



Binding of GS-461203 and its Halogen Derivatives to HCV Genotype 2a RNA Polymerase Drug Resistance Mutants

Muhammad Arba^{1*}, Setyanto Tri Wahyudi², Muhammad Sulaiman Zubair³, Dylan Brunt⁴, Mursalin Singh⁴ and Chun Wu^{4*}

¹ Department of Pharmacy, Faculty of Pharmacy, Universitas Halu Oleo, Kendari 93232, Indonesia; muh.arba@uho.ac.id

² Department of Physics, IPB University, Bogor 16680, Indonesia

³ Department of Pharmacy, Faculty of Science, Tadulako University, Palu 94118, Indonesia

⁴ Department of Molecular & Cellular Biosciences, College of Science and Mathematics, Rowan University, Glassboro, New Jersey 08028, United States; wuc@rowan.edu

* Correspondence: muh.arba@uho.ac.id (MA); wuc@rowan.edu (CW)

Abstract: Hepatitis C Virus (HCV) is reported to develop GS-461203 resistance because of multiple mutations within the RNA-dependent RNA Polymerase (RdRp) of HCV. The lack of a high-resolution structure of these RdRp mutants in complex with GS-461203 hinders efforts to understand the drug resistance. Here we decipher the binding differences of GS-461203 in the wild type and mutated systems T179A or M289L of HCV RdRp Genotype 2a using homology modeling, molecular docking, and molecular dynamics simulation. Key residues responsible for GS-461203 binding were identified to be Arg48, Arg158, Asp318, Asp319, and Asp220, and that mutations T179A or M289L have caused conformational changes of GS-461203 in the RdRp active site. The affinities of GS-461203 were reduced in T179A system, but it became slightly stronger in the M289L system. Furthermore, we designed two new analogues of GS-461203 which encouragingly induced more stable interactions than GS-461203, and thus resulted in much better binding energies. This present study reveals how a single mutation, T179A or M289L, will modulate GS-461203 binding in HCV RdRp Genotype 2a, while introducing two novel analogues to overcome the drug resistance which may be good candidate for further experimental verification.

Keywords: hepatitis C virus; drug resistance; RdRp; NS5B; MD simulation; drug mutation

Citation: Arba, M.; Wahyudi, S.T.W.; Wu, C. Title. *Sci. Pharm.* **2022**, *90*, x. <https://doi.org/10.3390/xxxxx>

Academic Editor: Firstname Lastname

Received: date

Accepted: date

Published: date

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Hepatitis C virus (HCV) has caused a pandemic which currently infects approximately 180 million people worldwide. HCV has a single-stranded, positive sense RNA genome belonging to genus *Hepacivirus* of *Flaviviridae* family [1]. Based on their viral genome sequence, HCV is divided into eight distinct genotypes (GT1-GT8), differing by a maximum nucleotide sequence of 33% between GTs, and tens of subtypes (e.g. 2a-2r) [2-5], differing by less than 15% of the nucleotide sequence [5-8]. The GT difference has a direct impact on the antiviral agent chosen for therapy, which highlights the importance of investigating a drug binding mode to a specific GT of HCV [9]. In addition, geographic distribution of GT varies. For instance, subtype 1a is common in the U.S. and in Europe, while subtype 1b is found worldwide [7, 10]. The GT2 particularly accounts for 11.0% infections worldwide, which is predominantly found in African and Asian countries, each with 23.7% and 18.6% prevalence, respectively [2, 7].

Targeted antiviral agents have dramatically improved the treatment of HCV patients, which includes the drug Sofosbuvir that targets the RNA-dependent RNA polymerase (RdRp) of HCV [11, 12]. The RdRp constitutes the catalytic core of the HCV replicase system without any counterpart in mammals, and thus is a prime target for antiviral drug development [13-15]. Sofosbuvir is a prodrug which is metabolized into its active form (GS-461203), and is the only approved drug of nucleotide inhibitors which replaced interferon (IFN) for the treatment of HCV [16]. Sofosbuvir is clinically used in all GTs due to the highly conserved RdRp among all GTs and its high barrier of resistance is well known. However, its use in HCV GT2 was reported to cause resistance associated variants (RAV) in the RdRp which impaired GS-461203 activity. The prevalence of RAV in GT2 was reported to be 4.0% [17] with point mutations including T179A or M289L, in addition to S282T, which was common in GS-461203 use in all GTs [17-21]. The prevalence of RAV for T179A and M289L in GT2 were 4.89% and 2.72%, respectively, while the variant of S282T was 0.28% in GT1b [17]. The S282T or M289L substitutions were observed to cause ~3-fold increase EC₅₀s, while the T179A effect was not significant [18].

Studying drug resistance has been hindered by the unavailability of a high-resolution structure of these RdRp mutant-GS-461203 complexes. Molecular dynamics simulations (MDS) is well appreciated for its ability to provide the binding dynamics, structural and energetic properties at a molecular level within protein-ligand interactions [22]. A previous study revealed the binding and unbinding mechanism of GS-461203 and substrate UTP to HCV RdRp GT2a wild type (WT) and S282T mutation using 200 ns conventional and steering MDS [23]. Here we extended MDS to two more important mutants T179A and M289L in GT2a. More important, novel analogues were designed to overcome the drug resistance caused by these mutations.

Here we built the homology models using the PDB structure (PDB ID 4WTG) of Sofosbuvir diphosphate in complex with RdRp of the Hepatitis C Viral GT2a with S15G, E86Q, E87Q, C223H and V312I mutations [24], docked GS-461203 to the homology models, and performed MDS for 2x500 ns to evaluate the effects of the single mutations T179 or M289 on GS-461203 binding. The GS-461203 binding behaviours in both WT and mutated HCV RdRp was discussed. Based on the insights, we designed two new GS-461203 analogues and analyzed their binding stabilities to the T179A system using 500 ns MDS. Encouragingly, the two analogues had stronger affinities than GS-461203 toward T179A system, which could be good candidates for further experimental studies.

2. Materials and Methods

The PDB structure of Sofosbuvir diphosphate in complex with RNA-dependent RNA polymerase of the HCV GT2a was downloaded from RCSB Protein Data Bank with PDB ID 4WTG. As the structure contained a series of mutated residues (S15G, E86Q, E87Q, C223H and V312I), all the mutations were converted back to the WT sequence by homology modeling. The structure was then prepared using the Maestro's Protein Preparation Wizard [25] with default parameters, while structures of T179A or M289L

were each then prepared by introducing the point mutation to the WT system. We also prepared system for S282T mutant as this substitution is common in all GTs.

Structure of Sofosbuvir diphosphate was download from the RCSB Protein Data Bank via PDB ID 4WTG [24]. GS-461203 was prepared using Maestro including the addition of one phosphate group and generating ionization states at pH=7 by using Epik's pKa calculations [25]. Ligands were then relaxed through optimization and minimization. GS-461203 was then docked with XP precision into the WT, T179A, S282T, and M289L receptor protein structures.

2.1 Molecular Dynamics Simulation

Four separate systems of the protein structure in complex with GS-461203 including WT, T179A, M289L, or S282T were prepared for two independent MDS runs (500 ns of each). Systems were solvated in a simple point-charge (SPC) [26] orthorhombic water box with a 10 Å water buffer and the addition of Na⁺ and Cl⁻ ions in order to reach 0.15 M NaCl. The OPLS_2005 force field [27] was used to model the protein using Desmond System Builder with Maestro's 2019-2 update on Linux operating system.

The MDS were performed with Schrodinger Maestro's Desmond simulation package [28, 29]. Relaxation and energy minimization were performed using the default protocol to reduce possible steric stress as explained in our previous paper [30].

2.2 Analysis of MD Simulation

Analysis of MDS trajectory was performed using the Desmond simulation interaction diagram (SID) analysis tool which reports protein-ligand root mean squared deviation (RMSD), protein-ligand contacts, protein root-mean squared fluctuation (RMSF), changes in secondary structure elements (SSE) during the simulation, and ligand torsion profiles. The protein and ligand RMSD plots were analyzed to ensure the convergence of the MDS. In addition, cluster analysis was performed using the Desmond trajectory clustering analysis tool [31]. The backbone RMSD matrix is used as the basis of structural similarity and the clustering with average linkage was cut off at 2.5 Å. The centroid structure of the protein-ligand complex was used to represent each structural family. Structural families with frames >1% of the total frames were considered separate structural families with separate centroid structures. To evaluate the effect of T179A and M289L mutations on whole protein structure, TM-align web-server was used, in which TM-score 0.0 to 0.3 and RMSD ≥ 5 Å were employed to assess the structural similarity between wild type and mutant systems [32].

2.3 Binding Energy Calculations

Binding energy calculation was performed using Molecular Mechanism-Generalized Born Surface Area (MM-GBSA) method employing the MM-GBSA Prime module of Schrodinger Maestro's Desmond simulation package. The validity of MM-GBSA method has been previously reported [30]. The snapshots of the last 50 ns of each three trajectories of the simulation were used for the calculation. The OPLS3e force field, a VSGB 2.0 solvation model, and the default Prime protocol was used [33]. The MM-GBSA_Prime protocol includes receptor, ligand, and receptor-ligand complex minimizations. The final binding

energy was calculated using equation: $\Delta E_{\text{bind}} = E_{\text{complex}} - (E_{\text{ligand}} + E_{\text{receptor}})$, in which the binding energy was broken down into three components: $E_{\text{electrostatic}}$, E_{vdW} , and $E_{\text{lipophilic}}$. $E_{\text{electrostatic}}$ was calculated by summing $E_{\text{H-bond}}$ and $E_{\text{coulombic}}$. E_{vdW} summed E_{vdW} , $E_{\text{pi-pi stacking}}$, and $E_{\text{self-contact}}$.

2.4 ADME Prediction

Prediction of ADME properties for compounds was performed using SwissADME web server [34].

3. Results

3.1 Ligand Docking

First, we performed molecular docking of the prepared GS-461203 into the RdRp receptor complex in order to evaluate the ability of the Schrodinger Maestro's Extra Precision (XP) Glide Docking methods to reproduce the experimental structure [35, 36]. The docked pose of GS-461203 and crystal structure of Sofosbuvir diphosphate was similar (Figure S1), which indicated that the dock2ing protocol was valid. Subsequently, GS-461203 was docked into T179A or M289L systems. The docked conformation of GS-461203 in WT and mutant systems are similar with little difference seen particularly in oxygen and phosphate atoms. Binding energies predicted by XP-docking were -17.673 kcal/mol, -18.549 kcal/mol, and -17.447 kcal/mol, for the WT, T179A, and M289L, respectively. Figure 1 depicts the docked conformations of GS-461203 in WT and mutants systems.

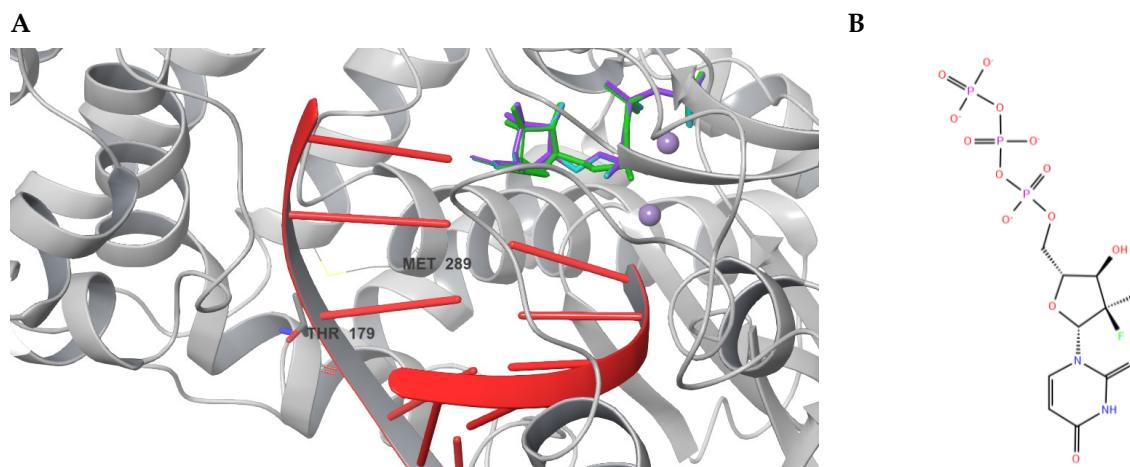


Figure 1. The docked poses of GS-461203 (A) in WT (green), T179A (blue) and M289L (purple), two manganese ions (purple), with the positions of mutated residues; and the chemical structure of the drug (B).

3.2 The RMSD Values

The RMSD of protein $C\alpha$ and GS-461203 of each system averaged over the two runs is depicted in Figure 2. It was showed that the protein $C\alpha$ was relatively more stable in WT compared to T179A and M289L, with mean values of 2.108 Å, which was lower than those in T179A (2.631 Å) and M289L (3.109 Å). Meanwhile, GS-461203 had a slight RMSD increase until ~ 250 ns in WT but became stable throughout the rest of the simulation, while it showed a higher RMSD movement since ~ 30 ns in the T179A. In M289L, GS-461203 seems to stabilize after ~ 100 ns in the M289L system. During MDS, the mean RMSD of GS-

461203 was 1.474 Å, 2.682 Å, and 2.083 Å, in WT, T179A, and M289L, respectively. The similar trend was observed in the individual trajectory that the mean RMSD of protein and ligand in WT system was lower than those in T179A and M289L (Table 1 and Figure 2), which indicated that the protein mutations has impacted the GS-461203 conformation to be more fluctuant.

Table 1. The mean RMSD of Protein C α atoms and Mean RMSD of ligand calculated for each trajectory.

System	Trajectory	Mean RMSD of Protein C α atoms (Å)	Mean RMSD of ligand (Å)
WT	Trajectory 1	2.039±0.218	1.079±0.216
	Trajectory 2	2.242±0.216	1.724±0.478
T179A	Trajectory 1	2.710±0.366	3.150±0.325
	Trajectory 2	2.552 ± 0.338	1.948±0.384
M289L	Trajectory 1	2.883±0.479	2.115±0.425
	Trajectory 2	3.417±0.544	1.928±0.619

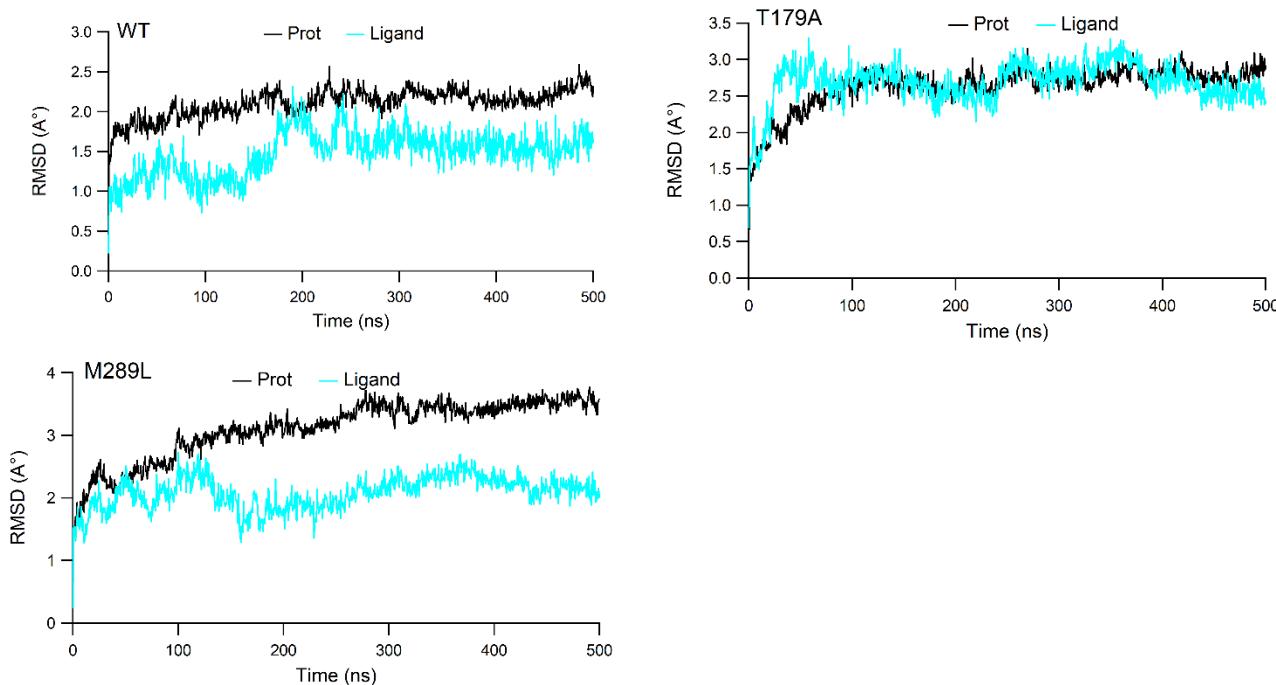


Figure 2. The root mean squared deviation (RMSD) of Protein C α and ligand in WT, T179A, S282T, and M289L systems.

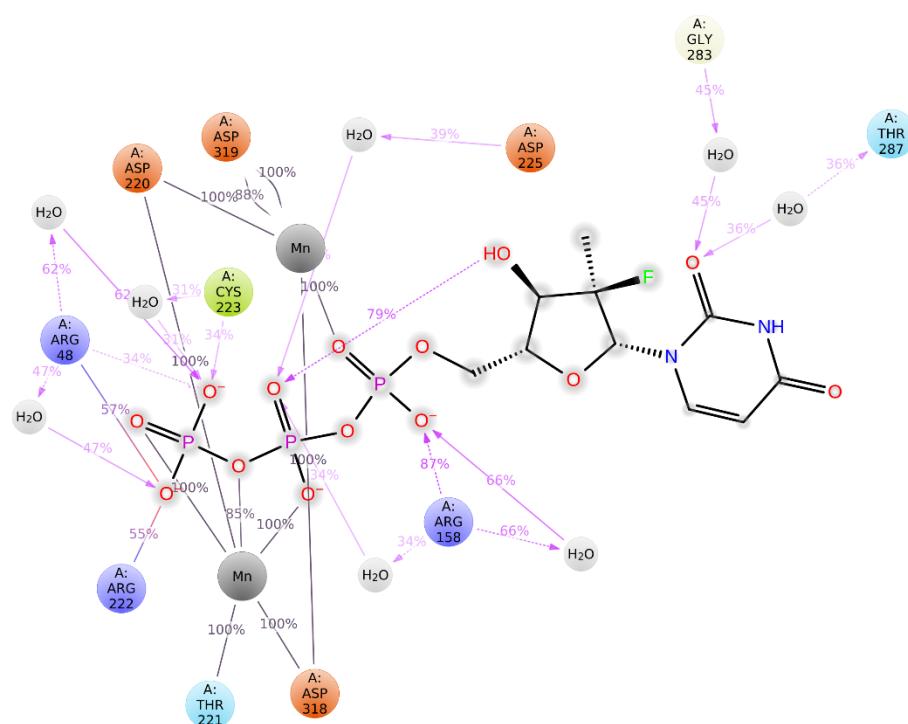
3.3 The Protein-Ligand Interaction Analysis

To acquire a protein-ligand interaction during MDS, analysis of the SID was performed. Interaction fraction with corresponding residues and protein-ligand interactions which sustained more than 20% of MDS is depicted in Figure 3 and Figure S2. In the WT system (Figure 3A), Hbond interactions were observed between the phosphate group of GS-461203, in addition to their interactions through water mediated Hbonds, at

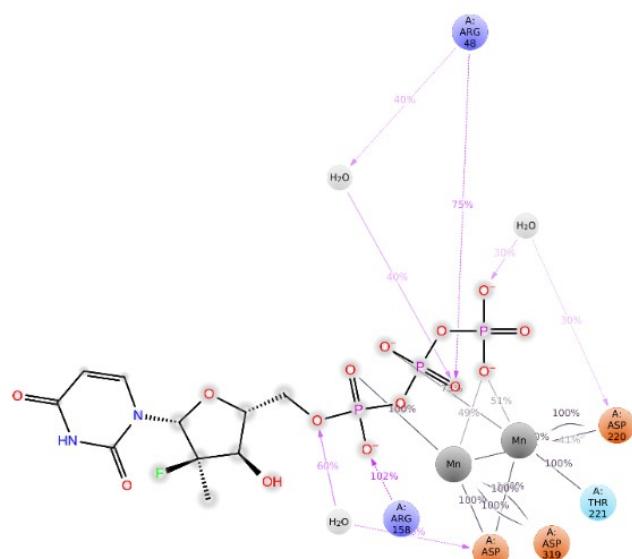
residues Arg158 (87%), Arg48 (57%), Arg222 (55%), and Cys223 (34%). The water mediated Hbond interactions was additionally observed between the GS-461203 uridine base with Gly283 (45%), Thr287 (36%) and between Asp225 (39%) with GS-461203 phosphate group. Manganese ions both interacted with 100% of simulation time with oxygen atoms of phosphate group aside from their interactions with Asp318, Asp319, Asp220, and Thr221, each with 100% simulation time, which indicates their role to stabilize GS-461203 binding.

171
172
173
174
175
176
177

A



B



C

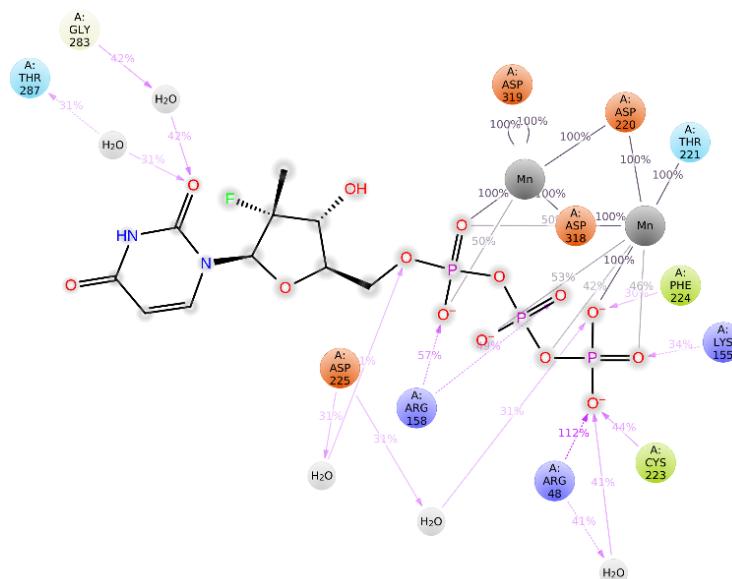


Figure 3. The 2D Protein-Ligand Interaction Diagrams lasting more than 20% of the MDS (A: WT-GS-461203; B: T179A-GS-461203; C: M289L-GS-461203) recorded during MD simulations.

In the T179A system (Figure 3B), Hbond interactions between GS-461203 with residues Arg158, and Arg48 were observed to increase to 102% and 75% simulation time, respectively. Additionally, the Hbond with Arg222 and Cys223 and the water mediated Hbond interactions between Gly283 and with the GS-461203 uridine base, and between Arg158 and Asp225 with GS-461203 phosphate groups disappeared. Manganese ion interaction with the phosphate group of GS-461203 were also decreased to 51% as compared to WT system, while ionic interactions with Asp318, Asp319, and Asp220 were kept at 100% simulation time. It was clear that the T179A mutation had diminished effects on ligand-protein interactions.

In the M289L system (Figure 3C), Hbond interactions between GS-461203 and Arg48, and GS-461203 and Cys223 were also observed with increased percentages of 112% and 44% simulation times, respectively. Alternatively, those with Arg158 were decreased to 57% simulation time. In addition, water mediated Hbond interactions between the GS-461203 uridine base with Gly283 and Thr287 were observed to slightly increase each with 42% and 31% simulation time, respectively. Additional Hbond interactions were observed with Phe224 (30) and Lys155 (34%), which did not exist in WT system. The Hbond with Arg222 disappeared in the M289L system. Just like in the WT and T179A systems, Thr221 contributed to Mn ions stabilization. Manganese ions were in close proximity with residues Asp318, Asp319, and Asp220 which were observed to stabilize the phosphate group of GS-461203. It is clearly seen that the binding modes of GS-461203 has changed in the two mutant systems.

3.4 Cluster analysis: the effect of mutation on whole protein structure

To acquire the dominant conformations of the protein-ligand complex, we

performed clustering analysis on the four systems. The protein alignment between T179A and WT systems from most populated cluster showed that the T179A mutant has the RMSD and TM-score values of 2.76 Å and 0.90268, respectively, while those between M289L and WT systems resulted in the RMSD and TM-score values of 3.95 Å and 0.82731, respectively. It is clear that the M289L movement was larger than that of T179A, each with reference to WT. Figure 4 displays comparison of protein structures between WT with each T179A and M289L taken from most populated clusters.

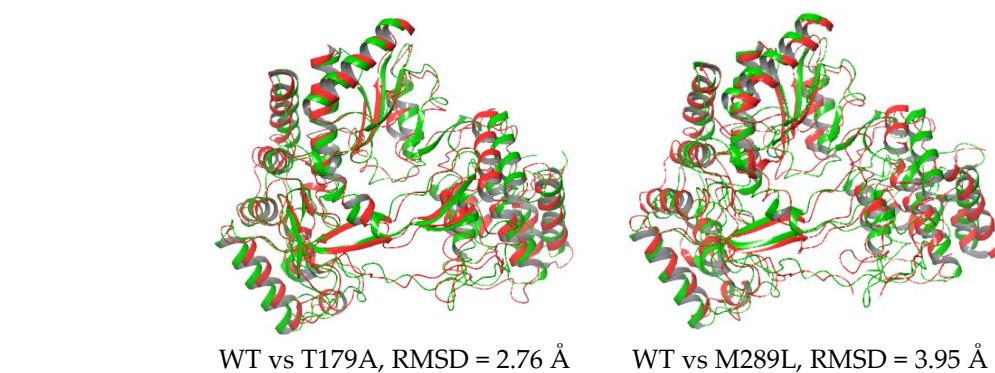


Figure 4. Comparison of protein structure between WT (green) and T179A (red) (A) and M289L (red) (B).

Further we analyzed the ligand binding sites from best cluster of each system. We found out that the GS-461203 maintained its interactions with key residues in WT, T179A, and M289L system. The residues D318, D319, D220 was in close distance with both Manganese ions which further interacted with phosphate group of GS-461203 as also found in experimental [24]. The residues R48 and R158 was also in close distance with phosphate groups of GS-461203. However, the interactions of fluorine atom of GS-461203 with N291 was more intense in M289L system as indicated by the distance of 2.73 Å, which was much lower than those in WT (5.36 Å) and T179A (5.95 Å). In addition, the phosphate group of GS-461203 in M289L system was involved in additional interactions with K155 and F224 with distances of 2.67 Å and 3.76 Å, respectively, which was absent in WT and T179A system. Figure 5 shows the active site conformation of WT, T179A, and M289L systems taken from most populated cluster.

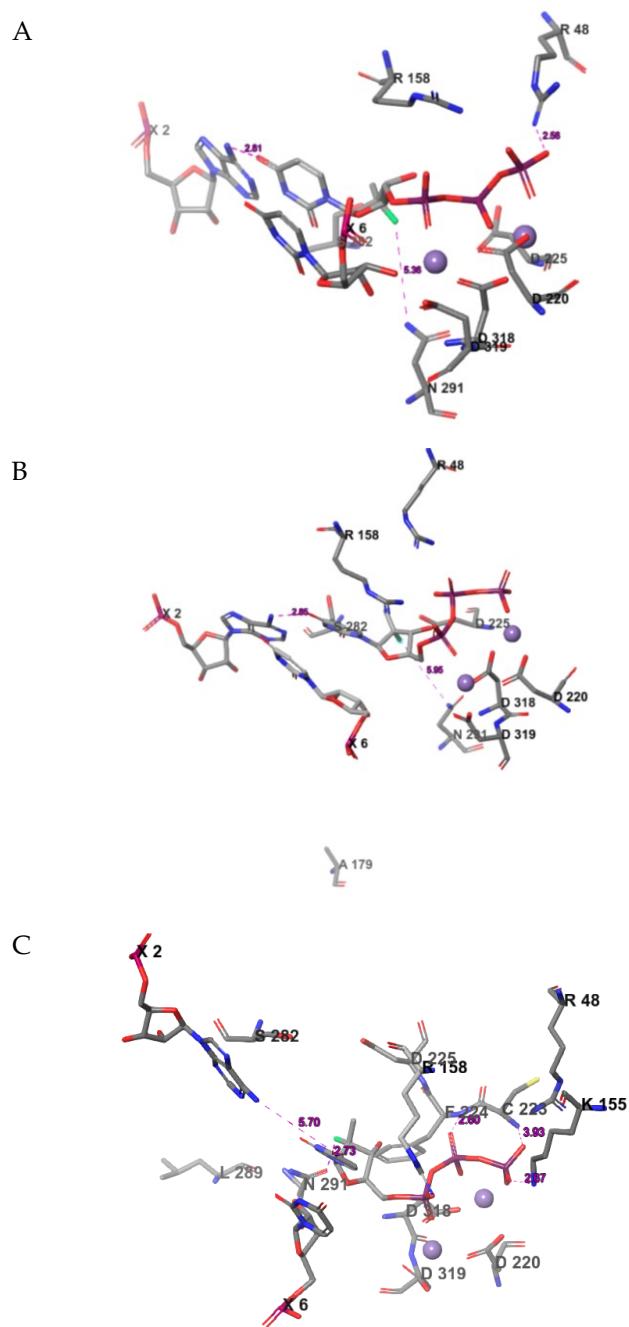
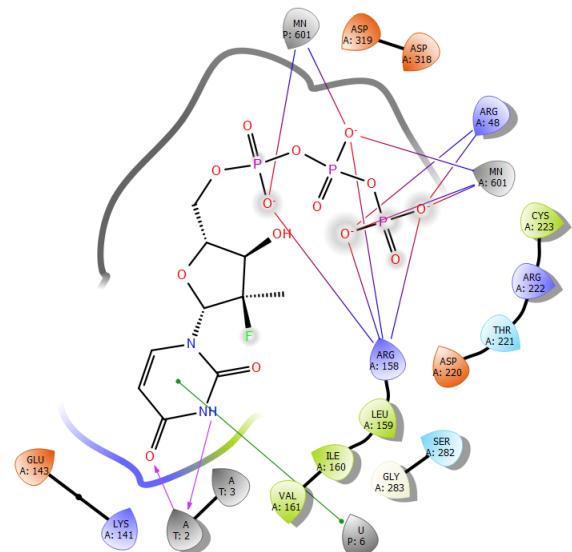


Figure 5. The active site conformation of WT (A), T179A (B), and M289L (C) systems taken from most populated cluster.

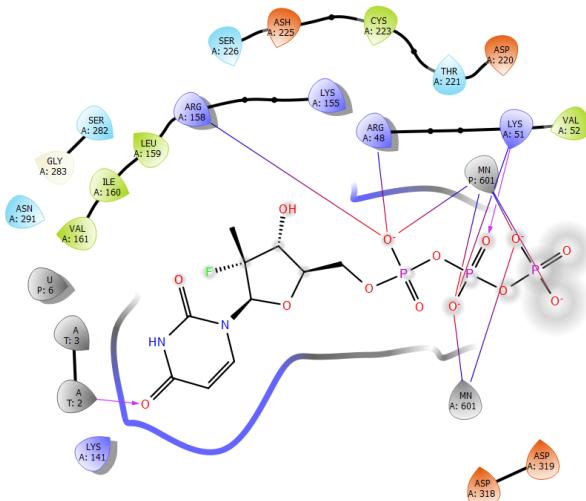
In Figure 6 depicts the 2D interactions of representative structure of the most dominant structural family for WT, T179A, and M289L systems. The pi-pi interactions were recorded in WT and M289L systems, however, it was absent in the T179 system. Salt bridge interactions with both Arg48 and Arg158 were recorded in WT, T179A, and M289L systems, while additional interactions with Lys155 and Phe224 were recorded in T179A and M289L systems, respectively. Manganese ions interacted through salt bridge interactions with the phosphate groups in all systems.

233
234
235
236
237
238
239
240
241
242
243

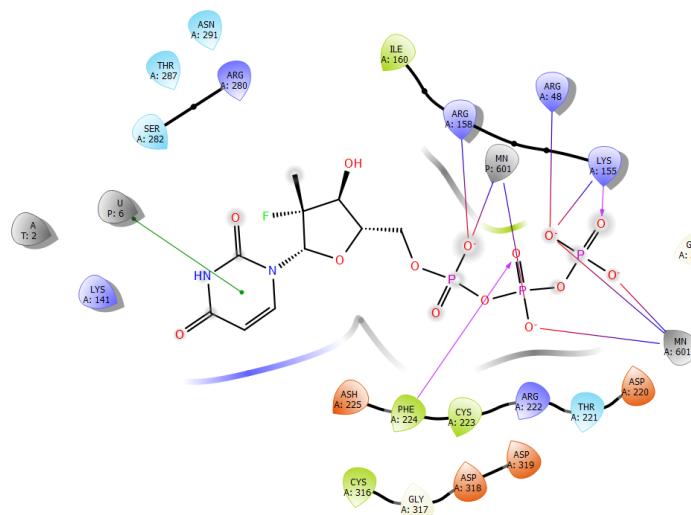
A



B



C

**Figure 6.** The representative structure of the most dominant structural family for **A**: WT; **B**: T179A; **C**: M289L systems.

3.5 MM-GBSA Binding Energy

The MM-GBSA binding energy was calculated to assess the effectiveness of the mutation on the binding energy of GS-461203. The total binding energy for the WT system was -45.5 kcal/mol, which was lower than that in the T179A (-35.0 kcal/mol) system, but was higher than that in the M289L system (-54.0 kcal/mol). The binding energy changes between the WT system and those in the T179A and M289L systems were 10.5 kcal/mol and 8.5 kcal/mol, respectively. Clearly, among the two mutations, T179A mutation resulted in the lower affinity of GS-461203, but a mutation in the M289L system resulted in a slightly stronger affinity of GS-461203. Table 2 shows the binding energy calculated for last 50 ns of each system.

Table 2. The binding energies calculated for the last 50 ns of each trajectory.

259	Systems	WT	T179A	M289L
260	ΔG_{bind}	-45.5 ± 5.8	-35.0 ± 5.3	-54.0 ± 7.7
261	$\Delta\Delta G_{\text{bind}}$	0.0	10.5	8.5
262	ΔE_{vdw}	-6.0 ± 6.8	-2.3 ± 6.1	4.5 ± 7.7
263	$\Delta\Delta E_{\text{vdw}}$	0.0	3.7	10.5
264	ΔE_{ele}	-35.3 ± 6.4	-28.7 ± 5.8	-55.1 ± 10.2
265	$\Delta\Delta E_{\text{ele}}$	0.0	6.6	19.8
266	ΔE_{lipo}	-4.3 ± 0.3	-4.0 ± 0.9	-3.4 ± 0.4
267	$\Delta\Delta E_{\text{lipo}}$	0.0	0.3	0.9

It was observed that the electrostatic energy (ΔE_{ele}) is the dominant factor when considering favorable binding contributions, each with -35.3 kcal/mol, -28.7 kcal/mol, and -55.1 kcal/mol in WT, T179A, and M289L, respectively. The differences of electrostatic interactions ($\Delta\Delta E_{\text{ele}}$) were 6.6 kcal/mol and 19.8 kcal/mol in T179A and M289L systems,

each compared to the WT system. The electrostatic energies were more positive in the T179A system, but it became more favorable in M289L compared to WT system. The possible explanations are the disappearance of Hbonds with Arg222 and Cys223, and water-mediated Hbond interactions with Gly283, Thr28, Arg158, and Asp225 in the T179A system (Figure 3B). The increased percentage of the electrostatic interactions with Arg48, Cys223, Gly282, Thr287 and the newly formed Hbond interactions with Phe224 and Lys155 in the M289L system (Figure 3C) seem to induce more favorable electrostatic interactions compared to the WT. We also calculated the binding energy for S282T system and we found out the total binding energy for the S282T system was -41.3 ± 5.4 kcal/mol (Table S1). The binding free energy difference between S282T and WT in our mmgbsa calculation is 4.2 kcal/mol, which is closer to the experimental value (3-fold difference in EC₅₀, $\Delta\Delta G=0.6$ kcal/mol) than that from the previous study (10.8 kcal/mol) [23].

Meanwhile, the van der Waals interactions were more favorable in the WT system (-6.0 kcal/mol) than in T179A (-2.3 kcal/mol) and M289L (4.5 kcal/mol). Clearly, the mutations have significantly reduced the van der Waals interactions, particularly in the M289L system with 10.5 kcal/mol change compared to the WT system, respectively. This may be due to the reduced van der Waals interactions with Ile160 in M289L system (Figure 3), which existed in the WT system. Meanwhile, the lipophilic interaction (ΔE_{lip}) did not change significantly after introducing mutations.

3.6 Protein RMSF Analyses

The RMSF of protein C α atoms were compared between protein WT and the mutants. Each prominent RMSF value was assigned with black, green, and blue arrows in WT, T179A, and M289L systems, respectively (Figure 7). The location of point mutation in each system was assigned with orange asterisks. The highest peaks are residues K151 of WT in both T179A and M289L systems excluding the carboxyl end of the protein. In M289L, the highest peak was at P22, while the highest peak of T179A was at A97, without considering the fluctuation of the carboxyl end. Concisely, mutation in T179A and M289L has the largest impact on residues A97, K151, and P22, respectively.

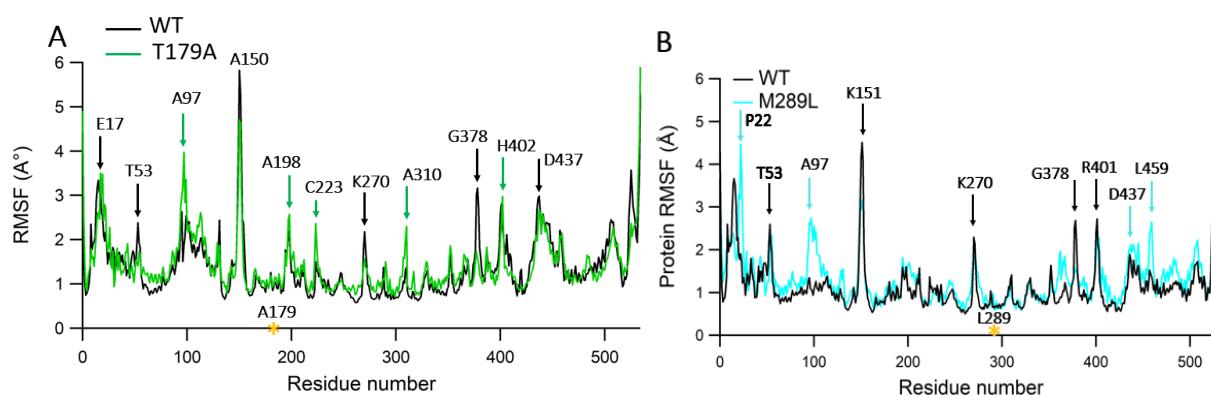


Figure 7. The RMSF values comparison between WT and T179A (A) and WT and M289L (B). The RMSF values of the C α atoms for residues in WT (black), T179A (green), and M289L (blue) are shown with the mutant positions are noted by red asterisks.

3.7 Ligand RMSF Values

The RMSF values of the ligand atoms were recorded as depicted in Figure 8. 306

The RMSF of GS-461203 in WT system was lower with mean RMSF values (1.093 Å) 307 compared to those in T179A (1.555 Å) and M289L (1.505 Å) systems, which indicated 308 that GS-461203 was more stable in the WT system compared to in T179A and M289L 309 systems. It is showed that the most fluctuant atoms were those in the the phosphate 310 group (atom numbers 28-30). 311

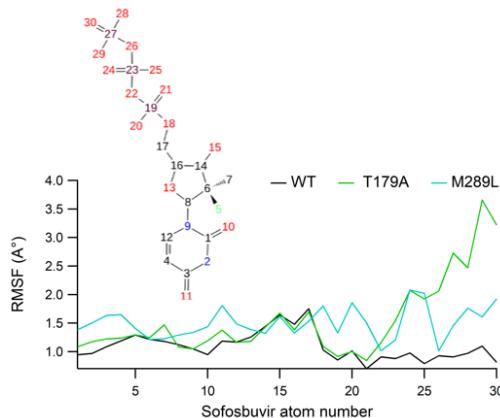
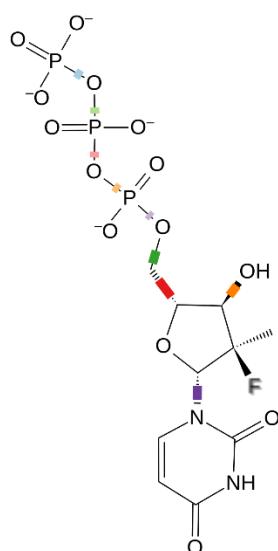


Figure 8. The RMSF values for atoms of GS-461203.

3.8 The GS-461203 Dihedral Angle Profiles

The dihedral angle of GS-461203 was depicted in Figure 9. It was showed that the 314 rotatable bonds, particularly those colored in light blue, light green, yellow, light purple, 315 red, and orange show a wider distribution in T179A and M289L systems, each as compared 316 to the WT system. The wider dihedral angle distributions were also observed in pink for 317 the T179A and M289L systems. Likewise, the green in T179A system, while the purple one 318 shows similar distribution in both WT and mutant systems. In brief, the mutant systems 319 induced a wider distribution of dihedral angles within GS-461203. 320



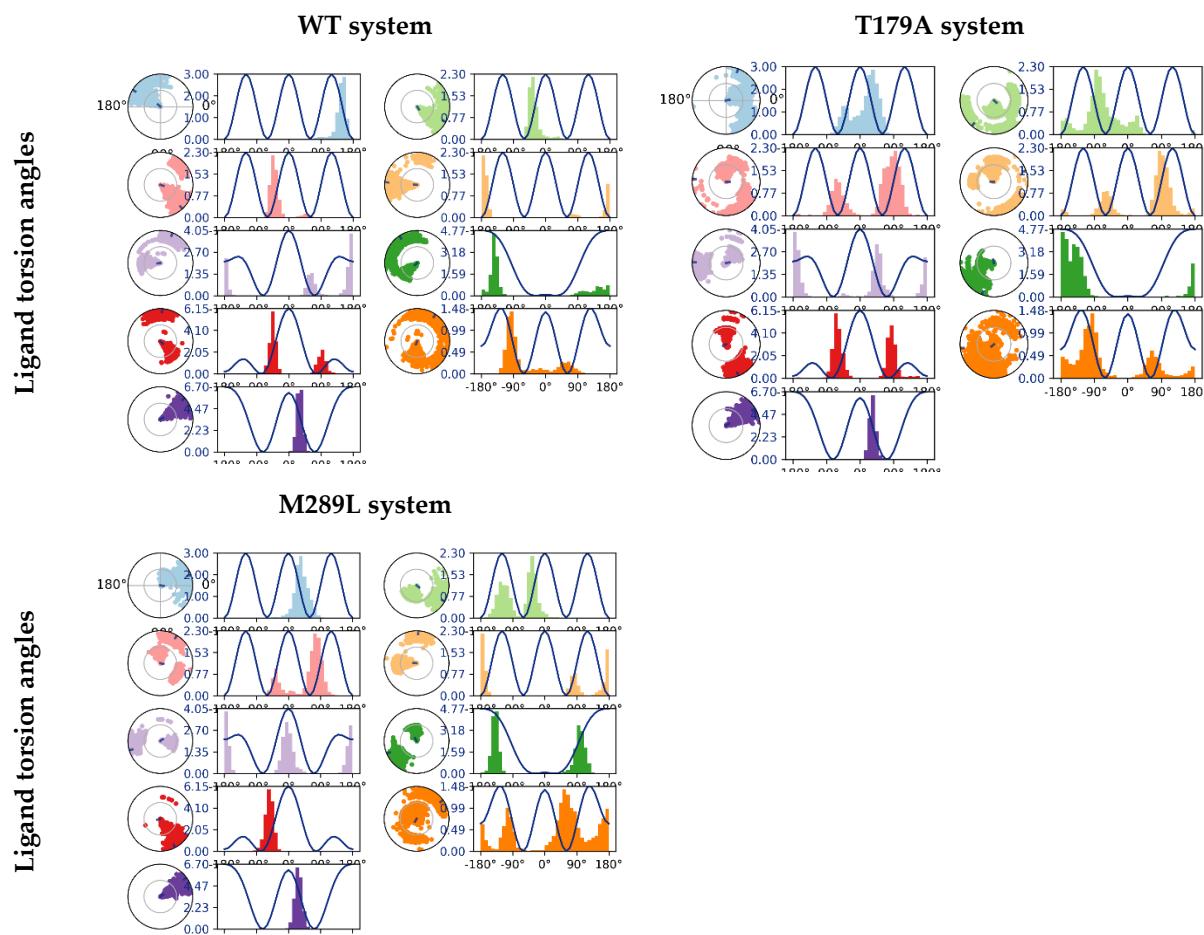


Figure 9. The dihedral angle of GS-461203 profiles during 500 ns MDS which is the conformational progression of the nine rotatable bonds of GS-461203. The dial plots describe the conformation of the torsion throughout the course of the simulation. The beginning of the simulation is in the center of the radial plot and the time evolution is plotted radially outwards. The bar plots summarize the data on the dial plots, by showing the probability density of the torsion. In addition, the 2D structure of GS-461203 in the top panel is for reference.

3.9 The secondary structure of protein

The SSE of protein monitored during MDS were plotted in Figure 10, which display the distribution of SSE including Alpha-helices and Beta-strands by residue. The total percentage of SSE for WT, T179A, and M289L were 47.40%, 46.09%, and 46.12%, respectively. The Alpha-helices composed 37.68% in WT system, which was comparable with 36.25% and 36.41% in T179A and M289L system, respectively, which indicated that the mutation had minor effects on SSE of protein.

326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341

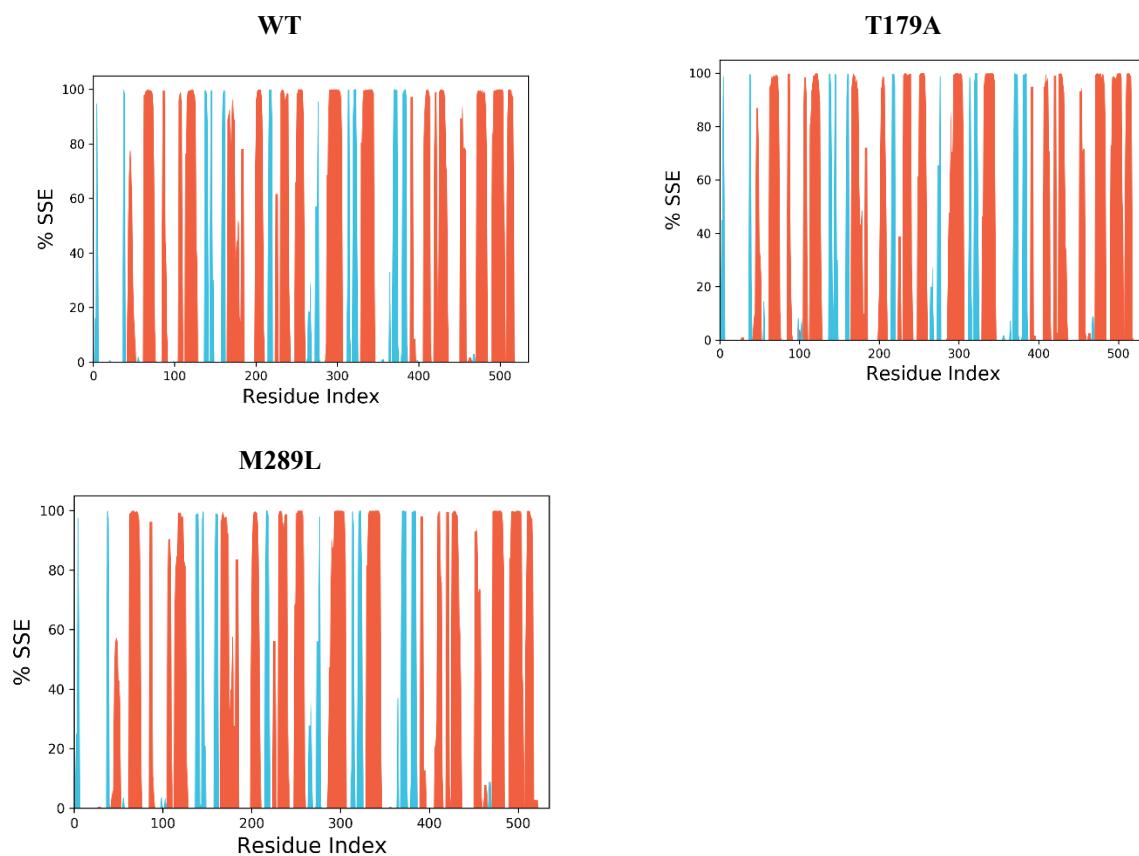
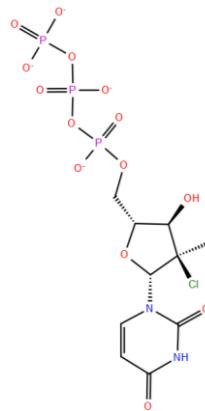


Figure 10. Protein SSE for the WT, T179A, and M289L systems during MDS. The alpha helices, beta sheets, and random coil were represented by red, blue, and white spaces.

3.10 Designing New Analogues

To improve GS-461203's binding to the T179A mutation, we designed new GS-461203 analogues by substituting the fluorine atom on the 2' position for either chloride or iodide atoms. The new structures of GS-461203 analogues are depicted in Figure 11.

CompA



CompB

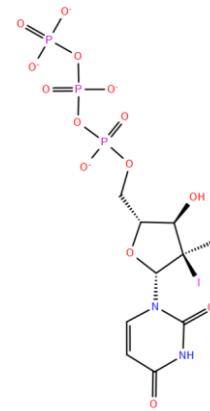


Figure 11. The designed GS-461203 analogues.

The two new analogues were submitted for MDS each for a duration of 500 ns. The RMSD values of $\text{C}\alpha$ protein are depicted in Figure S9. It was seen that the protein $\text{C}\alpha$

RMSD in T179A-CompA (2.324 Å) and T179A-CompB (2.199 Å) were lower than that of the T179A system (2.631 Å). Meanwhile, the ligand RMSD of GS-461203 in T179A was 2.682 Å, which was higher than those in T179A-CompA (1.869 Å) and T179A-CompB (2.173 Å).

In the T179A-CompA and T179A-CompB systems (Figure S10), Hbond interactions between GS-461203 and Arg158, and GS-461203 and Arg48 were also observed with high percentages of simulation time. The interaction between Manganese ions and the phosphate group of GS-461203 was increased to 100% simulation time, compared to the T179A system. In addition, Mn ions interactions with Asp318, Asp319, and Asp220 were kept at 100% simulation time. Moreover, the water-mediated Hbond interactions with Gly283 and Thr287 were observed with each 61% simulation in T179A-CompA system, which was absent in T179A system; while in T179A-CompB system, interactions with Thr287 were observed with a 58% simulation time. The newly formed Hbond with Thr221, Cys223, Asp225, and Phe224 were observed with high percentage of simulation time in T179A-CompA system while in the T179A-CompB system, interactions with Thr221, Asn291, Arg280, and Phe224 were recorded high percentages of simulation time.

The MM-GBSA binding energies are shown in Table 3. The total binding energy for the T179A system (-35.0 kcal/mol) was much higher than T179A-CompA (-73.2 kcal/mol) and T179A-CompB (-74.0 kcal/mol). The binding energy differences between T179A system and those in T179A-CompA and T179A-CompB systems were 38.2 kcal/mol and 39 kcal/mol, respectively. Clearly, the newly designed analogues have stronger affinities than that of GS-461203 in the T179A system.

Table 3 The binding energies calculated for the last 50 ns of each trajectory.

Systems	T179A	T179A- CompA	T179A-CompB
ΔG_{bind}	-35.0±5.3	-74.0±6.6	-73.2±6.9
$\Delta\Delta G_{\text{bind}}$	0.0	39	38.2
ΔE_{vdw}	-2.3±6.1	-13.3±6.9	-14.4±6.3
$\Delta\Delta E_{\text{vdw}}$	0.0	11.0	12.1
ΔE_{ele}	-28.7±5.8	-53.6±4.4	-53.0±6.0
$\Delta\Delta E_{\text{ele}}$	0.0	24.9	24.3
ΔE_{lipo}	-4.0±0.9	-7.1±0.4	-5.8±0.4
$\Delta\Delta E_{\text{lipo}}$	0.0	3.1	1.8

It is clear that the electrostatic energy (ΔE_{ele}) is the dominant factor in favorable binding contributions, each with -28.7 kcal/mol, -53.0 kcal/mol, and -53.6 kcal/mol in WT, T179A-CompA, and T179A-CompB, respectively. The differences of electrostatic interactions ($\Delta\Delta E_{\text{ele}}$) were 24.3 kcal/mol and 24.9 kcal/mol in the T179A-CompA and T179A-CompB systems, each compared to the T179A system. The electrostatic energies have largely increased in T179A-CompA and T179A-CompB systems. The possible

explanation of this is that the additional Hbond and Hbond water mediated interactions with Gly283 and Thr287, Thr221, Cys223, Asp225, and Phe224 in the T179A-CompA system, and interaction with Thr287, Thr221, Asn291, Arg280, and Phe224 in the T179A-CompB system.

Meanwhile, the van der Waals interactions were more favorable in the T179A-CompA and T179A-CompB systems (-14.4 kcal/mol and -13.3 kcal/mol, respectively) than in the T179A system (-2.3 kcal/mol). The newly designed analogues have significantly increased van der Waals interaction with the protein, with 12.1 kcal/mol and 11.0 kcal/mol differences, respectively, as compared to the T179A system (Figure S11). This was also seen for lipophilic interactions (ΔE_{lipo}) which were seen to be improved in the new analogues.

The RMSF of protein C α atoms was compared between the T179A system and the T179A-CompA and T179A-CompB systems (Figure S12), which showed similar modes of fluctuations. The RMSF values of ligand atoms were recorded as depicted in Figure S13. The RMSF of GS-461203 in T179A-CompA (1.022 Å) and T179A-CompB (1.036 Å) systems were lower than that in T179A system (1.555 Å), which indicate that the newly designed analogues induce more stable conformation of T179A.

3.11 Prediction ADME Properties

The ADME properties of compounds as predicted by SwissADME web server are displayed in Table 4 and Figure S15-S20. As the GS-461203 is a prodrug of sofosbuvir, the ADME properties were predicted for Sofosbuvir and its derivatives in which fluorine atom was substituted by chlorine or iodine atoms. The fluorine substitution by chlorine or iodine resulted in Log S (ESOL) values of -3.53 and -4.44 which grouped as soluble and moderately soluble, respectively, which were comparable with Sofosbuvir Log S (ESOL) value of -3.27 (soluble). The other ADME properties of Sofosbuvir derivatives including GI absorption, BBB permeant, cytochrome inhibition are all the same as Sofosbuvir. Further, the fluorine substitution to chlorine or iodine atoms in CompA and CompB, respectively, affected their solubilities. The Log S (ESOL) values for GS-461203, CompA, and CompB were 0.57, 0.31, and -0.61, respectively. Both GS-461203 and CompA were grouped as highly soluble, while CompB was categorized as very soluble, as predicted by the ESOL model [37]. It is clear that the ADME properties of designed compounds did not change significantly as compared to Sofosbuvir.

Table 4 The ADME properties of compounds predicted by SwissAdme.

Compound	Log S (ESOL)	GI absorption	BBB permeant	CYP1A2	CYP2C19	CYP2C9	CYP2D6	CYP3A4	Lipinski rule	PAINS
Sof ^a	-3.27 Soluble	Low	No	No	No	No	No	Yes	No; 2 violations: MW>500, NorO>10	0 alert
Sof-Cl ^a	-3.53 Soluble	Low	No	No	No	No	No	Yes	No; 2 violations: MW>500, NorO>10	0 alert

^aSof: Sofosbuvir; Sof-Cl and Sof-I: Sofosbuvir derivatives in which F was substituted by Cl and I, respectively.

4. Discussion

Sofosbuvir is an FDA-recommended drug used to treat HCV in all GTs through the inhibition of the virus RdRp. Previous experimental studies indicated that the residues R48 and R158 were shown to interact with phosphates of Sofosbuvir diphosphate [24, 38], while D220, D318, and D319 were shown to bind two catalytic Mn²⁺ ions [39]. In addition, the Hbonds with S282, N291, and D225, which exists when uridine diphosphate binds, were disrupted when Sofosbuvir diphosphate binds [24, 38]. Those experimental observations were in line with the present study found in the WT system.

Furthermore, despite the high conservation of RdRp among HCV GTs, researchers reported the emergence of resistance in GT2a due to GS-461203. The existence of a single substitution mutation of T179A or M289L in J6 GT2a replicons was suggested [18, 40-43]. Here we showed that the T179A mutation has reduced affinities of GS-461203 due to reduced electrostatic interactions. In the M289L system, the mutation induced much more positive van der Waals interactions, however, the effect was compensated by much more negative electrostatic interactions, which resulted in a slight increase of GS-461203 affinity. In case of M289L, which is located far enough from the active site of RdRp, the mutation effect slightly increased the GS-461203 affinity which was mostly due to increased electrostatic interaction.

In this study, the two newly designed halogen analogues of GS-461203 showed much stronger affinities than GS-461203 alone, which was due to more intense protein-ligand interactions as compared to the GS-461203. The previous systematic studies [44, 45] with a detailed database survey and quantum chemistry calculation have suggested that halogen bonds could play roles not only in drug-improving target binding affinity but also in tuning ADME/T property. Our binding and ADME prediction data is consistent with these

studies. In particular, the binding free energy predictions of each analogue in complex with T179A mutant showed a significant increase in to the ΔG_{bind} due to major contribution from the electrostatic interactions (ΔE_{ele}). This can be explained by the newly formed Hbond and Hbond water mediated interactions with Gly283 and Thr287, Thr221, Cys223, Asp225, and Phe224 in the T179A-CompA system, and interaction with Thr287, Thr221, Asn291, Arg280, and Phe224 in the T179A-CompB system. The van der Waals interactions (ΔE_{vdw}) were also more favorable in the analogues. This can be possibly explained by the increase in atomic/orbital size around the halogenated site seen with chlorine and iodine. This helps with shifting electron density towards the sugar ring in addition to the surrounding protein residues which can afford a higher opportunity for instantaneously induced dipole interactions. Additionally, the inclusion of more electron-rich atoms at the 2' position can have a large effect on the overall ligand geometry and surface area for opportunity of residue contact. This is seen with the substitution of a fluorine atom for chlorine or iodine in CompA or CompB, respectively, which further stabilized the sugar ring and CH₃ (atom 7) seen in the ligand RMSF plots in Figure S13. However, the improvement in stability and affinity of these analogues towards T179A comes with a predicted decrease in water solubility particularly with CompB.

5. Conclusions

This study was performed with the goal to understand the molecular details of GS-461203 binding to the HCV RdRp GT2a in WT and mutant systems using homology modeling, molecular docking and MDS. In this study, the key residues interacting with GS-461203 include Arg48, Arg158, Asp318, Asp319, and Asp220, which were shown to be in line with the experimental results. It is shown that the binding modes of GS-461203 has changed in mutant systems and resulted in reduced affinities in the T179A system. However, it induced a stronger affinity in the M289L system, which need further detailed studies to gain a deeper understanding of M289L system. In the case of T179A, the low affinity was due to weaker electrostatic interaction. Further, we designed two new analogues of GS-461203 and found out that the two compounds induced more stable interactions towards T179A than GS-461203. This was also demonstrated by the MM-GBSA prediction as much more favorable binding energies were obtained, which was due to the improved electrostatic, van der Waals, and lipophilic contributions, however further biological assays are needed to validate the potentials of the GS-461203 analogues.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Figure S1: title; Table S1: title; Video S1: title.

Author Contributions: Conceptualization, M.A. and C.W.; methodology, S.T.W.; software, C.W.; validation, M.S.Z, D.B. and M.S.; formal analysis, M.A.; investigation, D.B.; resources, M.S.; data

curation, M.A.; writing—original draft preparation, M.A.; writing—review and editing, C.W.; visualization, M.A.; supervision, C.W. All authors have read and agreed to the published version of the manuscript.	492 493 494 495
Funding: The APC was funded by World Class Professor (WCP) 2021 Ministry of Education, Culture, Research, and Technology, Republic of Indonesia.	496 497 498
Institutional Review Board Statement: Not applicable.	499
Informed Consent Statement: Not applicable.	500 501
Data Availability Statement: Not applicable.	502
Acknowledgments: MA would like to thank Ministry of Education, Culture, Research, and Technology, Republic of Indonesia, for supporting the research through Program World Class Professor (WCP) 2021 and the HPC Lab of Theoretical Physics Division, IPB University. CW thanks the support from the National Science Foundation of USA under Grants NSF RUI-1904797/ACI-1429467 and XSEDE MCB 170088.	503 504 505 506 507
Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.	508 509 510
References	511
1. Mohd Hanafiah, K., et al., <i>Global epidemiology of hepatitis C virus infection: New estimates of age-specific antibody to HCV seroprevalence</i> . Hepatology, 2013. 57 (4): p. 1333-1342.	512 513
2. Rabaan, A.A., et al., <i>Overview of hepatitis C infection, molecular biology, and new treatment</i> . Journal of Infection and Public Health, 2020. 13 (5): p. 773-783.	514 515
3. Ramirez, S. and J. Bukh, <i>Current status and future development of infectious cell-culture models for the major genotypes of hepatitis C virus: Essential tools in testing of antivirals and emerging vaccine strategies</i> . Antiviral Research, 2018. 158 : p. 264-287.	516 517 518
4. Bukh, J., <i>The history of hepatitis C virus (HCV): Basic research reveals unique features in phylogeny, evolution and the viral life cycle with new perspectives for epidemic control</i> . Journal of Hepatology, 2016. 65 (1): p. S2-S21.	519 520 521
5. Smith, D.B., et al., <i>Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: Updated criteria and genotype assignment web resource</i> . Hepatology, 2014. 59 (1): p. 318-327.	522 523
6. Anderson, J.C., et al., <i>Comparison of different HCV viral load and genotyping assays</i> . Journal of Clinical Virology, 2003. 28 (1): p. 27-37.	524 525
7. Petruzzello, A., et al., <i>Global epidemiology of hepatitis C virus infection: An up-date of the distribution and circulation of hepatitis C virus genotypes</i> . World journal of gastroenterology, 2016. 22 (34): p. 7824-7840.	526 527 528
8. Hedskog, C., et al., <i>Identification of 19 Novel Hepatitis C Virus Subtypes—Further Expanding HCV Classification</i> . Open Forum Infectious Diseases, 2019. 6 (3).	529 530
9. Chayama, K. and C.N. Hayes, <i>Hepatitis C virus: How genetic variability affects pathobiology of disease</i> . Journal of Gastroenterology and Hepatology, 2011. 26 (s1): p. 83-95.	531 532
10. Li, H.-C., C.-H. Yang, and S.-Y. Lo, <i>Hepatitis C Viral Replication Complex</i> . Viruses, 2021. 13 (3): p. 520.	533 534

11. Lam, A.M., et al., *PSI-7851, a pronucleotide of beta-D-2'-deoxy-2'-fluoro-2'-C-methyluridine monophosphate, is a potent and pan-genotype inhibitor of hepatitis C virus replication.* Antimicrobial agents and chemotherapy, 2010. **54**(8): p. 3187-3196. 535

12. Shah, N., T. Pierce, and K.V. Kowdley, *Review of direct-acting antiviral agents for the treatment of chronic hepatitis C.* Expert Opinion on Investigational Drugs, 2013. **22**(9): p. 1107-1121. 538

13. Patil, V.M., et al., *Current Perspective of HCV NS5B Inhibitors: A Review.* Current Medicinal Chemistry, 2011. **18**(36): p. 5564-5597. 540

14. Gentles, R.G., et al., *Discovery and Preclinical Characterization of the Cyclopropylindolobenzazepine BMS-791325, A Potent Allosteric Inhibitor of the Hepatitis C Virus NS5B Polymerase.* Journal of Medicinal Chemistry, 2014. **57**(5): p. 1855-1879. 541

15. Rosenberg, S., *Recent advances in the molecular biology of hepatitis C virus* Edited by F. Cohen. Journal of Molecular Biology, 2001. **313**(3): p. 451-464. 545

16. Götte, M. and J.J. Feld, *Direct-acting antiviral agents for hepatitis C: structural and mechanistic insights.* Nature Reviews Gastroenterology & Hepatology, 2016. **13**(6): p. 338-351. 547

17. Chen, Z.-w., et al., *Global prevalence of pre-existing HCV variants resistant to direct-acting antiviral agents (DAAs): mining the GenBank HCV genome data.* Scientific Reports, 2016. **6**(1): p. 20310. 549

18. Lam, A.M., et al., *Genotype and subtype profiling of PSI-7977 as a nucleotide inhibitor of hepatitis C virus.* Antimicrobial agents and chemotherapy, 2012. **56**(6): p. 3359-3368. 551

19. Lawitz, E., et al., *Sofosbuvir in combination with peginterferon alfa-2a and ribavirin for non-cirrhotic, treatment-naive patients with genotypes 1, 2, and 3 hepatitis C infection: a randomised, double-blind, phase 2 trial.* The Lancet Infectious Diseases, 2013. **13**(5): p. 401-408. 553

20. Jacobson, I.M., et al., *Sofosbuvir for Hepatitis C Genotype 2 or 3 in Patients without Treatment Options.* New England Journal of Medicine, 2013. **368**(20): p. 1867-1877. 555

21. Sorbo, M.C., et al., *Hepatitis C virus drug resistance associated substitutions and their clinical relevance: Update 2018.* Drug Resistance Updates, 2018. **37**: p. 17-39. 557

22. Arba, M., et al., *Mechanistic insight on the remdesivir binding to RNA-Dependent RNA polymerase (RdRp) of SARS-cov-2.* Computers in Biology and Medicine, 2021. **129**: p. 104156. 561

23. Pan, D., et al., *Computational study on the binding and unbinding mechanism of HCV NS5B with the inhibitor GS-461203 and substrate using conventional and steered molecular dynamics simulations.* Chemometrics and Intelligent Laboratory Systems, 2016. **156**: p. 72-80. 563

24. Appleby, T.C., et al., *Structural basis for RNA replication by the hepatitis C virus polymerase.* Science, 2015. **347**(6223): p. 771. 564

25. Madhavi Sastry, G., et al., *Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments.* Journal of Computer-Aided Molecular Design, 2013. **27**(3): p. 221-234. 566

26. Mark, P. and L. Nilsson, *Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K.* The Journal of Physical Chemistry A, 2001. **105**(43): p. 9954-9960. 569

27. Banks, J.L., et al., *Integrated Modeling Program, Applied Chemical Theory (IMPACT).* Journal of Computational Chemistry, 2005. **26**(16): p. 1752-1780. 571

28. Jorgensen, W.L., D.S. Maxwell, and J. Tirado-Rives, *Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids*. Journal of the American Chemical Society, 1996. **118**(45): p. 11225-11236. 575
576

29. Shivakumar, D., et al., *Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field*. Journal of Chemical Theory and Computation, 2010. **6**(5): p. 1509-1519. 578
579
580

30. Kumar, V., H. Liu, and C. Wu, *Drug repurposing against SARS-CoV-2 receptor binding domain using ensemble-based virtual screening and molecular dynamics simulations*. Computers in Biology and Medicine, 2021. **135**: p. 104634. 581
582
583

31. Bowers, K.J., et al. *Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters*. in *SC '06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*. 2006. 584
585

32. Zhang, Y. and J. Skolnick, *TM-align: a protein structure alignment algorithm based on the TM-score*. Nucleic acids research, 2005. **33**(7): p. 2302-2309. 586
587

33. Li, J., et al., *The VSGB 2.0 model: a next generation energy model for high resolution protein structure modeling*. Proteins, 2011. **79**(10): p. 2794-2812. 588
589

34. Daina, A., O. Michelin, and V. Zoete, *SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules*. Scientific Reports, 2017. **7**(1): p. 42717. 590
591
592

35. Friesner, R.A., et al., *Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy*. Journal of Medicinal Chemistry, 2004. **47**(7): p. 1739-1749. 593
594

36. Friesner, R.A., et al., *Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes*. Journal of Medicinal Chemistry, 2006. **49**(21): p. 6177-6196. 595
596
597

37. Delaney, J.S., *ESOL: Estimating Aqueous Solubility Directly from Molecular Structure*. Journal of Chemical Information and Computer Sciences, 2004. **44**(3): p. 1000-1005. 598
599

38. Sofia, M.J. and P.A. Furman, *The Discovery of Sofosbuvir: A Liver-Targeted Nucleotide Prodrug for the Treatment and Cure of HCV*, in *HCV: The Journey from Discovery to a Cure: Volume I*, M.J. Sofia, Editor. 2019, Springer International Publishing: Cham. p. 141-169. 600
601
602

39. Sofia, M.J., et al., *Nucleoside, Nucleotide, and Non-Nucleoside Inhibitors of Hepatitis C Virus NS5B RNA-Dependent RNA-Polymerase*. Journal of Medicinal Chemistry, 2012. **55**(6): p. 2481-2531. 603
604

40. Bagaglio, S., C. Uberti-Foppa, and G. Morsica, *Resistance Mechanisms in Hepatitis C Virus: implications for Direct-Acting Antiviral Use*. Drugs, 2017. **77**(10): p. 1043-1055. 605
606

41. Migliaccio, G., et al., *Characterization of Resistance to Non-obligate Chain-terminating Ribonucleoside Analogs That Inhibit Hepatitis C Virus Replication in Vitro**. Journal of Biological Chemistry, 2003. **278**(49): p. 49164-49170. 607
608
609

42. Svarovskaia, E.S., et al., *Infrequent development of resistance in genotype 1-6 hepatitis C virus-infected subjects treated with sofosbuvir in phase 2 and 3 clinical trials*. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 2014. **59**(12): p. 1666-1674. 610
611
612

43. Xu, S., et al., *In vitro selection of resistance to sofosbuvir in HCV replicons of genotype-1 to -6*. (2040-2058 (Electronic)). 613
614

44. Staroń, J., et al., *Tuning the activity of known drugs via the introduction of halogen atoms, a case study of SERT ligands – Fluoxetine and fluvoxamine*. European Journal of Medicinal Chemistry, 2021. **220**: p. 113533. 615
616
617

45. Xu, Z., et al., *Halogen Bond: Its Role beyond Drug–Target Binding Affinity for Drug Discovery and Development*. Journal of Chemical Information and Modeling, 2014. **54**(1): p. 69–78. 618
619
620
621
622