disease severity and its treatment response are greatly influenced by its nature [6]. Peg-interferon and ribavirin have been used with highest positive response rates for the treatment of genotypes 2 and 3 chronic HCV infection in developed countries [7–9]. It is ever challenging to characterize HCV heterogeneity

and its treatment in developing countries due to different aspects including immunologic stress, HCV relapse rate, insignificant epidemiological studies, therapeutic cost, and patient's access to treatment of disease complications over time [10]. This comparative study is therefore designed to define the influence of genetic variation on the structural architecture of HCV NS3 GT3a, that is, which was previously studied extensively by cloning of NS3-GT3a protease from local population [11].

NS3 protein has two separate independent functional domains. A serine protease domain localized to the 181 residues of N-terminal, which acts to cleave HCV polyprotein and RNA/DNA helicase composed of 456 amino acids in C-terminal. Helicase portion unwinds duplex RNAs formed during the viral RNA replication in a $3'$-$5'$ direction [12]. Inquisitively, the HCV helicase unwinds DNA more efficiently than RNA [13] even though the absence of any DNA intermediates in the HCV virulence cycle.

Several crystal structures of HCV NS3 including PDB IDs 3O8B [14], 1CU1 [15], 2ZJO [16], 1A1V [17], 1HEI [18], 3KQN [19], 2F55 [20], and 8OHM [21] and mechanistic studies [12, 21–27] helped to predict the ATP derived energy coupled helicase activity but all of the studies exclusively focused on

genotype 1a or 1b of HCV. However, no molecular architecture of full length HCV GT3a NS3 has been elucidated. Little is known about the impact of residue variations on the HCV GT3a NS3 activities.

Objectives. The NS3 was chosen for this study because it has a clear biological relevance and can be relatively modeled and characterized easily by using the predetermined crystal structures of HCV genotype 1 NS3. This study described the characterization of HCV NS3 GT3a in parallel to genotypes 1a and 1b NS3. The goal is to understand the structural architecture of HCV genome variability and to further define the conserved targets for antiviral agents.

2. Methods

The sequence of NS3 protein belonged to Pakistani strain of HCV genotype 3a previously reported (PK-NCVI/Pk3a NS3) with accession number (FJ839678) and NZL1 (NC_009824) was obtained from online database of Genbank maintained by NCBI. Only the sequence of PK-NCVI/Pk3a NS3 (FJ839678) was submitted to position-specific iterative BLAST [28] at NCBI to obtain homologous 3-dimensional (3D) structures of proteins deposited in the protein data bank (PDB). The further analysis was done based on highest sequence identity and smallest distance from PK-NCVI/Pk3a as potential templates with comparative modeling, including PDB Ids 3KQN [19] (X-ray structure of hepatitis C virus NS3 helicase in complex with ssRNA), 3O8B [14] (Crystal structure of HCV NS3 protein), and 1CU1 [15]. Sequences of crystal structures having PDB ids 3O8B [14], 1CU1 [15], 2ZJO [16], 1A1V [17], 1HEI [18], 2F55 [20], and 8OHM [21] were the other entries for NS3 protein in PDB and studied for protein-protein comparison. Multiple sequence analysis was carried out following an established protocol using ClustalW [29] and T-Coffee [30] along with rapid scanning and correction of multiple sequence alignments [31, 32]. The resulting multiple alignments were visualized and annotated using the sequence alignment protocol in CLC sequence viewer (CLCBIO, Aarhus, Denmark). The homology model for consensus sequence was produced using MODELLER [33, 34]. The different structural models were produced and clustered into groups by using NMRCLUST [35] based on the root mean square distance (RMSD) between the corresponding residues in their structures. Best model was selected using the results of NMRCLUST, ERRAT score [36], and MODELLER calculated energy score of the model, as criteria. All of the protein assessment reported optimized comparative model of PK-NCVI/Pk3a NS3. Molecular operating environment (MOE) software: chemical computing group's molecular operating environment (MOE) and discovery studio were used for further protein analysis [37].

3. Results

3.1. Phylogenetic Analysis and Structural Comparison of HCV NS3 Protein. We have done the pairwise alignment and phylogenetic analysis in order to identify percentage identity of reported NS3 GT3a (NZL1) with our target sequence PK-NCVI/Pk3a NS3 by comparative modeling. The target sequence and crystallographic analysis of PDB Id. 3O8B and 1CU1 shared the phylogenetic branch of target sequence as compared to other crystallographic studies (Figure 1(a)). Phylogenetic analysis was further confirmed by the multiple alignment percentage identity of 76.06% among their residues by using ClustalW and T-Coffee along with rapid scanning and correction of multiple sequence alignments (Figure 1(b)). The resulting multiple alignments were visualized and annotated using the sequence alignment protocol in CLC sequence viewer. It was found that most of the sequence variations were divided among all of the domains but N-terminal and middle portion of the reported NS3 protein were comparatively rich in variations (Figure 2). NS3-NS4A is composed of six subdomains (Figure 3(a)). The helicase, usually referred to as α-helical subdomain, consists of two structurally related β-α-β subdomains and a third subdomain of three short β strands and seven helices. The protease domain is folded into dual barrel fold, like other members of chymotrypsin serine protease family. The 13-residue NS4A proved to be a protease activation domain and is considered to be the sixth subdomain (Figure 3(a)). By using NMRCLUST, sixty models were produced and clustered into groups. The best structural model was selected using the results of NMRCLUST, ERRAT score, and MODELLER. Inspection of the final structural model indicates that some major mutations are present in the helicase and non-functional regions of PK-NCVI/Pk3a which presents some new non-covalent intermolecular (IM) interactions alongside of conserved sequence (Figure 3(b)). Major domains including motifs I (Walker motif: G207SGKSTK), Ia (Y223KVLVLNPSVA), II (Walker B: D290ECH), III (V319LATATPP), IV (L365IFCHSKKK), V (A410TDALMTGYTGDF), and VI (V456SRSQRRGRTGR) were conserved in the all HCV NS3 proteins of different genotypes. Structural architecture of these motives presented that those motifs make a border to ATP binding cleft and project some residues into the nucleic acid binding site. The residues of these motives transform the chemical energy derived from ATP hydrolysis into a mechanical force necessary for helicase movements for the disruption of ssRNA or RNA base pairs. Analogous IM interaction pattern of those domains was seen in all of the crystal studies and comparative model of reported NS3 except for mutations of Lys224Asn in motif Ia and Tyr418Phe in motif V (Supplementary Figure 1 available online at http://dx.doi.org/10.1155/2014/749254). Results implicated that mutation in motif 1a was strengthening its confirmation while motif V was loosely bound.

The role of most of the key conserved residues including Ala204, Thr324, Lys210 Ser231, Ser370, Thr411, Thr266, Tyr267, Met288, Thr269, Asp290, Glu291, Cys292, His293, Thr322, His369, Glu493, Tyr392, Trp501, Arg39, Val432, Phe438, Phe444, Thr450, Gln460, Arg461, Arg462, Gly463, Arg464, Thr465, Gly466, and Arg467 of HCV genotype 1a NS3, using site directed mutagenesis, has already been investigated. The comparison of these residues along with

Figure 1(a): Neighbour-joining tree

Clade 1 (100):
- 86: 100 → 1HEI, 1A1V; 2ZJO
- 100: 76 → 2F55; 3KQN; 8OHM

Clade 2 (100):
- 100 → Pk3a HCV NS3, NZL1
- 100 → 3O8B, 1CU1

(a)

	1CU1	3KQN	3O8B	2F55	8OHM	2ZJO	1A1V	1HEI	NZL1	Pk3a HCV NS3
	1	2	3	4	5	6	7	8	9	10
1CU1		67.51	100.00	67.72	66.40	71.95	68.98	67.19	75.77	76.06
3KQN			67.51	96.34	95.42	88.63	88.06	90.69	54.16	54.31
3O8B				67.72	66.40	71.95	68.98	67.19	75.77	76.06
2F55					94.94	87.37	86.57	89.36	53.58	53.58
8OHM						86.95	86.57	89.14	52.85	53.43
2ZJO							91.02	88.63	56.20	56.35
1A1V								95.31	56.35	56.35
1HEI									54.89	54.89
NZL1										97.08

(b)

FIGURE 1: Sequence comparisons of the HCV NS3 genotypes. (a) Neighbour-joining tree based on phylogenetic analysis showing the relationships of HCV NS3 GT3a with GT1a/1b (b). Diagonal represents the comparison of percentage identities among crystal structure of HCV NS3 with Pk3a HCV NS3 and HCV NS3 of NZL1 strain. Dark colored values showed a higher percentage identity among cross matched sequence.

their interactions in modeled and template protein has been depicted in Table 1.

Some residues were seen to make different interaction pattern with their neighbors especially in the presence of ssRNA, ATP and solvent molecules. In the crystallographic studies of helicase interaction with ssRNA PDB Id, 3KQN Gln460 interacts with Thr324 and Arg464 but in the model of PK-NCVI/Pk3a and other studies without ATP and ssRNA, this interaction was seen with His293 and Arg464. Some amino acids like Glu291, Cys292, His293, and Thr324 were

evaluated to be on surface and create a hydrogen bond with water used as solvent during crystallographic studies which was absent in the model of PK-NCVI/Pk3a. Some of the residues, which are not part of the functional regions of protein and nonconserved in PK-NCVI/Pk3a NS3, were actually involved in the structure stability. These residues are Arg9Gln, Leu332Pro, Leu354Ile, Ile605Val, and Ser622Cys [11]. The comparison of the IM interactions among nonconserved residues of PK-NCVI/Pk3a NS3 and of HCV NS3 of other genotypes is listed in Table 2.

```
                                    20              40              60              80
       1CU1 GSITAYSQQT RGLLGCIITS LTGRDKNQVE GEVQVVSTAT QSFLATCVNG VCWTVYHGAG SKTLAGPKGP ITQMYTNVDQ DLVGWQAPPG 90
       3KQN .......... .......... .......... .......... .......... .......... .......... .......... ..........
       3O8B .......... .......... .......... .......... .......... .......... .......... .......... ..........
       2F55 .......... .......... .......... .......... .......... .......... .......... .......... ..........
       8OHM .......... .......... .......... .......... .......... .......... .......... .......... ..........
       2ZJO .......... .......... .A. M..... .......... .......... .......... .......... .......... ..........
       1A1V .......... .......... .......... .......... .......... .......... .......... .......... ..........
       1HEI .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZL1(NC_009824) AP...A... ....T.V.. ....V.T ....L.... .T.G.T.G. .I. .R....A.H. AL....... .....P....
NS3-GT3aP (FJ839678) AP...A.R. ....T.V.. ....V.T ....L.... .T.G.T.G. .M. .R....V.H. AL....... .....P....
                                    100             120             140             160             180
       1CU1 ARSLTPCTCG SSDLYLVTRH ADVIPVRRRG DSRGSLLSPR PVSYLKGSSG GPLLCPSGHA VGIFRAAVCT RGVAKAVDFV PVESMETTMR 180
       3KQN .......... .......... .......... .......... .......... .......... .......... .......... ..........
       3O8B .......... .......... .......... .......... .......... .......... .......... .......... ..........
       2F55 .......... .......... .......... .......... .......... .......... .......... .......... ..........
       8OHM .......... .......... .......... .......... .......... .......... .......... .......... ..........
       2ZJO .......... .......... .......... .......... .......... .......... .QQ M..... .......... ..........
       1A1V .......... .......... .......... .......... .......... .......... .......... ..I. ...NL.....
       1HEI .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZL1(NC_009824) .K..E..A. ....D .....A.... .TA... .LAC... ..VM.....V A .........SLQ.I ...TLS.QA.
NS3-GT3aP (FJ839678) .K..E.... .A....D .....A.... .TA... .LAC... ..VM.....V A .........LR.I ...TLS.QA.
                                    200    (Motif I)  220     (Motif Ia)    240             260
       1CU1 SPVFTDNSSP PAVPQSFQVA HLHAPT|GSGK STK|VPAAYAA QG|YKVLVLNP SVA|ATLGFGA YMSKAHGIDP NIRTGVRTIT TGAPVTYSTY 270
       3KQN .......... .......... ...T..... .......... .......... .......... .......... ...I..... ..........
       3O8B .......... .......... .......... .......... .......... .......... .......... .......... ..........
       2F55 .......... .......... ...T..... .......... .......... .......... .......A. .GSI..... ..........
       8OHM .......... .......... .......... .......... .V. .......... .......A. .G.I..... ..........
       2ZJO .......... .......... ...A..... .......... .......... .V. . .......... ...I..... ..........
       1A1V .......... .......... .......... .......... .......... .V. . .......... ..S.I..... ..........
       1HEI .......... .......... .......... .......... .......... .V. . .......... ..S.I..... ..........
NZL1(NC_009824) .S.S...T. ......Y..G Y .......... .V. . ..N... .....S F..R.Y.... ..N..V. .KL.....
NS3-GT3aP (FJ839678) .S.S...T. ......Y..G Y .......... .V. . ..N... .....S F..R.Y.... ..N..V. .KL.....
                                    280    (Motif II)   300             320 (Motif III)   340             360
       1CU1 GKFLADGGCS GGAYDIIICD ECH|STDSTTI LGIGTVLDQA ETAGARLV|VL ATATPP|GSVT VPHPNIEEVA LSNTGEIPFY GKAIPIEAIR 360
       3KQN .......... .......... .......... .......... .......... .......... .......... ...S..... ......T.K
       3O8B .......... .......... .......... .......... .......... .......... .......... .......... ..........
       2F55 .......... .......... .......... .......... .......... .......... .IG ...N.... .......... .......K
       8OHM .......... .......... ....S..... .......... .......... .......... .......... .......... .......V.
       2ZJO .......... ....M..... ....A.S. .......... .......... .......... .I. .T. .......... ......T.K
       1A1V .......... .......... ....A.S. .......... .......... .......... .T. .......T. .......L.V.K
       1HEI .......... .......... ....A.S. .......... .......... .......... .S. .......T .......L.V.K
NZL1(NC_009824) .......V.. .AQ.A.S. .......... .......... ..V..T... .......I. ..S. .......GSE .......ALLK
NS3-GT3aP (FJ839678) .......V.. .AQ.A.S. .......... .......... ..V..T... .......I. ..L.S. .......GSE .......L..AL.K
       (Motif IV)   380             400             (Motif V)   420             440
       1CU1 GGRH|LIFCHS KKK|CDELAAK LSGLGINAVA YYRGLDVSVI PTIGDVVVVA| TDALMTGYTG DF|DSVIDCNT CVTQTVDFSL DPTFTIETTT 450
       3KQN .......... .......... ..L.... .......... .S...I. .......F. .......... .......... ..........
       3O8B .......... .......... .......... .......... .......... .......... .......... .......... ..........
       2F55 .......... ..T..... .T. ...L.... .......P.. .......F. .......... .......... ..........
       8OHM .......... ..A..L. .......... .......S. .......F. .......... .......... .......D...
       2ZJO .......... ..A....H. .......... .AS.N.. .......F. .......... .......... ..........
       1A1V .......... ..VA..... .......... ....S..... ...S ......F. .......... .......... ..........
       1HEI .......... ..VA..... .......... ....N..... ...S ......F. .......... .......... ..........
NZL1(NC_009824) .......... ..I.S. .R.M.L. .......... .T....C. .......F. .......... ..V A.E.Y. ...S..R.
NS3-GT3aP (FJ839678) .......... ..I.S. .R.M.L. .......... .S....C. .......... .......... ..V A.E.Y. ...S..R.
                                    460   (Motif VI)   480             500             520             540
       1CU1 VPQDA|VSRSQ RRGRTGR|GRR GIYRFVTPGE RPSGMFDSSV LCECYDAGCA WYELTPAETS VRLRAYLNTP GLPVCQDHLE FWESVFTGLT 540
       3KQN .......... ....M..... .......... .......... .......... .......... .......... .......... ..........
       3O8B .......... ....S..... .......... .......... .......... .......... .......... .......... ..........
       2F55 .......... ....S..... .......... .......... .......... .......... .......... .......... ..........
       8OHM .......... .......... .......... .......... .......... .......... .......... .......... ..........
       2ZJO M..... ...S..... .......... .......... .......... .......... .......... .......... ..........
       1A1V L.....T. ...KP .......A. .......... .......... ....M... .T. ..M. .......... .....G....
       1HEI L.....T. ...KP .......A. .......... .......... ....M... .T. ..M. .......... .....G....
NZL1(NC_009824) A.......... ...L .T..Y.AS. ...V. .......... ..S .D.Q...T .......S. .......D.. ..........
(FJ839678) A.......... ...L ...Y.S. ...V. .......... ..S .D.Q...T .......S. .......D.. ..........
                                    560             580             600             620
       1CU1 HIDAHFLSQT KQAGDNFPYL VAYQATVCAR AQAPPPSWDQ MWKCLIRLKP TLHGPTPLLY RLGAVQNEVT LTHPITKYIM ACMSADLEVV 630
       3KQN .......... .......... .......... .......... .......... .......... .......... .T. ..........
       3O8B .......... .......... .......... .......... .......... .......... .......... .......... ..........
       2F55 .......... .......... .......... .......... .......... .......... ..I. .......... ..........
       8OHM .......... .......... .......... .......... ..S.T. .......... ..L. ..V..F. ..........
       2ZJO .......... .......... .......... .......... ..T. .......... .......... .......... ..........
       1A1V .......... .S.E. .......... .......... .......... .......... .......... .T. ..........
       1HEI .......... .S.E. .......... .......... .......... .......... .......... .T. ..........
NZL1(NC_009824) .......... .Q.L.S.. T. .......... ..E... ..V... ...P....TC. ..L. .......T
NS3-GT3aP (FJ839678) .......... .Q.L.S.. T. .......... ..E... ..V... ...PI...C .......... .S......T
                                    640             660             680
       1CU1 T.......... .......... .......... ..... 685
       3KQN .......... .......... .......... .....
       3O8B .......... .......... .......... .....
       2F55 .......... .......... .......... .....
       8OHM .......... .......... .......... .....
       2ZJO .......... .......... .......... ...S
       1A1V .GSG...... .......... .......... ...S
       1HEI .......... .......... .......... .....
NZL1(NC_009824) .STWVLLGGV LAALAAYCLS VGCVVIVGHI ELEGKPALVP DKEVLYQQYD EMEEC
NS3-GT3aP (FJ839678) .STWVLLGGV LAALAAYCLS VGCVVIVGHI ELGGKPALVP DKEVLYQQYD EMEEC
```
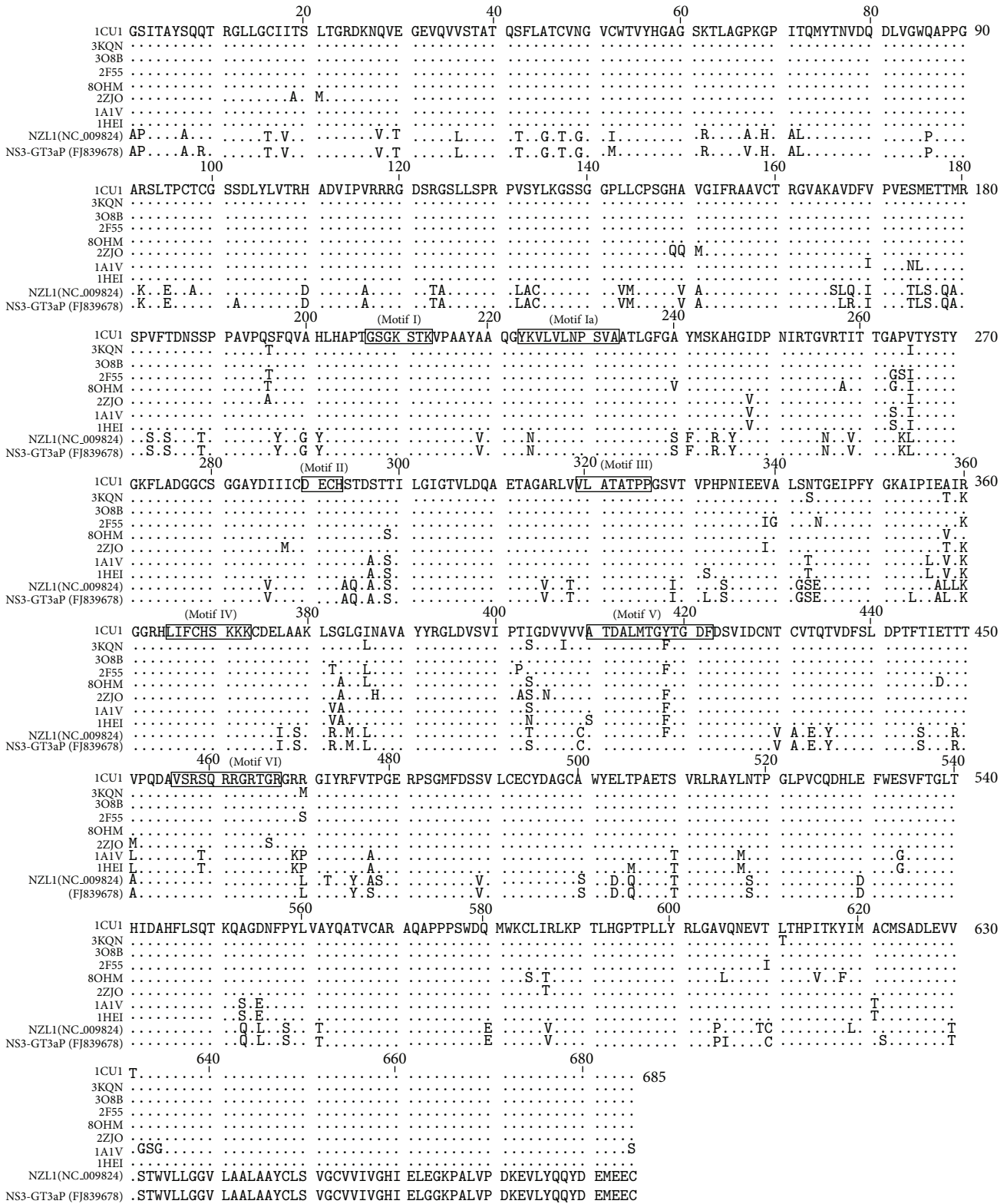
Figure 2: Multiple sequence alignment of different reported NS3 structures and target sequences of PK-NCVI/Pk3a NS3. Motifs I–VI are marked by boxes on the first sequence and labeled, respectively. The conserved sequences are shown as dots.

(a)   (b)
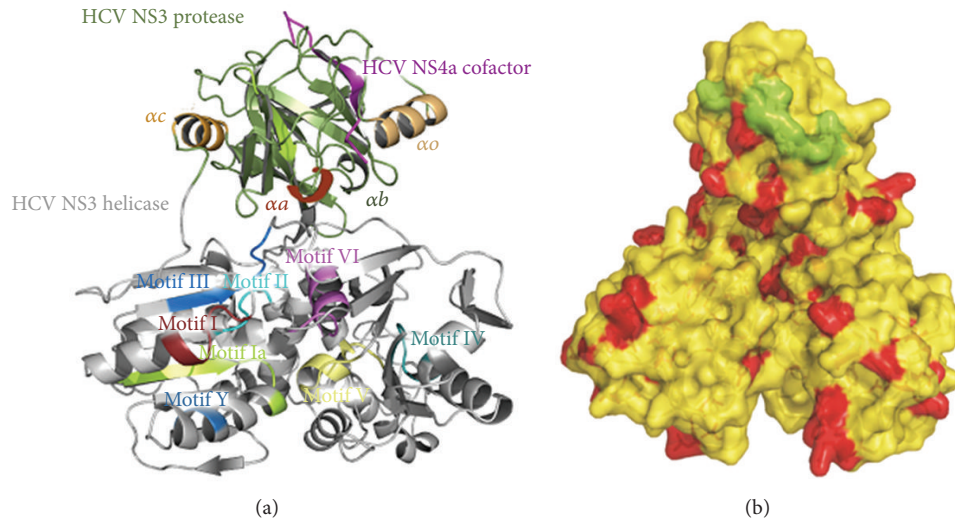
FIGURE 3: (a) Comparative structure of HCV NS3 GT3a (FJ839678). Different domains of protein along with its both protease and helicase sections. (b) Light yellow portion shows the conservations of residues while dark red portion shows the subsitutent residues in differnet crystal strucutres of HCV NS3 and NS3 GT3a.
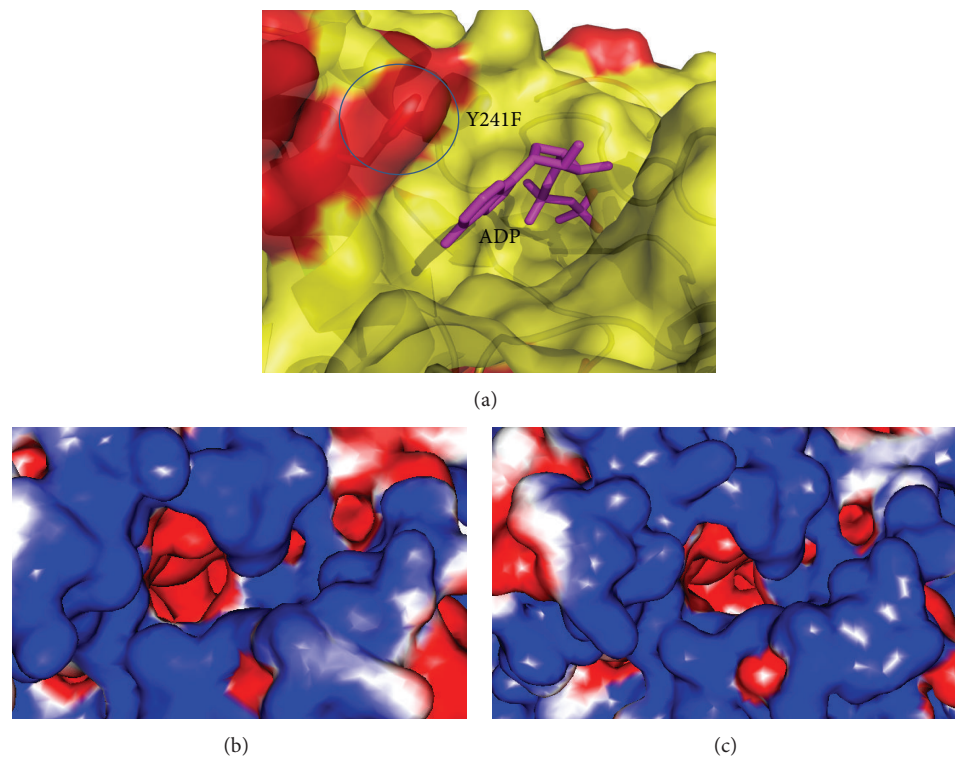

(a)

(b)   (c)

FIGURE 4: (a) ADP binding site in helicase portion of NS3 GT3a. (b) Comparison of electrostatic potentials of ADP binding site in NS3 GT3a and (c) template (PDB Id 3KQN).

The catalytic triad located in a cleft between two subdomains (or barrels), with His57 and Asp81 in the N-terminal subdomain and Ser139 in the C-terminal, has the similar behavior of IM interactions with their environment as shown in genotype 1a/1b. Along with other amino acids involved in making an active site of ATP molecules, Tyr241 and Thr419 sandwich the ADP adenine base (Figure 4). We found the mutation of Tyr241Phewhich can affect the ATP orientation and its effect on the helicase activity although the electrostatics of site produce changed environment. Main-chain nitrogen (from Gly207 to Thr212) and side-chain atoms (Lys210 and Thr212) were

TABLE 1: The intermolecular interactions shared between NS3 of HCV genotypes 1 and 3.

Residue	Interactions shared among model and templates (bonding → atom of residue)
Ala204	HB → NZ of Lys210
Lys210	HB → O of Ala204
	HB → O of Pro205
Ser231	HB → OG1 of Thr235
	HB → O of Thr416
Thr266	Hb → OH of Tyr284
Tyr267	HB → O of Ser211
Ile288	HYD → CB of Val319
Asp290	HB → OG of Ser211
	HB → OH of Tyr270
His293	HB → OG1 of Thr322
His369	HB → OH of Tyr350
Tyr392	HB → OD2 of Asp375
Phe438	HYD → CB of Phe444
	HYD → CE2 of Phe531
	HYD → CE of Met623
	HYD → CZ2 of Trp532
	HYD → CD1 of Leu627
Phe444	HYD → CZ3 of Trp532
	HYD → CD1 of Leu560
	HYD → CG2 of Ile619
	HYD → CE of MET623
Gln460	HB → NE of Arg464
Arg461	HB → OD1 of Asp410
	HB → OD1 of Asp425
Arg462	HB → O of Val331
	HB → OE2 of Glu338
Gly463	HB → NH1 of Arg467
Arg464	HB → O of Tyr418
	HB → O of Gly420
Thr465	HB → O of Phe422
	HB → O of Asp423
	HB → OG1 of Thr465
Gly466	HB → ND2 of Asn335
Arg467	HB → O of Thr419

HB = hydrogen bond, Hb = hydrogen bond, CB = carbon atom with number, O = oxygen in hydroxyl group, OG = oxygen of amino acid with some number, that is, G, OG1 = oxygen with number, that is, G1, OH = oxygen with distinct number, that is, H, CE = carbon atom of amino acid with some number, that is, E, CE2 = carbon atom with some number, that is, E2, CD1 = carbon atom with some distinct number, that is, D1, NE = nitrogen atom with number, that is, E, NH1 = nitrogen atom with some number, that is, H1, HYD = hydrophobic bond.

TABLE 2: List of nonconserved residues in HCV NS3 genotype 3a structure in comparison with other HCV NS3 genotypes. Bold residues were seen shared among templates and models.

Position of the residue	Crystal structures		Model of NS3 GT3a	
	Residue	Interactions	Residue	Interactions
219	ALA	—	VAL	VAL225, ILE248, LEU265
224	LYS	ALA283, **ASP285**	ASN	**ASP285**
286	ILE	**LEU317, VAL319**	VAL	**LEU317, VAL319**
297	SER	GLU493	ALA	—
315	ALA	—	VAL	ILE285
354	ILE	ILE359, ILE366, **ILE426**	LEU	LEU341, **ILE426**, PHE349, LEU358
358	THR	—	LEU	LEU358
377	LEU	**LEU381**, VAL409	ILE	**LEU381**
384	LEU	—	MET	ILE345
407	ILE	—	VAL	ILE357, ILE364, LEU379
430	THR	**ARG461**	VAL	**ARG461**
433	THR	—	GLU	ARG479
445	THR	—	SER	SER439
477	THR	—	SER	ASN429, ARG458
489	SER	—	VAL	PHE557, VAL436, ILE446
500	ALA	LYS551	SER	ASP496
505	THR	—	GLN	THR509
510	SER	GLU533	THR	—
530	GLU	—	ASP	SER534
553	ALA	—	GLN	LYS583
561	VAL	MET581, TRP582, LEU602	THR	PHE557
605	VAL	VAL609	ILE	VAL609, LEU598

conserved and showed similar interactions in the model of PK-NCVI/Pk3a NS3. Previous mutagenesis and arginine methylation studies have already confirmed the importance of Lys210, Arg464, and Arg467 in the ATP binding site [38–41]. Gln460 and Ala323 along with three positively charged side chains of Lys210, Arg464, and Arg467 were conserved and showed the similar interactions with γ-phosphate of ATP.

3.2. Structural Comparison of HCV NS3 Protease Domain. Protease domain in HCV NS3 is covalently connected to helicase by a loop of residues including Pro182, Val183, Phe184, Thr185, and Asp186 which forms a solvent-exposed
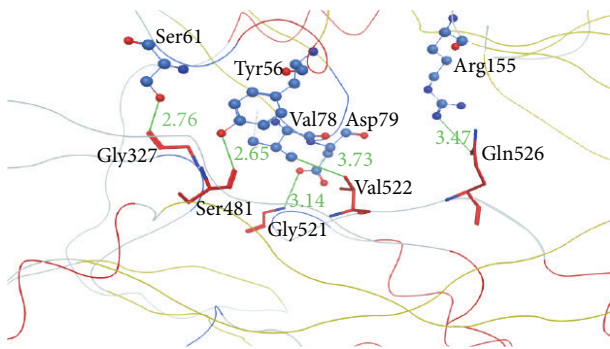
FIGURE 5: Interaction between HCV NS3 protease and helicase domains. Protease is showing in upper part of figure with atoms of residues as balls and helicase domain is the lower part of figure with atoms and bonds of involved residues shown as sticks. Secondary structure has been shown as lines.

strand in all of the crystal structures of HCV NS3. Due to the flexible nature of this loop, different residues of protease and helicase domains get involved in making intradomain interaction. Mutations at position 183 and 185 with Ser in this loop of PK-NCVI/Pk3a NS3 presented almost conserved residues for protease-helicase interaction.

The interaction site for the protease and helicase was also studied in different genotypes along with PK-NCVI/Pk3a NS3 (Figure 5). Interaction between helicase and protease normally involves Tyr56, Gly60, Ser61, Val78, and Asp79, Arg155 from the protease domain and Gly327, V329, Ser481, Pro482, Ser483, Gly484, Met485, Pro520, Gly521, Val522, Val524, and Gln526 from the side of helicase portion of NS3. These residues are conserved in all of the sequences except Val329 which is substituted with Ile329 in comparative model. The only conserved interactions seen in comparative model of NS3 for genotype 3a involved Tyr56, Ser61, Val78, and Asp79, Arg155, Gly327, Ser481, Gly521, Val522, and Gln526 (Figure 5) while rest of the IM contacts were not seen. This difference of interaction pattern between these two domains computationally confirms the orientation of protease domain with respect to helicase domain due to a flank motif connecting helicase and protease domain (Figure 5).

The residues of NS3 C-terminal 626–631 made an antiparallel β-sheet in extended conformation along the brink of protease involving His57, Lys136, Ser139, Arg155, Lys165, and Arg168 from protease domain and Asp625, Asp626, Glu628, Thr630, and Thr631 from the C-terminal in their interaction. Residues Lys136 and Ser139 were found in this interaction for the first time in our study and implicate some strong interaction of C-terminal with protease domain. Mutation of Val to Thr at position 630 showed conserved interaction pattern with protease domain. NS4a protein is involved in the completion of N-terminal protease of NS3 into barrel fold and the assembly of catalytic triad. Mutations of Ser7Ala, Gln9Arg, Gln28Val, and Glu30Thr were present at the borders of NS4a cofactor. Effects of these mutations were examined by running energy minimization protocol, which showed a different conformation for its folding. Similarly, mutations of Gln580Glu, Asp555Leu, Pro558Ser, Ala153Gln,
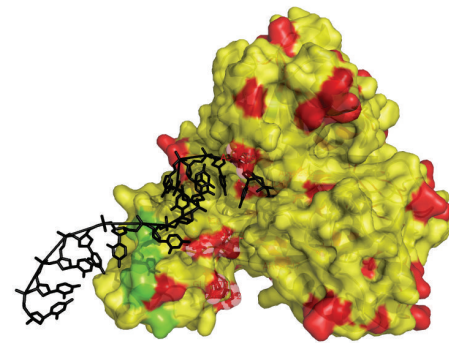
(a)
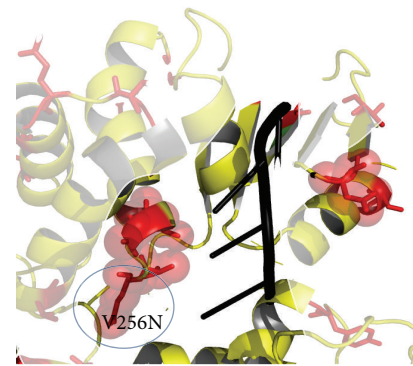
(b)

FIGURE 6: Two perspectives of ssDNA/ssRNA entry site in NS3. Black colored ssDNA was superimposed on the model using the template PDB Id 2F55. Red colored residues are the difference of residues seen in the multiple alignment of NS3GT3a with crystal studies.

Ala604Pro, Val256Asn, Thr449Arg, Thr430Val, Thr295Gln, Ser294Ala, Thr299Ser, and Ser297Ala were found to be present on either side of ssRNA pathway in helicase.

3.3. Structural Comparison of HCV NS3 Helicase Domain. The model of PK-NCVI/Pk3a was superimposed to the structure of NS3 (PDB Id. 2F55) and the residues involved in the helicase active site were characterized. The proposed ssDNA entry site in one of the dimer shown in PDB ID 2F55 can be seen in (Figure 6(a)). The entry site for ssDNA had a substitution of Val256Asn in homology model of PK-NCVI/Pk3a (Figure 6(b)). Mutations of Gln580Glu, Asp555Leu, Pro558Ser, Ala153Gln, Ala604Pro, Val256Asn, Thr449Arg, Thr430Val, Thr295Gln, Ser294Ala, Thr299Ser, and Ser297Ala were found to be present on either side of the entry of ssDNA (Figure 7). The NTP binding site was located at the periphery of the NTPase domain. The G*207*SGKST and the D*290*ECH sequences were forming lining of active site. One of the two loops connecting the RNA binding domain and NTPase had an invariant T*322*AT. His293, Thr322, and Thr324 are thought to function as a triad in coupling the helicase activity and NTP hydrolysis. We found that the
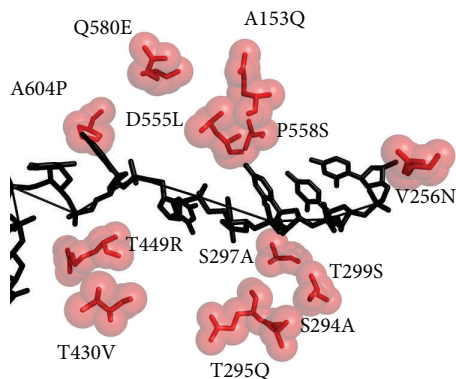
FIGURE 7: Superimposition of ssDNA from PDB Id 2F55 and NS3 GT3a model. Red sticks and spheres present the mutations seen in NS3 GT3a.
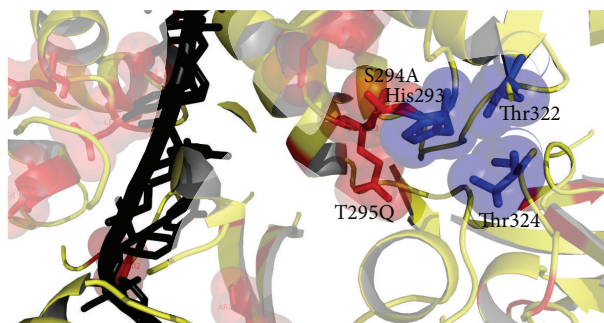
FIGURE 8: Residues Ser294Ala and Thr295Gln around coupling triad consist of His293, Thr322, and Thr324.

neighbors of His293 (i.e., Ser294Ala and Thr295Gln) were non-conserved, which were affecting its flexibility to make hydrogen bonds, NTP hydrolysis, and the helicase activity (Figure 8). In the electrostatics computations for the ssDNA entry site, the probability of nonconserved residues in both genotypes presented a difference in the size of orifice for ssDNA/ssRNA entry (Figures 9 and 10). The probability for the small opening may result in a strong grip of helicase upon ssRNA which enhanced the helicase activity.

4. Discussion

The comparative structural modeling (CM) was used to predict the structure-function connection which remained undetectable at the sequences level. Several crystal structures of HCV NS3 and mechanistic studies helped to visualize helicase and protease activity but all of the studies exclusively focused on genotype 1a or 1b of HCV. Here, we used CM to investigate whether the PK-NCVI/Pk3a NS3 of Pakistani origin is employing some new key residues. The HCV NS3 protease along with 54 residues of NS4a protein is necessary to have a correct fold. In the absence of NS4a protein, NS3 protease is partially unfolded [42]. NS4a protein is involved in the completion of N-terminal protease of
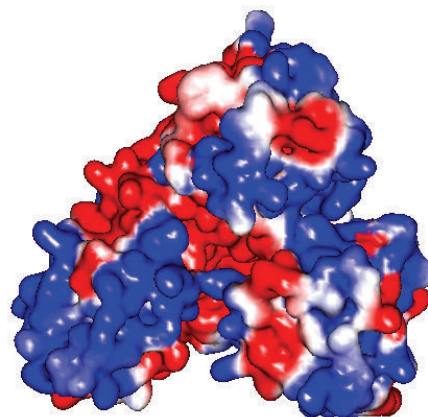
(a)
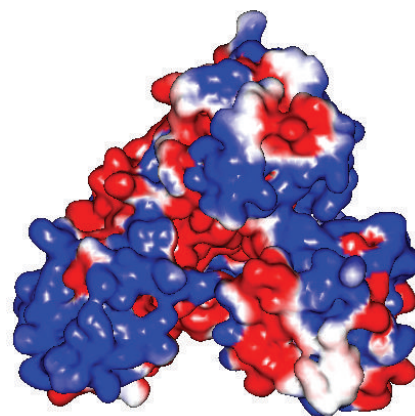
(b)

FIGURE 9: Electrostatic comparison of ssDNA/ssRNA entry site in (a) NS3 GT3a model and (b) crystal structure of HCV NS3 (PDB Id. 2F55).

NS3 into barrel fold and the assembly of catalytic triad. Mutations of Ser7Ala, Gln9Arg, Gln28Val, and Glu30Thr were present at the borders of NS4a cofactor. The energy differences were also present and crystal structures of NS3 were at higher energy level as compared to the energy of PK-NCVI/Pk3a. Although, this procedure somewhat defines the folding pattern of NS4a in NS3 protease domain but further molecular dynamics simulations are needed which will provide some new insights about the folding of NS4a and its involvement in the stabilization of NS3 protease domain. Similarly, mutations of Gln580Glu, Asp555Leu, Pro558Ser, Ala153Gln, Ala604Pro, Val256Asn, Thr449Arg, Thr430Val, Thr295Gln, Ser294Ala, Thr299Ser, and Ser297Ala were found to be present on either side of ssRNA pathway in helicase. The effect of these mutations will be confirmed and validated by running QM-MM computational techniques in the further enhancement of these studies. Crystallographic studies (PDB Id: 2F55) have already confirmed the role of NS3 dimer formation and a long stretch of ssDNA passing through this dimer [20]. The NTP binding site was located at the periphery of the NTPase domain. The G207SGKST and the D290ECH sequences were forming lining of active site. The NTPase

(a)

(b)
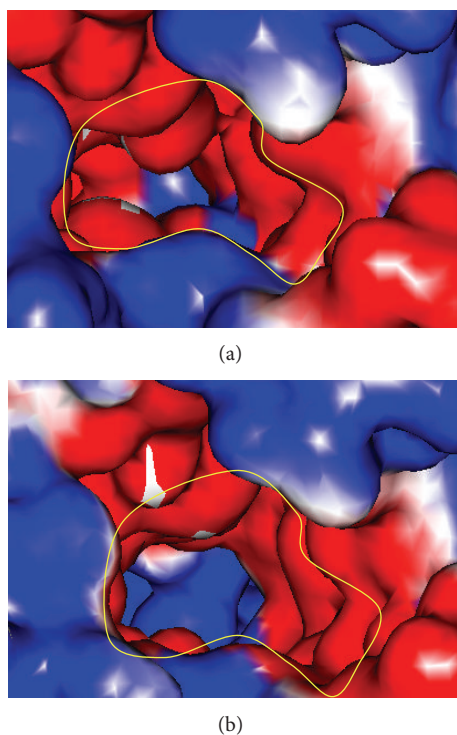
Figure 10: Comparison of internal space between the entry site of ssDNA/ssRNA in (a) NS3-GT3a model and (b) crystal structure of HCV NS3 without docked ssDNA/RNA (PDB Id. 3O8B).

had an invariant T322AT. His293, Thr322, and Thr324 are thought to function as a triad in coupling the helicase activity and NTP hydrolysis. These nonconserved amino acids in the helicase portion may be involved in the strong binding of ssDNA/ssRNA, which may enhance the helicase efficiency of enzyme. The strong grip of helicase might be involved in the rate of ssDNA/ssRNA unwinding. Involvement of nonconserved residues is necessary for the determination of their roles in the rate of unwinding and determination of key residues for helicase inhibitors.

Analysis of the final model and the underlying alignment among PK-NCVI/Pk3a NS3 and reported genotypes of NS3 indicates that the mutations occur in the nonfunctional regions of PK-NCVI/Pk3a NS3 and almost conserved interactions between the residues have been seen. Hence, the substituted residues in PK-NCVI/Pk3a NS3 are not part of the functional regions of protein and are involved in the structure stability [11]. However, experimental techniques should be used to confirm the implications of structural differences and to investigate the congruency of our *in silico* results.

The IFN treatment response against the HCV genotype 3 is higher (~80%) in comparison to other strains (40%) [43]. However the new class of protease inhibitors seems to be less efficacious against genotype 3a, requiring higher concentration to achieve threshold level of inhibition as measured by qPCR of viral load. This may be due to the differences in the amino acids composition between genotype 1 versus genotype 3, as we have reported thus generating variable architecture and variable response. The biological relevance

of the genotypic differences in NS3 has not been addressed in this study, but several intriguing ideas can be gleaned from the resultant data. These regions have less frequent mutations and possible substitutions are either positive in nature or have similar IM contact architecture. Although such ideas are highly speculative but they could be feasibly tested by measuring mutation rates in HCV replicon systems and could be measured in the presence and absence of ribavirin or other antiviral compounds. Altogether, this exercise reflected the interaction patterns of substituted residues along with implications for PK-NCVI/Pk3a NS3 in its role of disease profile and further studies with other genotypes should be carried out to define a complete map of IM noncovalent interactions responsible for the variable disease transmission, severity, and resistance to antiviral therapies.

Abbreviations

HCV: Hepatitis C virus
NS3: Nonstructural protein 3
kb: Kilo base
PK: Pakistan
3a: Genotype 3a
NS3: Nonstructural protein 3
DNA: Deoxy ribonucleic acid
ssRNA: Single stranded ribonucleic acid
QM/MM: Quantum mechanics/molecular mechanics
PDB: Protein data bank
IM: Intermolecular
ATP: Adenosine triphosphate
ADP: Adenine dinucleotide
NTPase: RNA nucleoside triphosphatases
NTP: Nucleoside triphosphate.

Conflict of Interests

The authors declare that they have no competing interests.

Authors' Contribution

Ishtiaq Qadri conceived the idea, designed the project, and analysed results. Kaneez Fatima performed the experimental work. Kaneez Fatima wrote the paper and Ishtiaq Qadri and Esam Azhar helped in literature review for the paper. Shilu Mathew and Ghazi Damanhouri did the copy editing and proof reading of the paper.

Acknowledgment

References

[1] A. L. Schmaljohn and D. McClain, "Alphaviruses (Togaviridae) and Flaviviruses (Flaviviridae)," in *Medical Microbiology*, S. Baron and T. X. Galveston, Eds., The University of Texas Medical Branch, Galveston, Tex, USA, 4th edition, 1996.

[2] K. Reed and C. Rice, "Overview of hepatitis C virus genome structure, polyprotein processing, and protein properties," *Current Topics in Microbiology and Immunology*, vol. 242, pp. 55–84, 2000.

[3] E. Gower, C. C. Estes, S. Hindman, K. Razavi-Shearer, and H. Razavi, "Global epidemiology and genotype distribution of the hepatitis C virus," *Journal of Hepatology*, 2014.

[4] J. F. Quarleri, M. V. Bussy, V. L. Mathet et al., "In vitro detection of dissimilar amounts of hepatitis C virus (HCV) subtype-specific RNA genomes in mixes prepared from sera of persons infected with a single HCV genotype," *Journal of Clinical Microbiology*, vol. 41, no. 6, pp. 2727–2733, 2003.

[5] S. Z. Safi, Y. Badshah, Y. Waheed et al., "Distribution of hepatitis C virus genotypes, hepatic steatosis and their correlation with clinical and virological factors in Pakistan," *Asian Biomedicine*, vol. 4, no. 2, pp. 253–262, 2010.

[6] A. M. Di Bisceglie and J. H. Hoofnagle, "Optimal therapy of hepatitis C," *Hepatology*, vol. 36, supplement 5, pp. S121–S127, 2002.

[7] N. S. Raja and K. A. Janjua, "Epidemiology of hepatitis C virus infection in Pakistan," *Journal of Microbiology, Immunology and Infection*, vol. 41, no. 1, pp. 4–8, 2008.

[8] M. Idrees and S. Riazuddin, "A study of best positive predictors for sustained virologic response to interferon alpha plus ribavirin therapy in naive chronic hepatitis C patients," *BMC Gastroenterology*, vol. 9, article 5, 2009.

[9] H. Aziz, M. L. Gil, Y. Waheed et al., "Evaluation of prognostic factors for Peg Interferon alfa-2b plus ribavirin treatment on HCV infected patients in Pakistan," *Infection, Genetics and Evolution*, vol. 11, no. 3, pp. 640–645, 2011.

[10] H. Aziz, U. Gill, A. Raza, and M. L. Gill, "Metabolic syndrome is associated with poor treatment response to antiviral therapy in chronic hepatitis C genotype 3 patients," *European Journal of Gastroenterology and Hepatology*, vol. 26, no. 5, pp. 538–543, 2014.

[11] K. Fatima, M. Tahir, and I. Qadri, "Development of robust in vitro serine protease assay based on recombinant Pakistani HCV NS3-4A protease," *Virus Research*, vol. 160, no. 1-2, pp. 230–237, 2011.

[12] C.-L. Tai, W.-K. Chi, D.-S. Chen, and L.-H. Hwang, "The helicase activity associated with hepatitis C virus nonstructural protein 3 (NS3)," *Journal of Virology*, vol. 70, no. 12, pp. 8477–8484, 1996.

[13] P. S. Pang, E. Jankowsky, P. J. Planet, and A. M. Pyle, "The hepatitis C viral NS3 protein is a processive DNA helicase with cofactor enhanced RNA unwinding," *The EMBO Journal*, vol. 21, no. 5, pp. 1168–1176, 2002.

[14] T. C. Appleby, R. Anderson, O. Fedorova et al., "Visualizing ATP-dependent RNA translocation by the NS3 helicase from HCV," *Journal of Molecular Biology*, vol. 405, no. 5, pp. 1139–1153, 2011.

[15] N. Yao, P. Reichert, S. S. Taremi, W. W. Prosise, and P. C. Weber, "Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease-helicase," *Structure*, vol. 7, no. 11, pp. 1353–1363, 1999.

[16] S. H. Liaw, S. J. Chen, C. Y. Hu et al., "Crystal structure of hepatitis C virus NS3 helicase with a novel inhibitor," http://www.pdb.org/pdb/explore/explore.do?structureId=2ZJO.

[17] J. L. Kim, K. A. Morgenstern, J. P. Griffith et al., "Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding," *Structure*, vol. 6, no. 1, pp. 89–100, 1998.

[18] N. Yao, T. Hesson, M. Cable et al., "Structure of the hepatitis C virus RNA helicase domain," *Nature Structural Biology*, vol. 4, no. 6, pp. 463–467, 1997.

[19] M. Gu and C. M. Rice, "Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 2, pp. 521–528, 2010.

[20] S. G. Mackintosh, J. Z. Lu, J. B. Jordan et al., "Structural and biological identification of residues on the surface of NS3 helicase required for optimal replication of the hepatitis C virus," *The Journal of Biological Chemistry*, vol. 281, no. 6, pp. 3528–3535, 2006.

[21] H.-S. Cho, N.-C. Ha, L.-W. Kang et al., "Crystal structure of RNA helicase from genotype 1b hepatitis C virus. A feasible mechanism of unwinding duplex RNA," *Journal of Biological Chemistry*, vol. 273, no. 24, pp. 15045–15052, 1998.

[22] M. K. Levin and S. S. Patel, "The helicase from hepatitis C virus is active as an oligomer," *The Journal of Biological Chemistry*, vol. 274, no. 45, pp. 31839–31846, 1999.

[23] M. K. Levin and S. S. Patel, "Helicase from hepatitis C virus, energetics of DNA binding," *The Journal of Biological Chemistry*, vol. 277, no. 33, pp. 29377–29385, 2002.

[24] G. A. Locatelli, G. Gosselin, S. Spadari, and G. Maga, "Hepatitis C virus NS3 NTPase/helicase: different stereoselectivity in nucleoside triphosphate utilisation suggests that NTPase and helicase activities are coupled by a nucleotide-dependent rate limiting step," *Journal of Molecular Biology*, vol. 313, no. 4, pp. 683–694, 2001.

[25] K. A. Morgenstern, J. A. Landro, K. Hsiao et al., "Polynucleotide modulation of the protease, nucleoside triphosphatase, and helicase activities of a hepatitis C virus NS3-NS4A complex isolated from transfected COS cells," *Journal of Virology*, vol. 71, no. 5, pp. 3767–3775, 1997.

[26] D. J. T. Porter, S. A. Short, M. H. Hanlon et al., "Product release is the major contributor to k_{cat} for the hepatitis C virus helicase-catalyzed strand separation of short duplex DNA," *The Journal of Biological Chemistry*, vol. 273, no. 30, pp. 18906–18914, 1998.

[27] F. Preugschat, D. R. Averett, B. E. Clarke, and D. J. T. Porter, "A steady-state and pre-steady-state kinetic analysis of the NTPase activity associated with the hepatitis C virus NS3 helicase domain," *Journal of Biological Chemistry*, vol. 271, no. 40, pp. 24449–24457, 1996.

[28] S. F. Altschul, T. L. Madden, A. A. Schäffer et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Research*, vol. 25, no. 17, pp. 3389–3402, 1997.

[29] M. A. Larkin, G. Blackshields, N. P. Brown et al., "Clustal W and Clustal X version 2.0," *Bioinformatics*, vol. 23, no. 21, pp. 2947–2948, 2007.

[30] C. Notredame, D. G. Higgins, and J. Heringa, "T-coffee: a novel method for fast and accurate multiple sequence alignment," *Journal of Molecular Biology*, vol. 302, no. 1, pp. 205–217, 2000.

[31] J. D. Thompson, J. C. Thierry, and O. Poch, "RASCAL: rapid scanning and correction of multiple sequence alignments," *Bioinformatics*, vol. 19, no. 9, pp. 1155–1161, 2003.

[32] F. Plewniak, L. Bianchetti, Y. Brelivet et al., "PipeAlign: a new toolkit for protein family analysis," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3829–3832, 2003.

[33] M. A. Marti-Renom, A. C. Stuart, A. Fiser, R. Sanchez, F. Melo, and A. Sali, "Comparative protein structure modeling of genes and genomes," *Annual Review of Biophysics and Biomolecular Structure*, vol. 29, pp. 291–325, 2000.

[34] A. Sali and T. L. Blundell, "Comparative protein modelling by satisfaction of spatial restraints," *Journal of Molecular Biology*, vol. 234, no. 3, pp. 779–815, 1993.

[35] L. A. Kelley, S. P. Gardner, and M. J. Sutcliffe, "An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies," *Protein Engineering*, vol. 9, no. 11, pp. 1063–1065, 1996.

[36] C. Colovos and T. O. Yeates, "Verification of protein structures: patterns of nonbonded atomic interactions," *Protein Science*, vol. 2, no. 9, pp. 1511–1519, 1993.

[37] Accelrys, *Discovery Studio. 2.8 Edition*, Accelrys, San Diego, Calif, USA, 2008.

[38] C.-T. Tai, W.-C. Pan, S.-H. Liaw, U.-C. Yang, L.-H. Hwang, and D.-S. Chen, "Structure-based mutational analysis of the hepatitis C virus NS3 helicase," *Journal of Virology*, vol. 75, no. 17, pp. 8289–8297, 2001.

[39] G. M. Heilek and M. G. Peterson, "A point mutation abolishes the helicase but not the nucleoside triphosphatase activity of hepatitis C virus NS3 protein," *Journal of Virology*, vol. 71, no. 8, pp. 6264–6266, 1997.

[40] D. W. Kim, J. Kim, Y. Gwack, J. H. Han, and J. Choe, "Mutational analysis of the hepatitis C virus RNA helicase," *Journal of Virology*, vol. 71, no. 12, pp. 9400–9409, 1997.

[41] F. H. T. Duong, V. Christen, J. M. Berke, S. H. Penna, D. Moradpour, and M. H. Heim, "Upregulation of protein phosphatase 2Ac by hepatitis C virus modulates NS3 helicase activity through inhibition of protein arginine methyltransferase 1," *Journal of Virology*, vol. 79, no. 24, pp. 15342–15350, 2005.

[42] J. L. Kim, K. A. Morgenstern, C. Lin et al., "Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide," *Cell*, vol. 87, no. 2, pp. 343–355, 1996.

[43] P. J. S. Provazzi, H. A. Arcuri, I. M. V. G. de Carvalho-Mello et al., "Structural studies of Helicase NS3 variants from Hepatitis C virus genotype 3 in virological sustained responder and non-responder patients," *BMC Research Notes*, vol. 3, article 196, 2010.