



OPEN

## An in silico approach to analyze HCV genotype-specific binding-site variation and its effect on drug–protein interaction

Ramsha Khalid<sup>1</sup>, Muhammad Faraz Anwar<sup>2</sup>, Muhammad Aanish Raees<sup>3</sup>, Sadaf Naeem<sup>1</sup>, Syed Hani Abidi<sup>1</sup> & Syed Ali<sup>4</sup>

Genotype variation in viruses can affect the response of antiviral treatment. Several studies have established approaches to determine genotype-specific variations; however, analyses to determine the effect of these variations on drug–protein interactions remain unraveled. We present an in-silico approach to explore genotype-specific variations and their effect on drug–protein interaction. We have used HCV NS3 helicase and fluoroquinolones as a model for drug–protein interaction and have investigated the effect of amino acid variations in HCV NS3 of genotype 1a, 1b, 2b and 3a on NS3-fluoroquinolone interaction. We retrieved 687, 667, 101 and 248 nucleotide sequences of HCV NS3 genotypes 1a, 1b, 2b, and 3a, respectively, and translated these into amino acid sequences and used for genotype variation analysis, and also to construct 3D protein models for 2b and 3a genotypes. For 1a and 1b, crystal structures were used. Drug–protein interactions were determined using molecular docking analyses. Our results revealed that individual genotype-specific HCV NS3 showed substantial sequence heterogeneity that resulted in variations in docking interactions. We believe that our approach can be extrapolated to include other viruses to study the clinical significance of genotype-specific variations in drug–protein interactions.

Genetic variability in the viral population can greatly accelerate the rate of viral evolution in response to different selection pressures<sup>1,2</sup>. Therefore, the characterization of genomic variations in viral variants is essential to understand various aspects of evolution, persistence, epidemiology, immune escape, and development of antiviral drug resistance<sup>1–4</sup>. Additionally, this analysis can identify crucial variations/mutations associated with variability in antiviral treatment response.

Hepatitis C virus (HCV) can serve as a model to study genotype-specific mutations as the clinical significance of HCV genotypes in treatment response is well established<sup>5</sup>. Studies have suggested that genetic variability in HCV genotypes can greatly affect T-cell mediated immune response<sup>6,7</sup>. Furthermore, the sequence heterogeneity in the HCV genome leads to the generation of HCV genotype-specific epitopes that can potentially evade immune surveillance<sup>6,7</sup>. Genotypic variation is also known to play a role in determining the course of disease progression and antiviral response<sup>6</sup>. In the recent past, the combination of pegylated interferon-alpha and a nucleotide analog ribavirin-based regimen was the standard care of treatment for HCV infection. HCV genotype 1 was considered to be the most challenging genotype to treat with this regimen. This regimen gave sustained virologic responses (SVRs) in 40%-50% of HCV genotype 1 infected patients as compared to 75%-80% in HCV genotype 2 and 3 infected patients<sup>5</sup>.

In this paper, we present an in-silico approach to explore genotype-specific amino acid variations and their effect on drug–protein interaction. In our study, HCV NS3 helicase was used as a model drug target whereas fluoroquinolones were used as model drugs. HCV NS3 is a multifunctional enzyme that belongs to DExH Box RNA helicases of superfamily 2 and exhibits NTP-mediated nucleic acid unwinding activity. NS3 plays a pivotal role in HCV biology particularly in the viral replicative cycle, viral assembly, persistence, and pathogenesis<sup>8–10</sup>. Additionally, NS3 is well characterized in terms of structure, domains, functionally active amino acid residues,

<sup>1</sup>Department of Biochemistry, University of Karachi, Karachi, Pakistan. <sup>2</sup>Department of Biological and Biomedical Sciences, Aga Khan University, Karachi, Pakistan. <sup>3</sup>Cincinnati Children's Hospital Medical Center, Ohio, USA. <sup>4</sup>Department of Biological Sciences, Nazarbayev University School of Medicine, Nazarbayev University, Astana, Kazakhstan. ✉email: m.haniabidi@gmail.com; syed.ali@nu.edu.kz

HCV genotypes	1a	1b	2b	3a
<b>Full length</b>				
1a	–	93.4	82.1	81.5
1b	93.4	–	82.9	82.8
2b	82.1	82.9	–	80.8
3a	81.5	82.8	80.8	–
<b>Fluoroquinolone binding region</b>				
1a	–	94.0	87.3	83.3
1b	94.0	–	88.1	85
2b	87.3	88.1	–	85.3
3a	83.3	85.0	85.3	–

**Table 1.** Identity matrix of amino acid sequences of NS3 HCV of genotype 1a, 1b, 2b and 3a: Sequence identity was estimated in full length and fluoroquinolone binding region sequences by constructing an ID matrix using the ‘BioEdit’ software. The identity scores are expressed as percentages.

nucleic acid interaction and unwinding mechanism. However, the genotype-specific variations in HCV NS3 and their effect on NS3 structure, and subsequently on drug–protein interactions, are relatively unexplored.

Fluoroquinolones are broad-spectrum antibiotics that target DNA gyrase and topoisomerase IV—bacterial enzymes involved in DNA replication—forming an irreversible drug–protein–nucleic acid complex<sup>11</sup>. Previously reported studies from our group and others have demonstrated that fluoroquinolones actively inhibit HCV replication by targeting viral protein NS3helicase. Since bacterial gyrase and viral helicase are functionally homologous; it has been hypothesized that fluoroquinolones inhibit viral helicases in a similar fashion<sup>12–15</sup>. To explore how genotype-specific variations in HCV NS3 might affect the inhibition of its activity by fluoroquinolones, we have investigated the effect of active site residue variations in HCV genotype 1a, 1b, 2b and 3a on NS3–fluoroquinolone interactions.

## Results

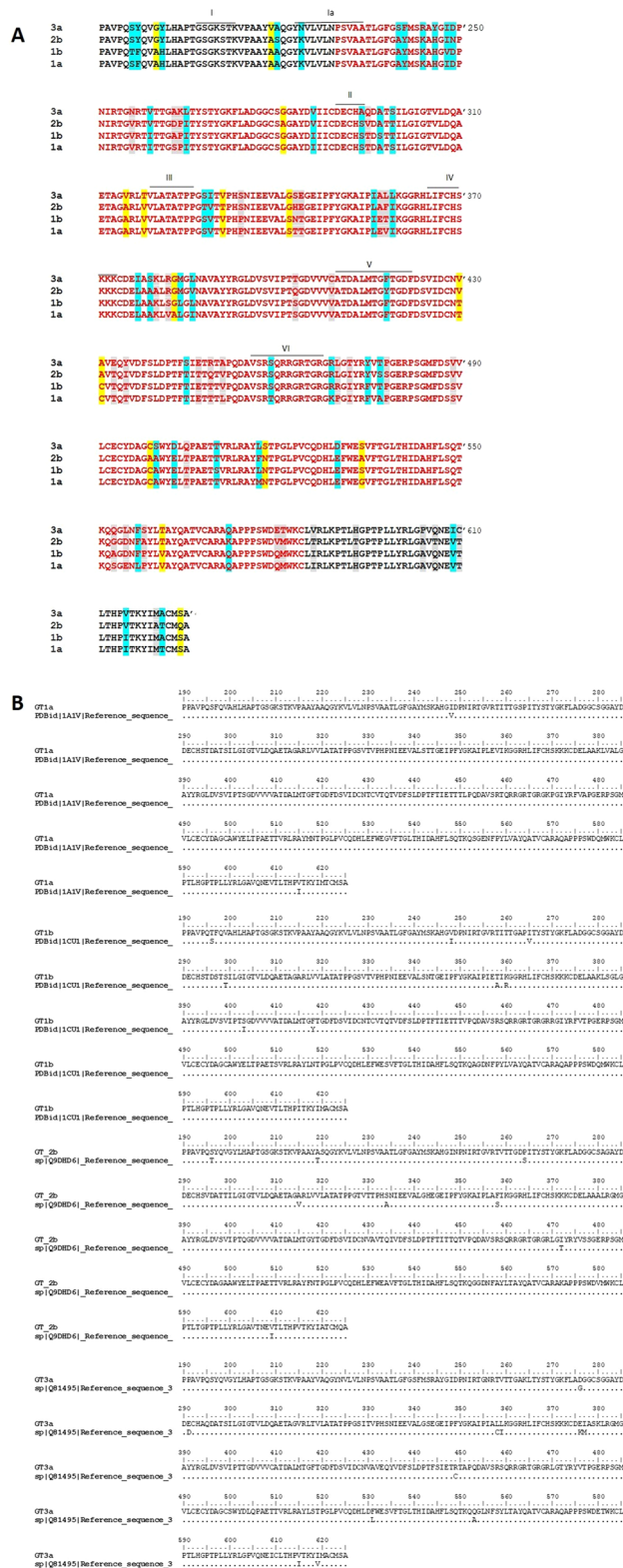
**Inter-genotype sequence comparison.** To explore the inter-genotype sequence similarity in NS3 protein, multiple sequence alignment was performed using NS3 sequences from HCV genotypes 1a, 1b, 2b, and 3a, followed by the construction of sequence identity matrix. The analysis was carried out in two different ways, one using the full-length protein sequence (Table 1a) and the other employing only the fluoroquinolone binding region sequence (Table 1b). Both the approaches revealed an overall high degree of sequence similarity between NS3 helicases from genotypes 1a and 1b, i.e. 93.4% in the full-length protein sequence (Table 1a) and 94% in the fluoroquinolone binding region site (Table 1b). When compared to genotypes 2b and 3a, genotypes 1a and 1b revealed sequence identity scores ranging from 81.5–82.9% for full length protein sequence (Table 1a) and 83.3–88.1% for fluoroquinolone binding region site (Table 1b). Between genotypes 2b and 3a, sequence identity of 80.8% and 85.3% were observed, respectively, for full length and fluoroquinolone binding region site sequences (Table 1a and b).

Before inter-genotype mutation analysis (Fig. 1A), the genotype-specific consensus sequences were compared with the available reference sequences. The analysis revealed that the consensus and reference sequences were identical, with exception of 1–2 sites (Fig. 1B), validating that the consensus sequences used in our study represent ‘true’ genotype-specific sequences.

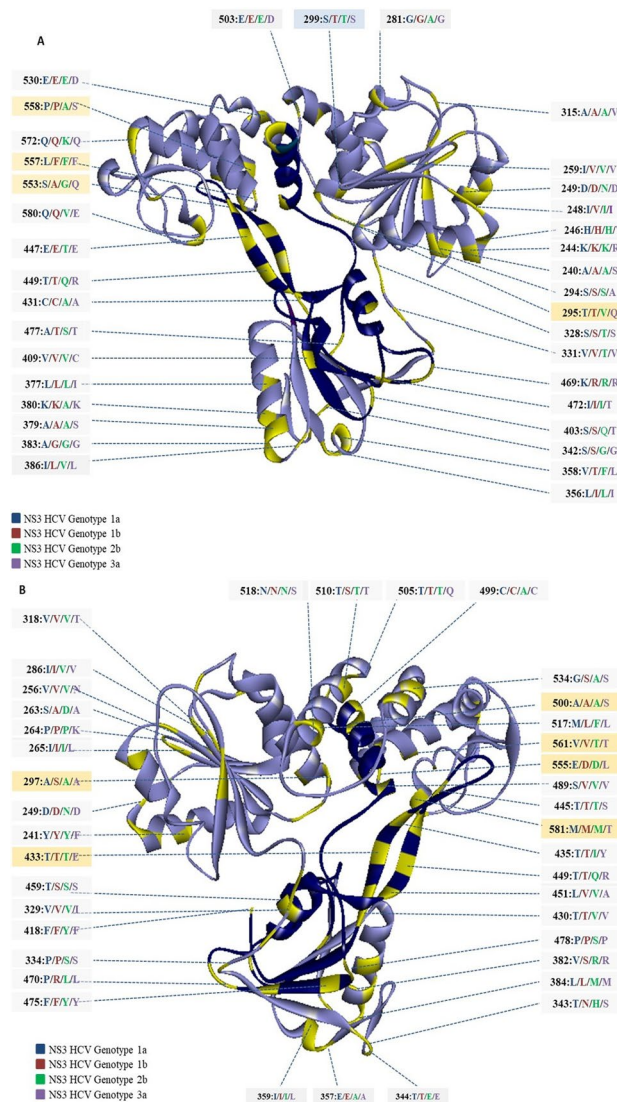
Mutational analysis of the fluoroquinolone binding region from genotypes 1a, 1b, 2b, and 3a, revealed a total of 90 and 73 variable sites, where genotype 3a revealed the most genotype-specific variations (Figs. 1A and 2). Fluoroquinolone binding region of NS3 sequences contained 33 conservative, 12 semi-conservative and 28 non-conservative amino acid substitutions amongst the four genotypes. Out of these, 3 positions namely 343 T/N/H/S (in genotype 1a/1b/2b/3a), 358V/T/F/L and 553S/A/G/Q showed variable amino acids across all genotypes (Figs. 1A and 2).

Our analysis revealed a high degree of conservation in helicase signature motifs across the four genotypes, except for motif V and motif VI (Fig. 1A), wherein certain genotypes a single amino acid substitution was observed (Fig. 1A). In motif V, F418Y variation was noted, where Y variant was found in genotype 2b; similarly, in motif VI, T459S was observed in genotype 1b, 2b, 3a, while K469R was observed in genotype 1b, 2b, 3a (Figs. 1A and 2). As mentioned in Methods, for this and other analyses 1A1V was used as the reference sequence.

**Validation of 3D structures and docking strategy and genotype-specific variations in NS3–fluoroquinolone interactions.** Before docking the NS3 structures were verified and validated using Verify 3D, GROMACS, and Ramachandran plot analysis. Structures constructed from three different tools were comparable, where structures constructed using Phyre 2 and Swiss-Model gave the lowest RMSD values of 0.41 and 0.0, respectively for genotype 2b and 0.42, and 0.07 for genotype 3a, respectively, when compared to the template structure. CPH tool has higher RMSD values of 1.55 and 1.53 for genotypes 2b and 3a, respectively, when compared to the template structures. All structures passed the 3D verification (performed using Verify 3D software) as at least 80% of the amino acids have scored  $\geq 0.2$  in the 3D/1D profile. Similarly, the structures were valid on the Ramachandran plot as most of the amino acids were under the permissible regions (Supplementary Fig. 1).



**Figure 1.** Sequence alignment of HCV NS3 genotypes 1a, 1b, 2b, 3a: **(A)** Multiple sequence alignment was performed employing ‘Clustal W’ using the amino acid sequence in the fluoroquinolone binding region (Pro230-Cys584). The fluoroquinolone binding region is highlighted in the red font. Conservative helicase motifs (I-VI) are indicated by black lines on top of the alignment. Conservative, semi-conservative and non-conservative amino acid variations are marked in blue, yellow and grey colors, respectively. **(B)** Pairwise sequence alignment was performed employing ‘Clustal W’ using 1a, 1b, 2b, 3a amino acid sequence and their respective reference sequences 1a (PDB ID:1A1V), 1B (PDBid:1CU1), 2B (UNIPROT ID: Q9DHD6) and 3A (UNIPROT ID: Q81495).



**Figure 2.** Structure of HCV NS3 with genotype-specific amino acid substitutions across HCV variants 1a, 1b, 2b, and 3a. Ribbon Diagram of (A) front view, and (B) posterior view of NS3 reference structure (PDB ID: 1A1V). Within the fluoroquinolone binding region (P230-C584), amino acid residues variable across HCV genotypes are highlighted in yellow color, while the substrate-binding site, R393-W501, is shaded dark blue. Amino acids of genotype 1a, 1b, 2b, and 3a are shown in the blue, red, green and purple font, respectively. A total of 73 amino acid variations were observed. To improve visualization, both the front and posterior views of NS3 were examined. Amino acids that were found to interact with fluoroquinolones are shown in orange boxes.

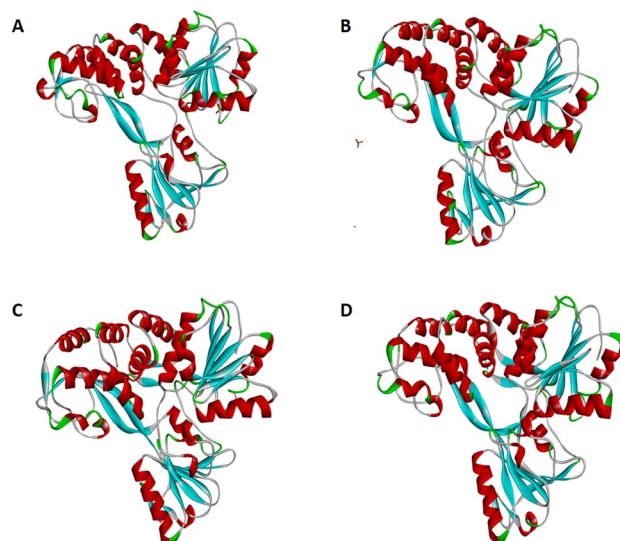
Based on the lowest RMSD values, overall better quality (based on Verify 3D and Ramachandran plot assessment) models developed using Swiss-Model was used in further analyses (Fig. 3). The structural comparison revealed the structures of the four genotypes to be quite similar (Fig. 3). The RMSD values between structures of genotype 1a-1b, 1a-2b, 1a-3a, 1b-2b, 1b-3a, 2b-3a were found to be 0.01, 0.03, 0.03, 0.03, 0.03, and 0.04, respectively, suggesting strong homology between the NS3 structures from the four genotypes.

Similarly, validation of the docking approach was done by performing blind docking (assuming drug binding site to be anywhere on the protein) on two previously reported complexes of NS3 bound to inhibitors, one a natural analog and other inhibitor ITMN-3479. Our results revealed that for both the models Molegro predicted the exact binding site/pose and gave the similar drug-protein interaction (for natural analog: Asp 454, Gln 481 and Cys 431; for ITMN-3479: Gly255, Thr269, Trp501) as reported for the reference models (Fig. 4 A and B).

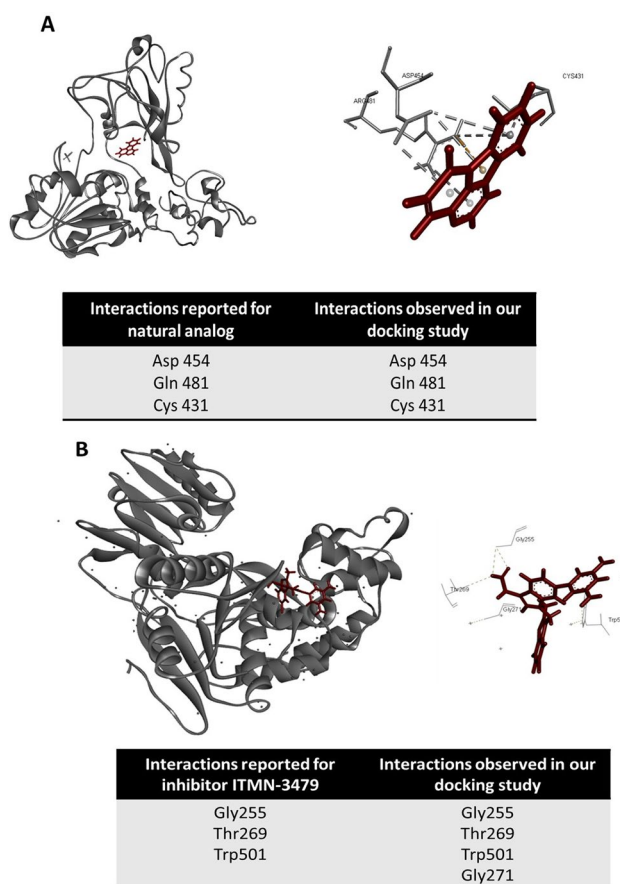
To investigate the effect of genotype-specific variations on NS3-fluoroquinolones interactions, molecular docking simulations were performed. Eight fluoroquinolones were docked individually on the consensus NS3 helicase structure constructed for each genotype.

Comparative analysis of the drug-protein interactions and binding poses revealed that almost all fluoroquinolones interacted with NS3 from all genotypes essentially in the same binding pocket, comprising amino acids





**Figure 3.** Structures of NS3 HCV: Structures of HCV NS3 used in the study are given. Structures for (A) 1a (PDB ID: 1A1V) and (B) 1B (PDBid:1CU1) were retrieved from the PDB database, while structures of (C) 2b and (D) 3a were constructed using homology modeling approach using CPH model, Swiss model, and Phyre 2 software. Models shown in this figure were constructed using Swiss model.



**Figure 4.** Validation of Docking strategy: Docking strategy was validated by re-docking the two previously published ligands (A) natural analog and (B) inhibitor ITMN-3479 on their respective ligands. (A) and (B) Poses of ligands bound to their receptors are shown, while tables at the bottom show amino acid interactions reported for each ligand and observed in our study. Ligand is shown in red, while protein is shown in dark grey color.

Pro230-Cys584 (Figs. 5 and 6; Table 2; Supplementary Table 2). However, in some of the poses, fluoroquinolones, such as Balofloxacin, Ciprofloxacin, and Levofloxacin, etc., exhibited a more scattered interaction with NS3 from genotype 1b (Fig. 6). No amino acid was found to form interactions with all fluoroquinolones, however, certain amino acids appeared to be crucial in the drug–protein interaction (Table 2; Supplementary Table 2). The amino acid residues that frequently made interactions with most fluoroquinolones were Arg481, Asp454, His293, and Gln434 (Table 2). Position 295 in all genotypes appeared to be important in drug–protein interaction as amino acids at this position, despite inter-genotype variation (genotypes 1a and 1b: Thr295; genotype 2b: Val295; genotype 3a: Gln295), interacted with almost all fluoroquinolones (Figs. 5 and 6; Table 2). In addition to position 295, certain genotype-specific residues commonly interacted with each of the fluoroquinolones; these residues include Leu451, Ser457, Asp454, and Arg481 in genotype 1a, and Glu493, and Ser297 in 1b. In contrast, for genotypes 2b and 3a, different residues were involved in forming interactions with different fluoroquinolones (Figs. 5 and 6; Table 2; Supplementary Table 2).

Analysis of the docking scores between NS3 (from different genotype) and different fluoroquinolone showed the docking scores for Balofloxacin-NS3 ranged from  $-98$  to  $-111$ ; Ciprofloxacin-NS3 ranged from  $-103$  to  $-114$ ; Enrofloxacin-NS3 ranged from  $-121$  to  $-100$ ; Lomefloxacin-NS3 ranged from  $-90$  to  $-101$ ; Levofloxacin-NS3 ranged from  $-90$  to  $-115$ ; Ofloxacin-NS3 ranged from  $-88$  to  $-106$ ; Sparfloxacin-NS3 ranged from  $-87$  to  $-104$ ; and Pefloxacin-NS3 ranged from  $-87$  to  $-109$  (Table 3).

## Discussion

In the current study, we present an in-silico approach to assess the effect of genotype-specific amino acid variations on protein folding and structural architecture and its interaction with the drugs. Our strategy involved computational analysis of NS3 helicase sequence and structures followed by molecular docking of fluoroquinolones on genotype-specific NS3 structures.

In the past era, a substantial number of studies explored the genetic heterogeneity of viral genotypes and its correlation with various aspects of the viral life cycle, including replication, and pathogenesis. Additionally, efforts have been made to unravel inter-genotype clinical and serological variations, functional differences in viral proteins and their effect on host immune system, generation of escape variants and viral epitopes, and on antiviral treatment response<sup>16–20</sup>. For example, the study conducted by Kaneez et al. showed amino acid variability in HCV NS3 1 and 3a and its impact on the structural architecture of the protein. The study also emphasized on the functionally active and conservative residues across selected HCV genotypes. To do so, several bioinformatics-based approaches and computational methods were employed. They reported that genotype-specific variations affected the structural architecture of NS3<sup>21</sup>. Similarly, Ahmed et al.<sup>22</sup> established the impact of sequence heterogeneity within NS5A and core regions on the response to pegylated interferon/ribavirin (PEG-IFN/RBV) therapy against HCV. The results showed a significant association between heterogeneity in IFN/RBV resistance-determining region IRRDR of NS5A and SVR, indicating that genetic heterogeneity in IRRDR can potentially serve as a predictor for SVR in HCV-infected patients treated with PEG-IFN/RBV combinatorial therapy<sup>22</sup>.

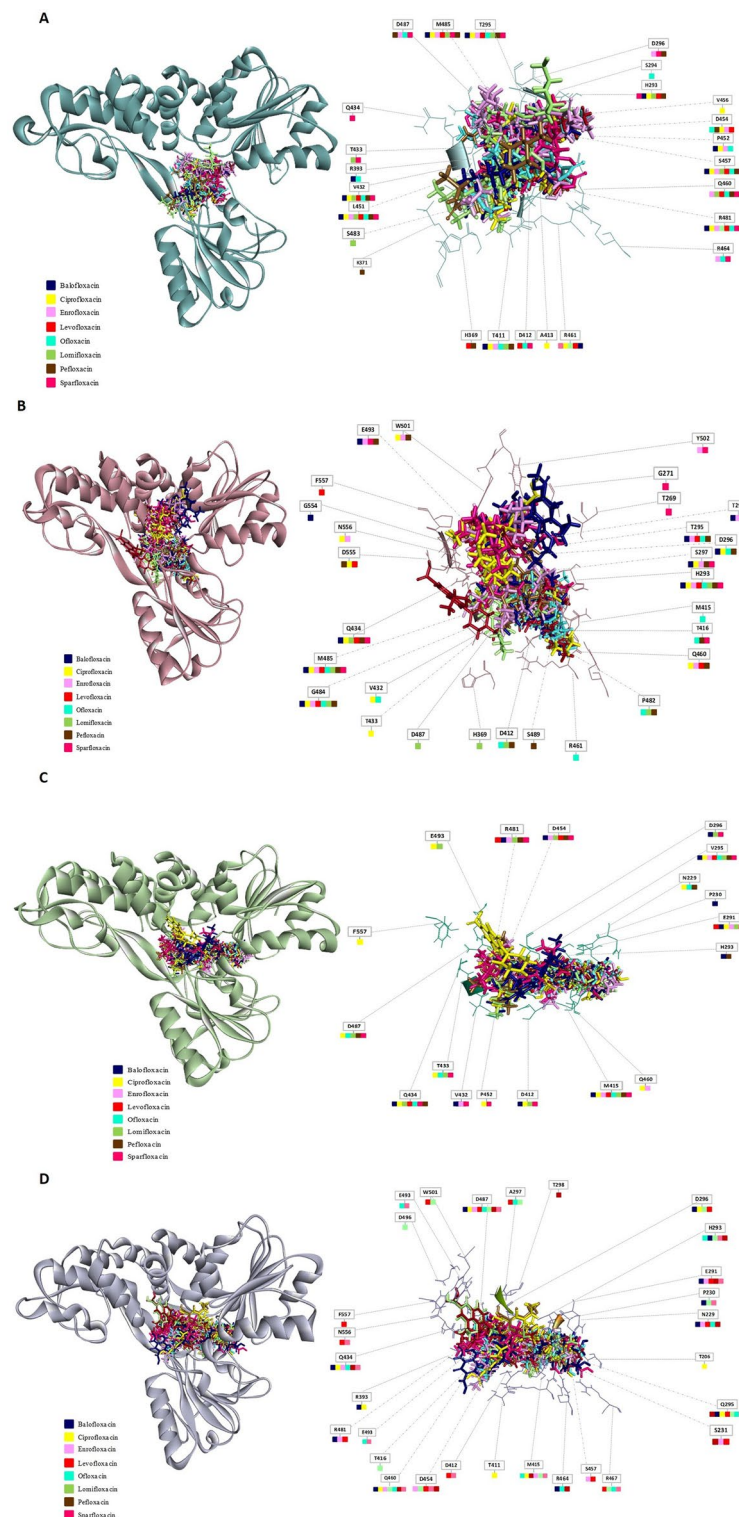
Another study published by Di Maio et al. implemented an in-silico approach to identify NS5B genetic variability in HCV genotypes and its potential effect on the genetic barrier for drug resistance (nucleoside inhibitors (NI) and non-nucleoside inhibitors (NNI))<sup>23</sup>. The impact of these mutations was analyzed via docking of sofosbuvir with the NS5B protein of HCV genotypes. This study suggested that HCV sequence variability in NS5B can potentially alter the efficacy of NS5B inhibitors<sup>23</sup>.

The approach we have used here has the advantage that it analyzes the effect of naturally occurring genotype-specific amino acid variations on drug–protein interactions in the absence of drug-induced selection pressure. This method allowed us to correlate genotype-specific sequence heterogeneity with its influence on protein's structural architecture, and with its effect on variable interactions with selected drugs. This approach can be extended to other viruses to study the clinical significance of genotype-specific sequence variations that influence drug–protein interactions.

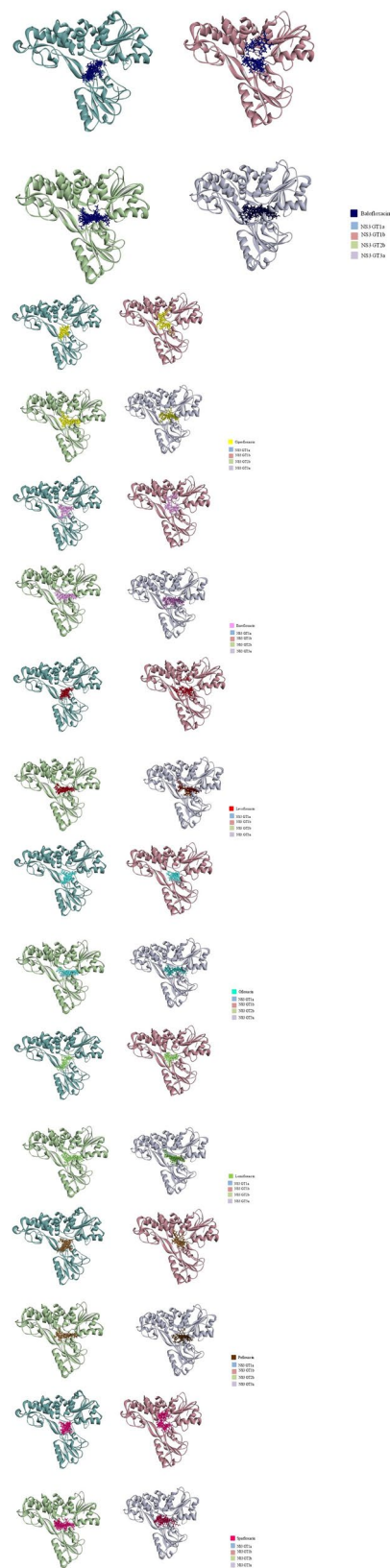
In our study, sequence to structure comparison identified the following genotype-specific substitutions: Y241F in genotype 3a (Motif Y), F418Y in genotype 2b (Motif V), and S459T in genotype 1a (Motif VI), whereas V256N in genotype 3a, S294A in genotype 3a, T295V/Q in genotypes 2b and 3a, A297S in genotype 1b, S299T in genotype 1b and 2b, T430V in genotype 2b and 3a, E555D/L in genotype 2b and 3a, P558A/S in genotype 2b and 3a and Q580V/E in genotype 2b and 3a (Figs. 1 and 2).

In the previous structural–functional studies of NS3 HCV, it has been proposed that the residues that play a critical role in NTP and substrate binding are found in the conserved helicase motifs: Motif I/Walker A (207'-GSGKSTK-213'), Ia (223'-YKVLVLPNSVA-233'), Walker B/Motif II (290'-DECH-293'), III (319'-VLA-TATPP-326'), IV (365'-LIFCHSKKK-373'), V (410'-ATDALMTGYTGDF-422'), and VI (456'-VSRQRRTG R-467') (Fig. 1)<sup>21,24,25</sup>. These helicase signature motifs are oriented in such a manner that their spatial arrangement forms the lining of NTP and substrate binding pocket, located in the cleft between domains I and 2<sup>25,26</sup>. Motif Y (Y241) is a characteristic motif of NS3 HCV located between motifs 1a and 1b. Tyrosine241 plays an important role in stacking the adenine base of ADP in correspondence with T419<sup>26</sup>. In addition to these, amino acids at positions 256V/V/V/N, 294S/S/S/A, 295T/T/V/Q, 297A/S/A/A, 299S/T/T/S, 430T/T/V/V, 555E/D/D/L, 558P/P/A/S and 580Q/Q/V/E exhibited genotype-specific variation. These amino acids form the entry side of ssRNA/DNA<sup>21</sup>.

In light of our observations, we speculate that A240S in genotype 3a, A431C in genotype 1a, T433E in genotype 3a, T435I/Y in genotype 2b and 3a, T445S in genotype 3a, E447T in genotype 2b, T449Q/R in genotype 2b and 3a, A500S in genotype 3a and F557L in genotype 1a might be of functional importance. These amino acids are present in the vicinity of the residues involved in the catalytic center and/or substrate binding loop, and therefore, can influence nucleic acid binding and NS3 catalytic activity. Residue Y241 is crucial for forming interactions with the adenine base of ADP<sup>21,26</sup>. In genotype 3a, we observed substitutions at both positions



**Figure 5.** Molecular docking simulation of HCV NS3 with fluoroquinolones constructed for four HCV genotypes. HCV NS3 protein each from genotype (A) 1a, (B) 1b, (C) 2b, and (D) 3a, was docked with the batch of 8 fluoroquinolones using the Molegro software. In the left panels, a cluster of 8 fluoroquinolones is shown docked into the NS3 binding site, where structures of genotype 1a, 1b, 2b, and 3a are represented with blue, red, green and purple, respectively. The right panels give schematic representations of the fluoroquinolone-interacting amino acids. Binding of fluoroquinolones with each of the amino acids is indicated with color-coded squares where red, purple, light green, dark green, yellow, blue, orange and pink represents Ciprofloxacin, Lomefloxacin, Enrofloxacin, Levofloxacin, Ofloxacin, Pefloxacin, Sparfloxacin, and Balofloxacin, respectively. The grey dashed lines show amino acids located in the back of the fluoroquinolone cluster.



**Figure 6.** Docking poses of HCV genotype-specific NS3 models interacting with different fluoroquinolones: The structures of NS3 from genotype 1a, genotype 1b, genotype 2b, and genotype 3a are shown in blue, purple, green, and grey, respectively. The docked poses of ligands (top 10 poses) are shown in the following order: (1) Ciprofloxacin, (2) Levofloxacin, (3) Ofloxacin, (4) Pefloxacin, (5) Balofloxacin, (6) Enrofloxacin, (7) Sparfloxacin, and (8) Lomefloxacin.



Drugs	1a	1b	2b	3a
Balofloxacin	-	-	-	ASN229
	-	-	PRO230	PRO230
	-	-	GLU291	GLU291
	HIS293	HIS293	HIS293	HIS293
	THR295	THR295	VAL295	GLN295
	-	ASP296	ASP296	ASP296
	-	SER297	-	-
	-	THR298	-	-
	ARG393	-	-	ARG393
	THR411	-	-	-
	-	-	ASP412	-
	-	-	MET415	-
	VAL432	-	VAL432	-
	-	-	-	GLU433
	-	GLN434	GLN434	GLN434
	LEU451	-	-	-
	PRO452	-	-	-
	-	ASP454	-	-
	SER457	-	-	-
	-	-	-	GLN460
	-	-	-	ARG464
	-	-	-	ARG467
	ARG481	ARG481	ARG481	-
	GLY484	GLY484	-	-
	MET485	MET485	-	-
-	-	-	ASP487	
-	GLU493	-	-	
-	GLY554	-	-	
Ciprofloxacin	-	-	-	THR206
	-	-	ASN229	-
	-	-	GLU291	-
	HIS293	HIS293	-	-
	THR295	-	VAL295	GLN295
	ASP296	ASP296	-	ASP296
	-	SER297	-	-
	-	-	-	ARG393
	THR411	-	-	THR411
	-	-	ASP412	-
	ALA413	-	-	-
	-	-	MET415	MET415
	VAL432	VAL432	-	-
	-	THR433	THR433	-
	-	GLN434	GLN434	GLN434
	LEU451	-	-	-
	PRO452	-	PRO452	-
	ASP454	-	-	-
	VAL456	-	-	-
	SER457	-	-	-
	-	GLN460	GLN460	GLN460
	ARG461	-	-	-
	-	-	-	ARG467
	ARG481	-	-	-
	GLY484	GLY484	-	-
MET485	MET485	-	-	
-	-	ASP487	ASP487	
-	-	GLU493	-	
-	TRP501	-	-	
-	ASP555	-	-	
-	ASN556	-	-	
-	-	PHE557	-	
Continued				

Drugs	1a	1b	2b	3a
Enrofloxacin	-	-	-	ASN229
	-	-	-	SER231
	-	-	GLU291	GLU291
	-	HIS293	-	-
	THR295	THR295	VAL295	-
	-	SER297	-	-
	-	THR298	-	-
	THR411	-	-	-
	-	-	MET415	MET415
	-	-	VAL432	-
	-	-	-	GLU433
	-	-	-	GLN434
	LEU451	-	-	-
	PRO452	-	-	-
	ASP454	ASP454	ASP454	ASP454
	SER457	-	-	SER457
	GLN460	GLN460	GLN460	GLN460
	ARG464	-	-	-
	ARG481	ARG481	ARG481	ARG481
	-	GLY484	-	-
MET485	MET485	-	-	
ASP487	-	-	ASP487	
-	GLU493	-	-	
-	TRP501	-	-	
-	TYR502	-	-	
-	ASN556	-	-	
Lomifloxacin	-	-	-	PRO230
	-	-	GLU291	-
	HIS293	HIS293	-	HIS293
	THR295	-	VAL295	GLN295
	-	-	ASP296	ASP296
	-	-	-	ALA297
	-	HIS369	-	-
	THR411	-	-	-
	-	ASP412	ASP412	-
	-	MET415	MET415	-
	-	-	-	THR416
	VAL432	-	-	-
	THR433	-	THR433	-
	-	GLN434	GLN434	-
	LEU451	-	-	-
	-	ASP454	ASP454	ASP454
	SER457	-	-	-
	-	-	-	GLN460
	ARG461	-	-	-
	-	PRO482	-	-
SER483	-	-	-	
GLY484	-	-	-	
MET485	MET485	-	-	
ARG481	ARG481	ARG481	-	
-	ASP487	ASP487	ASP487	
-	-	GLU493	-	
-	-	-	ASP496	
-	-	-	TRP501	
-	-	-	SER558	
Continued				

Drugs	1a	1b	2b	3a
Levofloxacin	-	-	-	ASN229
	-	-	-	SER231
	-	-	GLU291	GLU291
	HIS293	HIS293	-	-
	THR295	THR295	VAL295	GLN295
	-	-	-	ASP296
	-	-	-	ALA297
	HIS369	-	-	-
	ASP412	-	-	ASP412
	-	-	MET415	-
	VAL432	-	-	-
	-	-	THR433	GLU433
	-	GLN434	GLN434	-
	LEU451	-	-	-
	ASP454	-	ASP454	ASP454
	-	ASP555	-	-
	SER457	-	-	SER457
	GLN460	GLN460	-	-
	ARG461	-	-	-
	ARG481	ARG481	ARG481	ARG481
	GLY484	GLY484	-	-
	MET485	MET485	-	-
	-	-	-	ASP487
-	-	-	TRP501	
-	-	-	ASN556	
-	PHE557	-	PHE557	
-	-	-	SER558	
Ofloxacin	-	-	ASN229	ASN229
	-	HIS293	-	HIS293
	SER294	-	-	-
	THR295	THR295	VAL295	GLN295
	-	ASP296	-	-
	-	-	-	ALA297
	THR411	-	-	-
	ASP412	ASP412	-	-
	-	MET415	MET415	MET415
	-	THR416	-	-
	VAL432	VAL432	-	-
	-	-	THR433	-
	-	-	GLN434	GLN434
	LEU451	-	-	-
	PRO452	-	-	-
	ASP454	-	-	-
	SER457	-	-	-
	QLN460	-	-	GLN460
	-	ARG461	-	-
	ARG464	-	-	ARG464
	-	-	-	ARG467
	ARG481	-	-	-
	-	PRO482	-	-
-	GLY484	-	-	
-	MET485	-	-	
ASP487	-	ASP487	ASP487	
-	-	-	GLU493	
Continued				

Drugs	1a	1b	2b	3a
Sparfloxacin	-	-	-	PRO230
	-	THR269	-	-
	-	GLY271	-	-
	-	-	GLU291	GLU291
	HIS293	HIS293	-	HIS293
	THR295	-	VAL295	GLN295
	ASP296	-	ASP296	-
	-	SER297	-	-
	-	THR298	-	-
	ASP412	-	ASP412	ASP412
	-	-	MET415	MET415
	-	THR416	-	-
	VAL432	-	VAL432	-
	THR433	-	THR433	-
	GLN434	GLN434	GLN434	GLN434
	LEU451	-	-	-
	-	-	PRO452	-
	-	-	ASP454	ASP454
	SER457	-	-	-
	GLN460	-	-	GLN460
	ARG464	-	-	-
	ARG467	-	-	ARG467
	ARG481	ARG481	ARG481	ARG481
	-	PRO482	-	-
	GLY484	-	-	-
	MET485	MET485	MET485	-
	ASP487	-	ASP487	ASP487
	-	GLU493	-	GLU493
	-	TYR502	-	-
	-	-	-	ASN556
Pefloxacin	-	-	ASN229	ASN229
	-	-	-	SER231
	-	-	-	GLU291
	HIS293	HIS293	HIS293	HIS293
	THR295	THR295	VAL295	GLN295
	ASP296	ASP296	-	-
	-	SER297	ALA297	-
	-	-	-	THR298
	HIS369	-	-	-
	LYS371	-	-	-
	THR411	-	-	-
	-	ASP412	-	-
	-	-	MET415	MET415
	THR416	-	-	-
	VAL432	-	-	-
	-	GLN434	GLN434	GLN434
	LEU451	-	-	-
	ASP454	-	ASP454	ASP454
	SER457	-	-	-
	GLN460	GLN460	-	GLN460
	-	-	-	ARG464
	ARG481	-	ARG481	-
	-	GLY484	-	-
	MET485	MET485	MET485	-
	ASP487	-	ASP487	ASP487
	-	SER489	-	-
	-	GLU493	-	-
	-	TRP501	-	-
	-	ASP555	-	-

**Table 2.** HCV genotype 1a, 1b, 2b and 3a NS3 amino acids residues interacting with Fluoroquinolones: Molegro Virtual Docker software was employed to dock the panel of 8 fluoroquinolones (Sparfloxacin, balofloxacin, Enrofloxacin, Ofloxacin, Pefloxacin, Lomefloxacin, Levofloxacin, and Ciprofloxacin) on each of the four NS3 helicases from different HCV genotypes. The table shows amino acid residues (bold) in the fluoroquinolone binding region of NS3 helicase from different genotypes that formed interactions with the fluoroquinolones. Absence of a residue in genotype(s) is indicated by ‘-’.



Ligand (top 10 poses)	IC <sub>50</sub> (uM) <sup>12</sup>	Genotype 1a (MolDock score)	Genotype 1b (MolDock score)	Genotype 2b (MolDock score)	Genotype 3a (MolDock score)
Balofloxacin	1.37	-99	-105	-112	-104
Ciprofloxacin	2.99	-104	-115	-108	-110
Enrofloxacin	1.15	-101	-102	-107	-122
Levofloxacin	-	-90	-116	-106	-102
Lomefloxacin	1.14	-96	-91	-102	-99
Ofloxacin	1.47	-88	-97	-107	-100
Pefloxacin	4.22	-103	-87	-102	-110
Sparfloxacin	1.22	-92	-88	-104	-92

**Table 3.** Molegro docking score for NS3-Fluoroquinolones interactions: Molegro docking score for top 10 poses observed for each fluoroquinolone-NS3 (specific to each genotype) interactions are shown. Previously determined in vitro IC<sub>50</sub> (μM) for each fluoroquinolone, except for Levofloxacin (shown as -) is also given.

240 and 241 simultaneously, while the Tyrosine241 was replaced with phenylalanine, alanine at position 240 that was substituted with serine. These substitutions concurrently can affect the binding of ATP molecule and subsequently the helicase activity due to changes in the electrostatic environment of the binding site. In close vicinity of Walker B Motif, lie residues 294 and 295. In this motif, H293 is of prime importance and is implicated in the coupling of ATPase and helicase activity<sup>21</sup>. Amino acid 294 and 295 are not only found in the proximity H293 but are also present at the entry side of the substrate. Therefore, substitutions at these positions, such as S294A and T295V/Q (Figs. 1 and 2), might not only influence entry and binding orientation of ssRNA/DNA but may also affect NTP binding and hydrolysis. Motif V corresponds to the coordination of ATP and substrate binding<sup>21,26,27</sup>. We observed a substitution of amino acid F418Y, adjacent to the functionally important residue 419, involved in NTP binding and catalysis<sup>21</sup>. Here, we speculate little or no effect on the activity because the amino acid change was conservative. Residues 431C/C/A/A, 433T/T/T/E, 435T/T/I/Y, 445T/T/T/S, 447E/E/T/E, 449T/T/Q/R, 500A/A/A/S, 557L/F/F/F may play either a direct or indirect role in ssDNA/RNA interaction and unwinding. This leads us to speculate that genotype-specific variations in these amino acids might influence substrate interaction and helicase activity.

A characteristic structural feature of NS3 HCV is a β-hairpin spanning the amino acids 430–452. This hairpin consists of two conserved phenylalanine residues F438 and F444. Residues V432 and T450 in the hairpin are known to be involved in nucleic acid interactions<sup>28</sup>. V432 intercalates between the DNA bases, stacks the nucleotide bases at 5' terminal, and locks the enzyme in position, whereas T450 is involved in ssDNA binding to the nucleic acid pocket of the NS3 helicase<sup>28</sup>. F438 and F444 are responsible for the release ssDNA upon ATP binding<sup>28</sup>. It may be suggested that substitutions in the vicinity of these two residues, at positions 430T/T/V/V, 431C/C/A/A, 433T/T/T/E, 435T/T/I/Y, 445T/T/T/S, 447E/E/T/E, 449T/T/Q/R and 451L/V/V/A can alter the flexibility of the loop and subsequently might affect substrate interaction. Q460, a conserved residue of helicase Motif VI, plays a key role in ATP binding and coupling ATP hydrolysis with helicase translocation<sup>26</sup>. W501 is a critical residue for nucleic acid binding<sup>26</sup>.

To gain insights in drug–protein binding pattern across the HCV genotypes; molecular docking simulations were performed with NS3 protein models of each candidate genotype against a batch of 8 fluoroquinolones (Sparfloxacin, Ciprofloxacin, Balofloxacin, Levofloxacin, Lomefloxacin, Enrofloxacin, Pefloxacin, and Ofloxacin) previously shown to exhibit inhibitory potential against HCV NS3 helicase under in vitro conditions<sup>12</sup>. In our analysis, it was observed that the fluoroquinolones interacted with amino acids that constitute the helicase catalytic core via H-bond as well as non-bond interactions (Supplementary Table 2). Several studies have demonstrated that both H-bonds and hydrophobic interactions stabilize the ligands at the target site and assist in altering binding affinity and drug efficacy<sup>29</sup>. Furthermore, biological activity of the drug increased with increase in the number of hydrophobic interactions in the core of drug–target interface<sup>29</sup>. The residues that are functionally important and were found involved in H-bond as well as non-bond interactions with our drugs are P230, S231, T269, K371, R393, W501 and Y502<sup>25,26</sup> (Figs. 5 and 6; Table 2; Supplementary Table 2). Residues S231, T269, W501, and Y502 are functionally important as they are known to facilitate the interactions of NS3 with a 3' segment of substrate/ssDNA. P230 and S231 are key residues involved in DNA binding. S231 interacts with substrate/ssDNA via bridging water molecule, while T269 is thought to be indispensable for helicase activity of NS3. K371 and R393, crucial residues of NS3 active site, are involved in substrate binding, where R393 makes interactions with the 5' segment of ssDNA and coordinates in translocation of NS3 helicase<sup>25,26</sup>; in addition to this, K371 participate in stabilizing the interaction with ssDNA/RNA<sup>30</sup>. W501 stacks the nucleotide bases near the 3' end of the substrate/ssDNA and acts as a bookend along with V432 that defines a central binding cavity<sup>26</sup>.

Fluoroquinolones, Sparfloxacin (in genotype 3a), Balofloxacin (in genotype 2b and 3a), and Lomefloxacin (in genotype 3a) targeted residue P230 (Fig. 5 and Table 2) that might hinder in substrate binding. Enrofloxacin, Levofloxacin, and Pefloxacin (in genotype 3a) were found to interact with residue S231. Interactions of the Enrofloxacin, Levofloxacin, and Pefloxacin with these residues might affect ATP binding and transition as well as enzyme–substrate interaction. Likewise, Sparfloxacin (in genotype 1b) targeted T269 (Table 2) that can potentially inhibit substrate binding and subsequently NS3 unwinding activity. R393 is thought to be a key residue that makes direct contact with ssDNA/RNA as well as facilitates nucleic acid unwinding<sup>28</sup>. Therefore, drugs such as

Balofloxacin (in genotype 1a and 3a), Ciprofloxacin (in genotype 3a) targeting residue R393 (Fig. 5 and Table 2) can greatly affect substrate binding, NS3 translocation and subsequently nucleic acid unwinding.

W501 is a critical amino acid of substrate binding cleft of NS3 HCV that serves as a bookend residue; stacking nucleotide of ssDNA/RNA at 3' terminal preventing the protein from sliding along the nucleic acid and facilitates helicase translocation<sup>26,28</sup>. Ciprofloxacin, Pefloxacin and Enrofloxacin (in genotype 1b) and Levofloxacin and Lomefloxacin (in genotype 3a) interacted with W501, while most fluoroquinolones with interacted residues in their proximity of W501, namely E493 (present in genotypes 1b, 2b, and 3a) (Fig. 5 and Table 2), which might cause hindrance in nucleic acid interaction and can potentially inhibit the helicase activity of NS3. Maga et al. identified a series of compounds out of which QU663 exhibited a strong binding affinity with NS3. QU663 binds to the RNA binding cleft and makes direct contact with residues R393 and W501<sup>31</sup> that are crucial for substrate binding and helicase translocation. The present study showed that fluoroquinolones we tested also targeted the same crucial residues in the substrate-binding catalytic cleft, namely, W501 and R393 (Fig. 5).

We identify a few limitations of our study. A major limitation of this study was that our analysis was performed entirely *in silico*. We used two crystal structures (for 1a and 1b) and used a homology modeling approach to construct NS3 structures for genotypes 2b and 3a. Homology modeling heavily relies on the identification of the correct template, selected based on sequence identity<sup>32</sup>. A sequence identity above 50% generally tends to generate reliable structures with limited errors in loops and side-chain positing, while structures below 30% can have serious folding errors<sup>32</sup>. Additionally, regions in the query sequence that share low sequence similarity with the template protein, even though the rest of the protein show a high sequence similarity, can lead to erroneous folding. We tried reducing this bias/error by using three different homology modeling programs, validating and checking constructed structures for various errors and using a template that exhibited excellent similarity with our query sequences. Interestingly, all three programs identified 1A1V as a suitable template sharing more than 90% similarity with the sequence, and all three programs gave the same structures. Additionally, our structures passed the 3D verification (performed using Verify 3D software) as at least 80% of the amino acids have scored  $> 0.2$  in the 3D/1D profile. Similarly, the structures were valid on the Ramachandran plot as most of the amino acids were under the permissible regions. The other limitation we anticipate is with the molecular docking approach we used. In absence of experimentally characterized binding sites and experimentally derived binding energies, molecular docking analyses can be unreliable and can potentially miss out several key residues, present artificial interaction with residues and/or under- or over-estimate the binding scores<sup>33</sup>. We tried to address this issue by validating our docking approach using two previously reported structures bound to the ligand. We adopted a 'blind docking' approach, where the binding site of the ligand was not defined. Our software predicted the same pose for the ligand and also identified all the residues previously reported for the two structures. Another limitation that we identify is use of docking score to rank the poses and selection the top pose with highest docking score. Docking programs produce one (or several) different poses for every ligand, and further rank different compounds based on their scoring functions<sup>34</sup>. Several studies suggest that binding energies/scores predicted by the docking might be incorrect, despite the correctly predicted binding pose<sup>34</sup>. These can be overcome by using more robust approaches such as MD simulations followed by WaterMap analysis; however, these approaches can be time-consuming and computationally demanding<sup>34</sup>. Nonetheless, as the experimentally solved structures continue to grow, the boundaries that differentiate between reliable and unreliable predictions will narrow and the capacity of the docking tools to predict correct poses/interactions will increase<sup>32</sup>. However, to gain further insights into the mechanism, site-directed mutagenesis of crucial residues followed by functional assays may be employed. These analyses should be supplemented by observations from the crystal structures of drug–enzyme complexes.

In conclusion, our approach provides an in-depth analysis of sequence variation in HCV NS3 protein across the selected genotypes and identifies active site amino acid residues crucial for drug–protein interactions. This approach can help study inter-genotype sequence heterogeneity in viral proteins and its correlation with the response to the antiviral treatment regimen.

## Methods

**Retrieval of HCV NS3 sequences and sequence analysis.** A total of 1703 HCV NS3 sequences, belonging to genotype 1a, 1b, 2b, and 3a were downloaded in FASTA format from Los Alamos HCV Database (<https://hcv.lanl.gov/content/index>)<sup>35</sup>. Other genotypes were not considered because their sequences were not present in the database in sufficient numbers to conduct a meaningful analysis. Out of these sequences, 687, 667, 101 and 248, belonged to, respectively, genotype 1a, 1b, 2b, and 3a. Sequences for each genotype were aligned separately and edited using the MEGA 6.0 software, implementing the ClustalW algorithm. The aligned nucleotide sequences were then used to generate consensus sequences for each of the candidate genotypes using the advanced 'Consensus maker' tool with 0.7 thresholds (i.e. residues represented in  $\geq 70\%$  of the sequences were used to generate a consensus) (<http://www.hiv.lanl.gov/content/sequence/CONSENSUS/consensus.html>). In all genotypes, almost all consensus residues/sites were present in 99% of the sequences, with exception of a few sites, where consensus residues were present in 70–85% sequences (Supplementary Table 1). The consensus sequences of NS3 for genotypes 1a, 1b, 2b, and 3a were subsequently translated to amino acid sequences using the ExPASy translate tool<sup>36</sup>. Additionally, reference sequences for genotypes 1a, 1b, 2b, and 3a were downloaded in FASTA format PDB (genotype 1a: PDBid:1A1V and genotype 1b: PDBid:1CU1) and UniProt (genotype 2b: UniProt ID Q9DHD6 and genotype 3a: UniProt ID Q81495) databases. For genotype variation analysis, the genotype sequences were aligned with each other using the ClustalW algorithm implemented in MEGA 7 software<sup>37</sup>. Additionally, each genotype consensus sequence was also aligned, using MEGA 7, with its reference sequence to compare the difference in consensus sequences and the reference sequences.

Inter-subtype sequence similarities and differences were determined by constructing a sequence identity matrix using 'Bioedit' software<sup>38</sup>. Two ID matrices were constructed: one with a full-length sequence of HCV NS3 helicase comprising 623 amino acids, and the other comprising of amino acids spanning Pro230-Cys584 (termed in this study as fluoroquinolone binding region). This region was selected because fluoroquinolones are found to form interactions with different amino acids within this region<sup>39</sup>. Additionally, this region also contains the previously reported helicase active site, spanning from Arginine393 to Tryptophan501<sup>8,40,41</sup>.

**Retrieval of HCV helicase structures and protein homology modeling.** PDB and UniProt databases were searched for the available genotype-specific NS3 helicase structures and crystal structures were only found for genotypes 1a and 1b. These structures (genotype 1a: PDBid:1A1V and genotype 1b: PDBid:1CU1) were downloaded on .pdb format. For genotypes 2b and 3a, the Homology Modeling approach was adopted for the generation of 3D protein structures<sup>10,42</sup>. For accuracy and reliability, 3D models of NS3 were generated using three different homology modeling tools, namely the CPH model, the Swiss model and Phyre 2<sup>43–45</sup>. The structures were visually inspected in Discovery Studio Visualizer version 4.0 (DSV4.0; Dassault Systèmes BIOVIA, Discovery Studio Visualizer, version 4.0, San Diego: Dassault Systèmes, 2016; Retrieved from <http://accelrys.com/products/discovery-studio/>), and thereafter saved in PDB format. Subsequently, the structures were verified using the Verify 3D tool, while the energy minimization and validation were performed using the GROMACS tool<sup>46</sup> and Ramachandran plot analysis implemented in DSV4.0.

**Fluoroquinolone structures.** In this study, we used a panel of 8 fluoroquinolones, namely Sparfloxacin, Ofloxacin, Balofloxacin, Pefloxacin, Levofloxacin, Lomefloxacin, Ciprofloxacin, and Enrofloxacin, that have previously been reported to effectively inhibit HCV NS3 helicase<sup>12</sup>. Structures of these drugs were retrieved from PubChem Database in 3D SDF format<sup>47</sup>. Before docking analysis, SDF structures were converted to PDB format using DSV4.0.

**Analysis of genotype-specific fluoroquinolone-NS3 interactions.** Molecular docking simulations were performed to analyze the genotype-specific protein–ligand interactions. The docking approach was validated by performing blind docking (assuming drug binding site to be anywhere on the protein) on two previously reported complexes of NS3 bound to inhibitors, one a natural analog M2<sup>48</sup> and other inhibitor ITMN-3479 (<http://www.rcsb.org/pdb/explore/litView.do?structureId=3RVB>). Molecular docking studies and conformational analysis were conducted by Molegro Virtual Docker<sup>49</sup> and DSV4.0 software. Molegro Virtual Docker binding site was set to default, where the Molegro identifies the binding ligand binding sites in the protein. For our analysis, Molegro identified two binding sites, one located in the DNA binding domain, which contained the key residues within the active site of NS3, as identified by previous reports<sup>8,40,41</sup>, and others in the ATP domain of the helicase. The 8 fluoroquinolones were docked individually on each of the four genotype-specific models of NS3, into the identified cartesian space. Molecular docking was performed using standard precision protocols with default parameters of Molegro Virtual Docker. A total of 2000 interactions were simulated, and out of those top 10 poses was selected based on docking energies. Using these poses, analysis of genotype-specific variations on NS3–fluoroquinolone binding was carried out. Visualization of docking poses and analysis of drug–protein interactions were performed using DSV4.0.

Received: 17 October 2019; Accepted: 18 September 2020

Published online: 30 November 2020

## References

- Rubio, L., Guerri, J. & Moreno, P. Genetic variability and evolutionary dynamics of viruses of the family Closteroviridae. *Front. Microbiol.* **4**, 151. <https://doi.org/10.3389/fmicb.2013.00151> (2013).
- Grigoras, I. *et al.* High variability and rapid evolution of a nanovirus. *J. Virol.* **84**, 9105–9117. <https://doi.org/10.1128/JVI.00607-10> (2010).
- Irwin, K. K., Renzette, N., Kowalik, T. F. & Jensen, J. D. Antiviral drug resistance as an adaptive process. *Virus Evol.* <https://doi.org/10.1093/ve/vew014> (2016).
- Lauring, A. S. & Andino, R. Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog.* **6**, e1001005. <https://doi.org/10.1371/journal.ppat.1001005> (2010).
- Cavalcante, L. N. & Lyra, A. C. Predictive factors associated with hepatitis C antiviral therapy response. *World J. Hepatol.* **7**, 1617–1631. <https://doi.org/10.4254/wjh.v7.i12.1617> (2015).
- Zein, N. N. Clinical significance of hepatitis C virus genotypes. *Clin. Microbiol. Rev.* **13**, 223–235 (2000).
- von Delft, A. *et al.* The broad assessment of hepatitis C virus genotypes 1 and 3 antigenic targets reveals limited cross-reactivity with implications for vaccine design. *Gut* **65**, 112–123. <https://doi.org/10.1136/gutjnl-2014-308724> (2016).
- Frick, D. N. The hepatitis C virus NS3 protein: A model RNA helicase and potential drug target. *Curr. Issues Mol. Biol.* **9**, 1–20 (2007).
- Raney, K. D., Sharma, S. D., Moustafa, I. M. & Cameron, C. E. Hepatitis C virus non-structural protein 3 (HCV NS3): A multi-functional antiviral target. *J. Biol. Chem.* **285**, 22725–22731. <https://doi.org/10.1074/jbc.R110.125294> (2010).
- Kim, J. L. *et al.* Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: The crystal structure provides insights into the mode of unwinding. *Structure* **6**, 89–100 (1998).
- Aldred, K. J., Kerns, R. J. & Osheeroff, N. Mechanism of quinolone action and resistance. *Biochemistry* **53**, 1565–1574. <https://doi.org/10.1021/bi5000564> (2014).
- Khan, I. A., Siddiqui, S., Rehmani, S., Kazmi, S. U. & Ali, S. H. Fluoroquinolones inhibit HCV by targeting its helicase. *Antiviral Ther.* **17**, 467–476. <https://doi.org/10.3851/IMP1937> (2012).

13. Ikeda, S., Yazawa, M. & Nishimura, C. Antiviral activity and inhibition of topoisomerase by ofloxacin, a new quinolone derivative. *Antiviral Res.* **8**, 103–113 (1987).
14. Portolani, M. *et al.* Suppression of BK virus replication and cytopathic effect by inhibitors of prokaryotic DNA gyrase. *Antiviral Res.* **9**, 205–218 (1988).
15. Witvrouw, M. *et al.* Broad-spectrum antiviral activity and mechanism of antiviral action of the fluoroquinolone derivative K-12. *Antiviral Chem. Chemother.* **9**, 403–411. <https://doi.org/10.1177/095632029800900504> (1998).
16. Cao, D., Ni, Y. Y. & Meng, X. J. Substitution of amino acid residue V1213 in the helicase domain of the genotype 3 hepatitis E virus reduces virus replication. *Virology*. **15**, 32. <https://doi.org/10.1186/s12985-018-0943-5> (2018).
17. Norouzi, M. *et al.* Identification of Hepatitis B Virus Surface Antigen (HBsAg) Genotypes and Variations in Chronic Carriers from Isfahan Province, Iran. *J. Public Health* **41**, 104–111 (2012).
18. Scheel, T. K. *et al.* Analysis of functional differences between hepatitis C virus NS5A of genotypes 1–7 in infectious cell culture systems. *PLoS Pathog.* **8**, e1002696. <https://doi.org/10.1371/journal.ppat.1002696> (2012).
19. Cento, V. *et al.* HCV genotypes are differently prone to the development of resistance to linear and macrocyclic protease inhibitors. *PLoS ONE* **7**, e39652. <https://doi.org/10.1371/journal.pone.0039652> (2012).
20. Imamura, T. *et al.* Distribution of hepatitis B viral genotypes and mutations in the core promoter and precore regions in acute forms of liver disease in patients from Chiba, Japan. *Gut* **52**, 1630–1637 (2003).
21. Fatima, K., Azhar, E., Mathew, S., Damanhour, G. & Qadri, I. Comparison of structural architecture of HCV NS3 genotype 1 versus Pakistani genotype 3a. *Biomed. Res. Int.* **2014**, 749254. <https://doi.org/10.1155/2014/749254> (2014).
22. El-Shamy, A. *et al.* Sequence heterogeneity of NS5A and core proteins of hepatitis C virus and virological responses to pegylated-interferon/ribavirin combination therapy. *Microbiol. Immunol.* **55**, 418–426. <https://doi.org/10.1111/j.1348-0421.2011.00331.x> (2011).
23. Di Maio, V. C. *et al.* Hepatitis C virus genetic variability and the presence of NS5B resistance-associated mutations as natural polymorphisms in selected genotypes could affect the response to NS5B inhibitors. *Antimicrob. Agents Chemother.* **58**, 2781–2797. <https://doi.org/10.1128/AAC.02386-13> (2014).
24. Tai, C. L. *et al.* Structure-based mutational analysis of the hepatitis C virus NS3 helicase. *J. Virol.* **75**, 8289–8297 (2001).
25. Zheng, W. & Tekpinar, M. Structure-based simulations of the translocation mechanism of the hepatitis C virus NS3 helicase along single-stranded nucleic acid. *Biophys. J.* **103**, 1343–1353. <https://doi.org/10.1016/j.bpj.2012.08.026> (2012).
26. Gu, M. & Rice, C. M. Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism. *Proc. Natl. Acad. Sci. USA* **107**, 521–528. <https://doi.org/10.1073/pnas.0913380107> (2010).
27. Linder, P. Dead-box proteins: a family affair—active and passive players in RNP-remodeling. *Nucleic Acids Res.* **34**, 4168–4180. <https://doi.org/10.1093/nar/gkl468> (2006).
28. Frick, D. N. In *Hepatitis C Viruses: Genomes and Molecular Biology* (ed S. L. Tan) (2006).
29. Patil, R. *et al.* Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathways of drug-designing. *PLoS ONE* **5**, e12029. <https://doi.org/10.1371/journal.pone.0012029> (2010).
30. Lin, C. & Kim, J. L. Structure-based mutagenesis study of hepatitis C virus NS3 helicase. *J. Virol.* **73**, 8798–8807 (1999).
31. Belon, C. A. & Frick, D. N. Helicase inhibitors as specifically targeted antiviral therapy for hepatitis C. *Fut. Virol.* **4**, 277–293. <https://doi.org/10.2217/fvl.09.7> (2009).
32. Rodrigues, J. P. *et al.* Defining the limits of homology modeling in information-driven protein docking. *Proteins* **81**, 2119–2128. <https://doi.org/10.1002/prot.24382> (2013).
33. Huang, S. Y. & Zou, X. Advances and challenges in protein-ligand docking. *Int. J. Mol. Sci.* **11**, 3016–3034. <https://doi.org/10.3390/ijms11083016> (2010).
34. Pansar, T. & Poso, A. Binding affinity via docking: Fact and fiction. *Molecules* **23**, 1899. <https://doi.org/10.3390/molecules23081899> (2018).
35. Kuiken, C., Yusim, K., Boykin, L. & Richardson, R. The Los Alamos hepatitis C sequence database. *Bioinformatics (Oxford, England)* **21**, 379–384. <https://doi.org/10.1093/bioinformatics/bth485> (2005).
36. Artimo, P. *et al.* ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* **40**, W597–W603. <https://doi.org/10.1093/nar/gks400> (2012).
37. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874. <https://doi.org/10.1093/molbev/msw054> (2016).
38. Hall, T. A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98 (1999).
39. Anwar, M. F. *et al.* Integrated cheminformatics-molecular docking approach to drug discovery against viruses. *Infect. Disord. Drug Targets* <https://doi.org/10.2174/1871526518666181019162359> (2018).
40. Shah, A. A., Siddiqui, S., & Ali, S. H. HCV helicase as a therapeutic target. In *Virology II: Advanced Issues* (2014)
41. Kim, D. W., Kim, J., Gwack, Y., Han, J. H. & Choe, J. Mutational analysis of the hepatitis C virus RNA helicase. *J. Virol.* **71**, 9400–9409 (1997).
42. Berman, H. M. *et al.* The protein data bank. *Nucleic Acids Res.* **28**, 235–242. <https://doi.org/10.1093/nar/28.1.235> (2000).
43. Nielsen, M., Lundegaard, C., Lund, O. & Petersen, T. N. CPH Models-30-remote homology modeling using structure-guided sequence profiles. *Nucleic Acids Res.* **38**, W576–W581. <https://doi.org/10.1093/nar/gkq535> (2010).
44. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protocols* **10**, 845–858. <https://doi.org/10.1038/nprot.2015.053> (2015).
45. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* **31**, 3381–3385. <https://doi.org/10.1093/nar/gkg520> (2003).
46. Van Der Spoel, D. *et al.* GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718. <https://doi.org/10.1002/jcc.20291> (2005).
47. Bolton, E. E., Wang, Y., Thiessen, P. A. & Bryant, S. H. Integrated platform of small molecules and biological activities. In *Annual Reports in Computational Chemistry, vol 4* (eds Ralph, A. W. & David, C. S.) 217–241 (Elsevier, Amsterdam, 2008).
48. Anwar, M. F. *et al.* Application of an integrated cheminformatics-molecular docking approach for discovery for physicochemically similar analogs of fluoroquinolones as putative HCV inhibitors. *Comput. Biol. Chem.* **84**, 107167. <https://doi.org/10.1016/j.compbiolchem.2019.107167> (2020).
49. Thomsen, R. & Christensen, M. H. MolDock: A new technique for high-accuracy molecular docking. *J. Med. Chem.* **49**, 3315–3321. <https://doi.org/10.1021/jm051197e> (2006).

## Acknowledgements

SA received funding from Nazarbayev University under the grant:110119FD4516, and by the National Institute On Drug Abuse of the National Institutes of Health under Award Number R03DA052179.



### Author contributions

R.K.: Performed most analyses and wrote initial draft. M.F.A.: Performed data analyses. M.A.R.: Performed sequence analyses. S.N.: Helped with molecular docking analysis. S.H.A.: Supervised the project, finalized the analyses and finalized paper writing. S.A.: Supervised the project, critically analyzed the manuscript and finalized paper writing.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-77720-9>.

**Correspondence** and requests for materials should be addressed to S.H.A. or S.A.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020