JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Detailed Computational Study of the Active Site of the Hepatitis C Viral RNA Polymerase to Aid Novel Drug Design

Khaled H. Barakat,*[,†,‡] John Law,[†] Alessio Prunotto,[§] Wendy C. Magee,[†] David H. Evans,[†] D. Lorne Tyrrell,[†] Jack Tuszynski,[⊥,∥] and Michael Houghton*[,†]

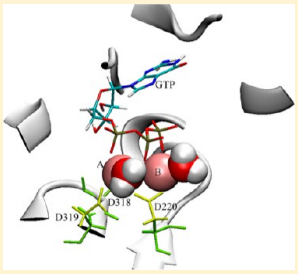
[†]Li Ka Shing Institute of Virology, Department of Medical Microbiology and Immunology, [⊥]Department of Oncology, and [∥]Department of Physics, University of Alberta, Edmonton, AB, Canada T6G 2E1

[‡]Department of Engineering Mathematics and Physics, Fayoum University, Fayoum, Egypt

[§]Department of Aerospace and Mechanical Engineering, Politecnico di Torino, 24-10129 Torino, Italy

Supporting Information

ABSTRACT: The hepatitis C virus (HCV) RNA polymerase, NS5B, is a leading target for novel and selective HCV drug design. The enzyme has been the subject of intensive drug discovery aimed at developing direct acting antiviral (DAA) agents that inhibit its activity and hence prevent the virus from replicating its genome. In this study, we focus on one class of NS5B inhibitors, namely nucleos(t)ide mimetics. Forty-one distinct nucleotide structures have been modeled within the active site of NS5B for the six major HCV genotypes. Our comprehensive modeling protocol employed 287 different molecular dynamics simulations combined with the molecular mechanics/Poisson−Boltzmann surface area (MM-PBSA) methodology to rank and analyze these structures for all genotypes. The binding interactions of the individual compounds have been investigated and reduced to the atomic level. The present study significantly refines our understanding of the mode of action of NS5B-nucleotide-inhibitors, identifies the key structural elements necessary for their activity, and implements the tools for ranking the potential of additional much needed novel inhibitors of NS5B.



1. INTRODUCTION

The social, clinical, and economic burden of the hepatitis C virus (HCV) on the world population, especially in the developing countries, is immense. Today, over 180 million people are living with chronic HCV infection,[1] a disease that does not always present symptoms until there is serious liver damage. The incidence of the disease is raising with an estimated 2.5 to 4 million new infections every year. The virus is the major cause of chronic liver disease in Europe and the USA, with nearly 10 000 deaths in the USA each year due to liver failure or hepatocellular carcinoma. In China, over 30 million people are chronically infected with the virus. Egypt has the highest prevalence in the world with approximately 25% of the population being infected with HCV.[2−4]

HCV is a single-stranded RNA virus belonging to the *Flaviviridae* family. Similar to other viruses, the HCV viral genome is translated into a long polypeptide that is cleaved by cellular and viral proteases into the necessary proteins for its replication and repackaging.[5] Among these proteins is the viral RNA polymerase (NS5B) that is responsible for replicating the HCV genome.[6] The enzyme lacks proofreading ability. As a result, its products usually experience high mutation rates, with approximately 10^{-5} nucleotide substitutions per replicated base.[7] These mutations have led to the emergence of at least six genotypes of the HCV virus that are less than ∼70% identical in nucleotide sequence.[8] This means that each of these genotypes can be viewed as a distinctive target for HCV treatment. To make the problem even more challenging, each genotype has its own family of subtypes, with differences in their nucleotide sequences in the 20−25% range.[8,9] Genotypes 1−3 are mainly concentrated in the western world, Australia, and East Asia, while genotypes 4−6 dominate the Middle East, South Africa, and Southeast Asia.[10]

A vital need is the development of a potent therapy that targets the different genotypes with minimal side effects. Since the discovery of HCV in the late-1980s,[11] considerable progress has been made toward achieving this goal. Until recently the standard of care (SOC) for HCV patients was PEGylated interferon-alpha (PEGIFN-α) in combination with the nucleoside analogue, ribavirin administered over a period ranging from 24 to 48 weeks depending on the genotype. This combination was effective in approximately 50% of genotype 1 patients, 90% of genotype 2, and 70% of genotype 3 patients. Side effects of this therapy frequently prevented their use in some patients or resulted in an incomplete therapeutic regimen. For genotype 1 patients, the newer SOC includes inhibitors of the viral serine protease used in combination with interferon-alpha and ribavirin and results in cure rates of ∼70%, often with shorter therapy durations. However, severe side effects can occur with their use.

Fortunately, recent advances in understanding the HCV life cycle and genomic organization revealed significant information about the structure and function of its different viral

Received: July 7, 2013

   dx.doi.org/10.1021/ci4003969 | *J. Chem. Inf. Model.* XXXX, XXX, XXX−XXX

components. This, in turn, has directed HCV research toward new avenues for rational drug design aimed at finding a novel HCV therapy. The newly developed therapies directly target specific enzymes that play important roles in HCV replication and propagation.[12] Among these targeted enzymes is the NS5B viral polymerase, which is the focus of this work. The enzyme has been the subject of intensive drug discovery investigations aimed at developing directly acting antiviral (DAA) agents that inhibit its activity and, hence, preventing the virus from replicating its genome.[13] These efforts led to the development of many small-molecule-inhibitors that can occupy the catalytic active site,[14] therefore blocking its interaction with the substrate, result in chain termination, or bind to one of four distinctive allosteric sites,[13] thus preventing the protein structure assuming an active conformation. Drugs that target the active site of the enzyme are typically nucleotide analogues that imitate the enzyme's natural substrate and act as chain terminators or inhibitors of chain initiation. On the other hand, those that bind to the allosteric sites span a large and diverse chemical space that varies according to the sites they occupy. Several NS5B-inhibitors are in various stages of clinical trails, but to date none have received market approval.[15]

The present investigation is focused solely on active site inhibitors and has two objectives. First, the technology used revealed a set of dynamic models for NS5B nucleotide inhibitors and illustrates their interactions with the different elements of the active site. These interactions were investigated computationally at the atomic level and ranked using the MM/PBSA (molecular mechanics/Poisson−Boltzmann surface area) technique.[16] Reassuringly, and similar to our previous work,[17−19] our modeling protocol was able to rank and reproduce the binding affinities of known nucleotide-mimetic and inhibitory structures with great accuracy. Second, the present study explores the HCV genome diversity at the active site as well as the common mutations that are usually correlated with an acquired resistance against NS5B nucleotide inhibitors. On the basis of these objectives, we focused on almost all known nucleotide mimetics that were reported in the literature as potential inhibitors of NS5B activity. Our goal was to build a detailed structural model that could be used to develop new and more potent inhibitors targeting NS5B. To the best of our knowledge, this is the first report offering such a comprehensive model for the NS5B polymerase. This study presents a broad investigation focused on understanding the mode of action of these compounds, identifies the key structural elements necessary for their activity, and provides the tools for ranking and developing novel inhibitors of NS5B activity.

2. MATERIALS AND METHODS

2.1. Modeling NS5B for Different Genotypes. The current study investigated six different major HCV genotypes observed around the globe (1b isolate BK, 1a isolate H77, 2a isolate JFH-1, 3a isolate NZL1, 4a isolate ED43, 5a isolate EUH1480, and 6a isolate EUHK2). All protein structures were created by modifying the NS5B-UTP crystal structure found in the Protein Data Bank (PDB code: 1GX6).[20] This structure describes a high-resolution (1.85 Å) conformation of genotype 1b bound to a nucleotide substrate at the active site. The two catalytic manganese ions included in the initial crystal structure were replaced by magnesium in our model (see Discussion). These two magnesium ions are coordinated by the phosphate groups of the NTP substrate and three aspartate residues, namely D220, D318, and D319. To generate structures for the other genotypes, their sequences were obtained from the Protein Knowledge Database (UniProtKB: http://www.uniprot.org/) and were aligned to the sequence of 1GX6 using the alignment utility of the same database. This was followed by mutational analysis using the Swiss-PdbViewer software (http://spdbv.vital-it.ch/) to replace every amino acid from 1GX6 with its counterpart that represents the targeted genome. Cocrystallized water molecules were preserved during the modeling process as they showed extreme importance in maintaining the stability of the bound nucleotides (see below). Rotamer libraries were used to avoid any steric clashes of the newly generated amino acid with its environment and to maximize its hydrogen bonding with the surrounding amino acids.

2.2. Preparation of Nucleotide−Polymerase Complexes. For all genotypes, the bound conformation of UTP within 1GX6 was retained as the expected common binding trend for all nucleotides. Nucleotide analogues were generated by keeping the triphosphate in its coordinates bound to the two Mg^{2+} ions while modifying the base and sugar moieties. The Molecular Operating Environment (MOE)[21] software was used to prepare and create the different nucleotides using its builder module followed by a short minimization of the active site and bound ligands to avoid any steric clashes with the protein side chains. In this way, for each genotype, 41 different nucleotides that represent the set of experimentally tested compounds in the literature were prepared in a clean and realistic binding conformation suitable for subsequent molecular dynamics simulations.

2.3. Molecular Dynamics (MD) Simulations. For all generated NS5B−nucleotide structures and all genotype protein structures, MD simulations were carried out using the NAMD program,[22] at a mean temperature of 300 K and physiological pH (pH 7) using the all-hydrogen AMBER99SB force field for the protein[23] and the generalized amber force field (GAFF) for the ligands.[24] The simulation protocol was similar to our previously described work.[25−27] Briefly, for each of the seven used NS5B structures, we created 41 nucleotide−NS5B complexes, which resulted in 287 different MD simulations. Protonation states of all ionizable residues were calculated using the program PDB2PQR.[28] Each bound system was then solvated in a cube of water residues, which provided at least a 12 Å-wide buffer followed by neutralizing the system with sodium and chloride ions. The 287 systems were then minimized, heated to 300 K with heavy restraints placed on all backbone atoms and equilibrated using periodic boundary conditions for 100 ps where energy restraints were reduced to zero in successive steps of the MD simulation. The simulations were then continued for 30 ns. During the last 10 ns of MD simulations, atomic coordinates were saved to the trajectory every 10 ps, producing 1000 snapshots for each protein−ligand we studied. This time interval was selected to be twice the largest correlation time (which ranged from 3 to 5 ps in our calculations) to ensure that all binding energy data are statistically independent and uncorrelated. Atomic fluctuation (B-factors) and root mean deviations from the reference structures (RMSD) were calculated over the simulation time using the PTRAJ utility within AMBER10.

2.4. MM-PBSA. We used the MM-PBSA technique[16] to predict the binding energies and study the energetic contributions of the different residues of the active site. Similar to the work described previously in the literature,[25−27] the total free energy for each system is estimated as the sum of the

dx.doi.org/10.1021/ci4003969 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

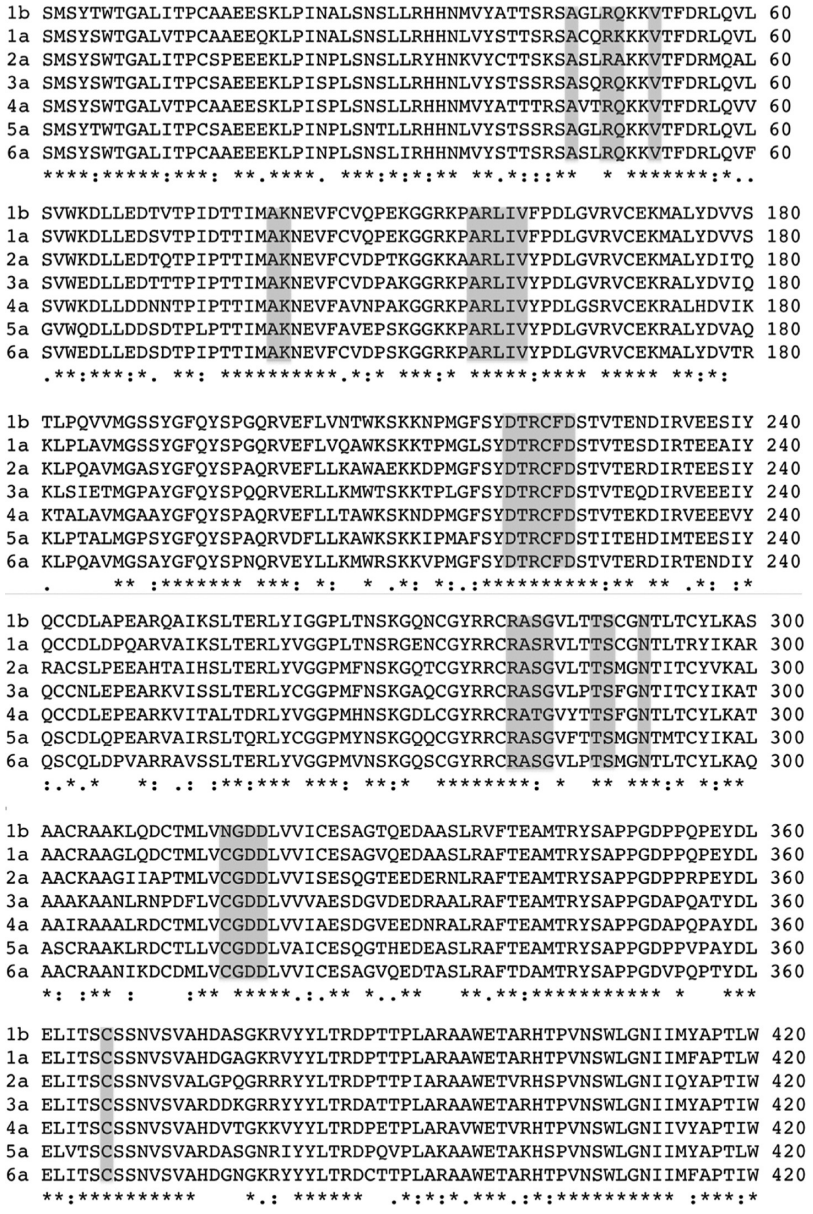


Figure 1. Sequence alignment for the genotypes included in this study. Residues forming the active site are highlighted in gray. Conserved residues are marked with a star underneath the residue names. The active site is conserved except for four mutations. UniProtKB IDs are P26663 for genotype 1b, P27958 for genotype 1a, Q99IB8 for genotype 2a, Q81258 for genotype 3a, O39929 for genotype 4a, O39928 for genotype 5a, and O39927 for genotype 6a.

average molecular mechanical gas-phase energies (E_{MM}), solvation free energies (G_{solv}), and entropy contributions (TS_{solute}) of the binding reaction:

$$G = E_{MM} + G_{solv} - TS_{solute} \quad (1)$$

The molecular mechanical (E_{MM}) energy of each snapshot was calculated using the SANDER module of AMBER10. The solvation free energy (G_{solv}) was estimated as the sum of electrostatic solvation free energy, calculated by the finite-difference solution of the Poisson−Boltzmann equation in the adaptive Poisson−Boltzmann solver (APBS) and nonpolar solvation free energy, calculated from the solvent-accessible surface area (SASA) algorithm. The solute entropy was approximated using the normal-mode analysis. For each protein−ligand complex, the binding free energy was approximated by the difference between the bound and free systems:

$$\Delta G^{\circ} = \Delta G_{gas}^{NS5B-ligand} + \Delta G_{solv}^{NS5B-ligand} - \{\Delta G_{solv}^{ligand} + \Delta G_{solv}^{NS5B}\} \quad (2)$$

The selection of the snapshots' frequency for calculating the MM-PBSA energies was based on estimating the correlation time similar to the work described by Genheden and Ryde.[29] That is, the delta MM-PBSA energy points from the whole MD trajectory (X) have been divided into blocks (Y_i) of equal time spaces (τ). The function Φ is then calculated according to the following equation:

$$\Phi = \frac{\tau\sigma^2(Y)_\tau}{\sigma^2(X)} \quad (3)$$

dx.doi.org/10.1021/ci4003969 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Where $\sigma^2(X)$ is the variance of the whole trajectory delta MM-PBSA energy points and $\sigma^2(Y)_\tau$ is the variance of the averages of the energy data points within the blocks of length τ (i.e., for each block the average delta energy is calculated then the variance of the n blocks generated is then used in eq 3 as $\sigma^2(Y)_\tau$ for a certain τ). The length of the block (τ) is then varied and the values of Φ is expected to be constant when the block averages are statistically independent and at this point the time correlation can be estimated.

2.5. Energy Decomposition Per Residue. The solvation and ligand−protein vdW and electrostatic interaction energies were broken down into their residue terms:

$$\Delta G_{\text{gas}}^{\text{NS5B-ligand}} \approx \Delta E_{\text{gas}}^{\text{vdW}} + \Delta E_{\text{gas}}^{\text{ele}} \quad (4)$$

Furthermore, the solvation free energy was considered as a sum of two terms, namely a nonelectrostatic and electrostatic terms:

$$\Delta G_{\text{solv}} \approx \Delta G_{\text{solv}}^{\text{nonele}} + \Delta_{\text{solv}}^{\text{ele}} \quad (5)$$

The nonelectrostatic term was approximated by a linear function of the (SASA). That is

$$\Delta G_{\text{solv}}^{\text{nonele}} = \gamma \text{SASA} \quad (6)$$

where γ = 7.2 cal/(mol Å^2).

3. RESULTS

3.1. Sequence Alignment. Figure 1 demonstrates the similarities among the different genotypes that were included in this study. The reference sequence is 1b, which represents the initial structure that was used to generate the other structures (see Materials and Methods). As shown, the active site is clearly conserved among the different genotypes except for five mutations. These mutations include S282T (genotype 4a), G283R (genotype 1a), C316N (genotype 1b), Q49K (genotype 1a), and Q49A (genotype 2a). The locations of these mutations around the active site are shown in Figure 2. The closest of these mutations to the bound nucleotide is S282T. This is expected, as the S282 mutation is a well-known drug-resistant mutation.[30] The most distal mutation is Q48A/K. The significance of these mutations for the binding affinity and the binding mode of the substrate nucleotide will be discussed in more details below.

3.2. Molecular Dynamics (MD) Simulations. MD simulations have been always used as a practical and fairly rapid tool to predict the most realistic interactions between a binding site and its bound ligand. Depending on the size of the system and the simulation time, an MD simulation may require extensive computational resources. Nevertheless, the output trajectory of a typical simulation reveals extremely valuable information about the system under investigation.[31,32] For example, by observing the trajectory one can understand the role of water molecules in coordinating the interactions between the ligand and the receptor and investigate the stability and flexibility of the system.

The high-resolution crystal structure that we used in this study helped us to accurately model the bound nucleotide within the active site. In particular, it was important to understand and confirm the location of the two Mg^{2+} ions and their interactions with the three-aspartate residues (D220, D318, and D319), which seem to play a crucial rule in the nucleotide transfer reaction along with the cocrystallized water molecules. Before we started transforming the 1b crystal

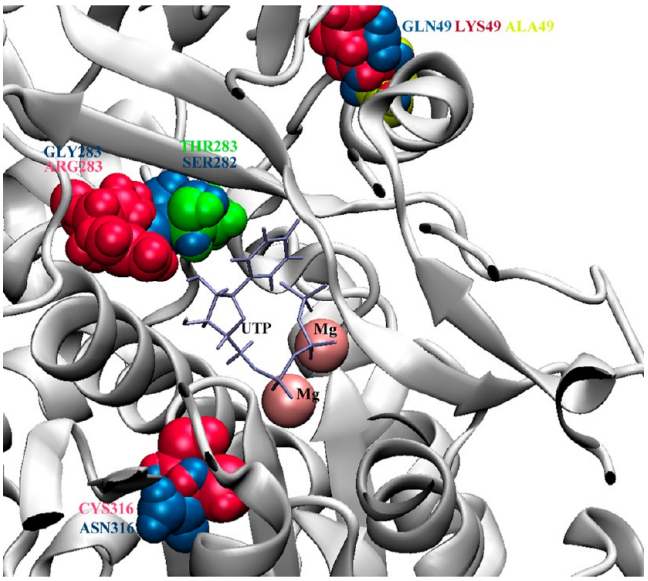


Figure 2. Location of the mutations within the active site. The protein is shown in white with a new cartoon representation. The bound nucleotide is colored in mauve, magnesium ions are colored in pink, and the mutated residues are shown in vdW representation with different colors for the different genotypes.

structure into the different genotypes and before modifying the bound UTP molecule into the different nucleotide analogues, we equilibrated the crystal structure through a 30 ns MD simulation. In our equilibrated model, the two Mg^{2+} ions were of great importance to stabilize the bound structure and seem to play important roles in screening the negative charge of the phosphate backbone allowing the active site to adopt the preferred catalytic conformation as well as coordinating the catalytic aspartate residues by bringing them closely together in order to perform their chemical function on the bound nucleotide.

Taking this equilibrated NS5B structure as a starting point, we modified the bound UTP molecule into the different 41-nucleotide inhibitor structures as well as created structures for the different genotypes. That is, 287 NS5B−ligand structures (7 genotypes × 41 nucleotide structures) were generated and were subjected to 30 ns MD simulations. Figure 3 shows two examples of such simulations. For each ligand studied in this work, we compared its flexibility in the bound form to that of its free solvated structure. This was important to elucidate which regions of the ligand are important in binding so one can preserve them in constructing future derivatives. On the other hand, highly flexible regions in the bound structure tended to be unimportant for the binding and, hence, could be substituted with other moieties in order to have them more tightly bound. For example, Figure 3 compares the flexibility of 2′-deoxy-2′-fluoro-2′-C-methyluridine (**8**) to that of 2′-methyl-7-deaza-7-fluoro-adenosine (**30**). In this regard, one can see that the two ligands presented in Figure 3 are favorably stabilized within the active site. This is clear in the RMSD data for the two compounds, where the free structures in both cases fluctuated and spanned different confirmations, while in the bound structures they were very stable with one binding mode fixed throughout the simulation. This was also apparent from the B-factors for the atomic coordinates of the two structures. While the free structures exhibited considerable fluctuations, the bound structures were extremely stable. The heterocycles for all

dx.doi.org/10.1021/ci4003969 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

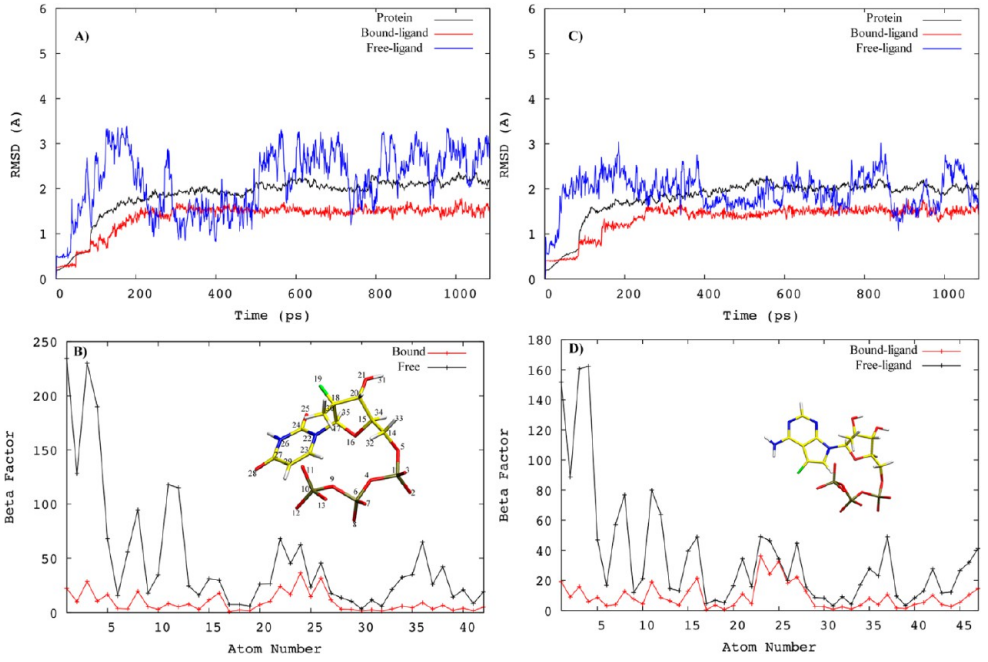


Figure 3. RMSD and atomic fluctuations for two different nucleotide inhibitors 8 and 30. A and B correspond to **8**, while C and D relate to **30**. The two ligands are highly flexible in their unbound form compared to the bound state within the protein.

the compounds we studied here were less rigid due to the absence of the template structure in our simulations. We expect them to be more stabilized by hydrogen bonding with the pairing bases from the template.

For the different genotypes, as their structures were obtained by mutating the initial 1b crystal structure 1GX6, we equilibrated each model for 30 ns before running binding energy analysis for the bound structures of the different nucleotides. It was expected that this would provide a sufficiently long simulation time to relax the modeled structures to their equilibrium conformations. To a large extent, the same pattern of interaction between the active sites with the different ligand structures was observed. All nucleotide structures were more rigid within the active site than in their free form in water. Further details of the atomic protein−ligand interaction for the different cases are presented below.

3.3. MM-PBSA Ranking. Analyzing the literature regarding the development of nucleotide analogues as potent and selective inhibitors for NS5B, one notices that almost none of them were rationally designed based on their energetic interactions with the enzyme. Consequently, to the best of our knowledge, the present study provides a unique piece of information that has not been used before by relating the observed efficacy of the published or patented compounds to their direct interaction with the NS5B active site. For a successful rational drug design protocol, one needs not only to predict the correct binding conformation of a ligand within its binding site; however, the precise calculation of the binding affinity of the ligand is also necessary. Here, we used the MM-PBSA method, introduced by Kollman et al.[16] as a tool to rank the 41 different nucleotide analogues described below. The MM-PBSA has been previously adopted as an accurate approach to predict binding energies at a reasonable computational cost.[25,33,34] As described in previous studies, the most computationally expensive step is the calculation of the entropy using NMODE method.[16,31] Therefore, for each system, we only used 50 snapshots extracted in successive steps from the overall 1000 snapshots representing the MD trajectories (see Materials and Methods).

Most of the experimental IC_{50} values that were found in the literature (see Table 1) were related to the 1b genotype. This was the main reason for choosing this genotype as a reference and a starting point in this study. The work presented here can therefore be considered to be an optimization problem for the 1b case and the data obtained for the remaining genotypes were based on this model. Very little binding data for the studied nucleotides were available for the rest of the genotypes. This makes it hard to test the predictability for our simulation protocol across HCV genotypes. However, the results presented below (shown in Table 2) are quite consistent with the few published experimental values.

Figure 4 compares the predicted binding energies for the tested nucleotides to their experimental values for the 1b case. The IC_{50} values represented by K_i can be directly related to the observed free energy change of binding, ΔG, using the equation:

$$\Delta G = RT \ln K_i \quad (7)$$

where R is the gas constant, $R = 1.987$ cal/(K mol) and T is the absolute temperature. Although the discrepancy in the MM-PBSA calculations for the interactions of the studied inhibitors with NS5B is about 1 kcal/mol for most of the cases, they are in an excellent agreement with the experimental data. The values for very few compounds were found to be off by ∼2 kcal/mol.

Table 2 compares the available experimental binding data for the other genotypes to the energies predicted by the simulation protocol presented here. Once again, the simulation protocol is seen to be successful in predicting the binding data for the NS5B inhibitors.

3.4. Energy Terms Per Residue. One major obstacle to developing a genotypewide NS5B inhibitor is the natural sequence heterogeneity and the emergence of escape mutations emerging during the course of treatment.[9,14] The latter mutations seem to correlate with resistance for many NS5B

dx.doi.org/10.1021/ci4003969 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 1. Comparison between Experimental and Predicted Energies for Genotype 1b

compound no.	predicted affinity kcal/mol	experimental affinity kcal/mol	ref
1	−8.97	−9.23	39,40
2	−9.31	−9.42	39,44
3	−8.64	−7.83	39,45
4	−7.54	−7.41	39
5	−8.60	−8.1	39
6	−7.06	−6.6	39,42
7	−9.33	−9.52	38,39
8	−7.05	−8.59	38,39,46
9	−8.93	−9.26	38,39,41
10	−7.36	−8.51	39
11	−8.82	−8.96	39
12	−8.92	−9.08	39
13	−7.76	−6.23	44
14	−7.54	−7.86	44
15	−7.39	−5.7	44
16	−6.37	−5.5	48
17	−8.37	−5.88	49
18	−7.03	−6.98	49
19	−5.58	−5.5	49
20	−7.97	−8.21	55
21	−8.11	−9.5	48
22	−9.60	−9.19	48
23	−8.28	−8.94	54
24	−7.66	−7.16	48
25	−8.74	−9.93	56
26	−7.03	−6.90	56
27	−8.09	−8.83	39
28	−7.54	−7.86	39
29	−8.89	−9.68	39
30	−9.83	−9.79	39
31	−7.86	−7.38	44
32	−6.80	−7.64	44
33	−6.40	−5.93	44
34	−8.69	−9.57	49
35	−7.26	−7.29	49
36	−8.45	−9.47	50
37	−8.87	−9.47	50
38	−8.154	−8.54	50
39	−7.13	−8.10	49
40	−7.74	−7.15	43
41	−7.98	−8.75	43

inhibitors currently under investigation. Here, we deal with this problem at the atomic level by breaking down the binding energies for the different nucleotide analogues studied into individual protein residue contributions. Our objective is to

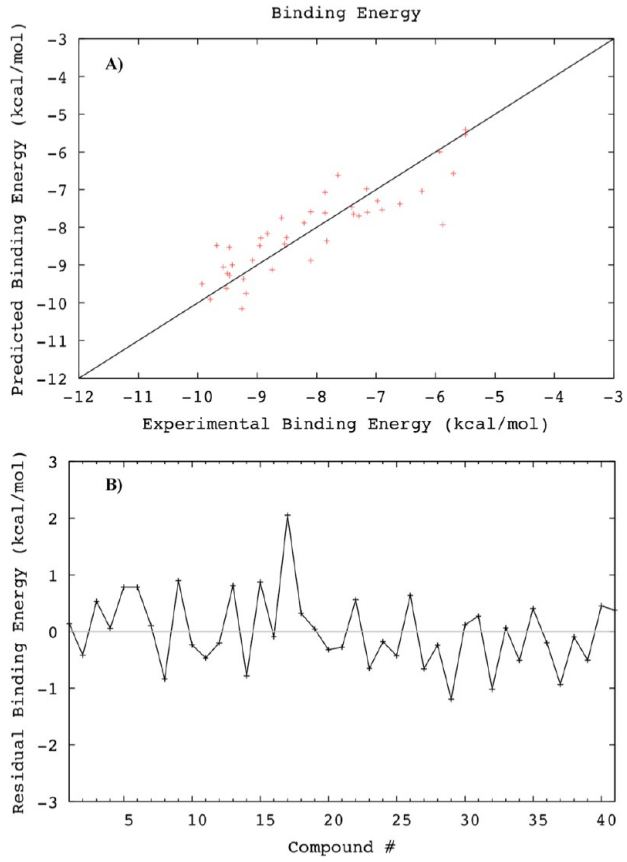


Figure 4. Ranking of the 41 compounds for the 1b case. (A) Correlation between predicted and experimental data. (B) Difference between the two.

examine the influence of these mutations on the binding of the ligand molecules to the active site, both structurally as they affect their binding mode and energetically as they lower the affinity to the protein. Figure 5 shows the binding energy decomposition for the different residues located within the active site of genotypes 1b and 4a including the two Mg^{2+} ions. Interestingly, for all the inhibitors, one can notice a common interaction pattern with the different protein residues. These fingerprints have three main features: (i) the two Mg^{2+} ions interact favorably with the bound nucleotide and are essential for stabilizing the triphosphate tail; (ii) the residues that contribute to the binding energy are 391, 352, 282, 224, 223, 222, 158, 155 and 143; and (iii) residues that are very close to the triphosphate tail have no preferred interaction with the nucleotide due to the electrostatic repulsion of their negatively

Table 2. Comparing Predicted Binding with Available Experimental Data for the Different Genotypes

	1a		2a		3a		4a		
compound no.	exp	pr	exp	pr	exp	pr	exp	pr	ref
1	−8.65	−7.66	NA	−7.20	NA	−6.75	NA	−8.53	38,39
3	−8.37	−7.33	−7.15	−7.09	−8.65	−5.97	−7.86	−8.79	39,45
5	−7.58	−7.18	−7.41	−6.05	NA	−6.79	NA	−8.36	45
7	−9.50	−7.40	NA	−7.24	NA	−6.73	NA	−8.39	39,42
8	NA	−8.61	−7.52	−7.11	−8.45	−7.61	−7.67	−8.79	38,39,46
10	−8.65	−5.92	−8.95	−7.69	NA	−6.85	NA	−8.67	38,39,41
11	−7.92	−7.26	NA	−7.85	NA	−7.99	NA	−8.96	39,41

dx.doi.org/10.1021/ci4003969 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

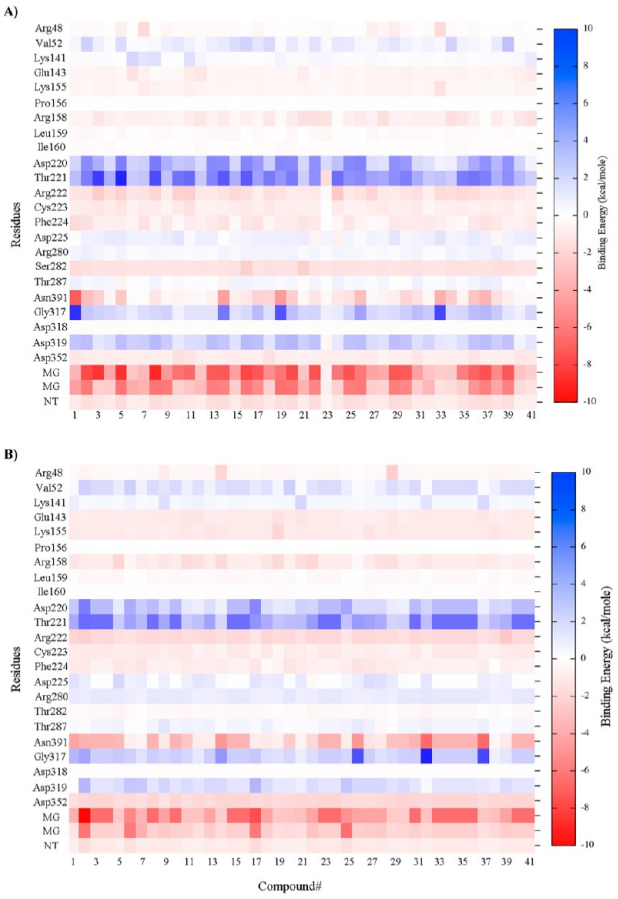


Figure 5. Energy terms for genotypes 1b (A) and 4a (B) revealing a common interaction pattern of all nucleotides with the different protein residues.

charged side chain with the phosphates, which is compensated by the presence of the Mg^{2+} ions.

With very few variations, this interaction pattern is seen to be similar among the different genotypes. For example, comparing the genotype 1b case to that of 2a (see Supporting Information Figure S1), we notice that although the binding trends for all the molecules are similar, most of the compounds showed more favorable interactions except for residues T221 and D220. The same behavior occurs for the genotype 3a and 5a cases (see Figure S1). For most of the residues, the interaction is the same as those for 1b and 2a while the only apparent differences are limited to the two residues T221 and D220.

4. DISCUSSION

4.1. All Nucleotide Inhibitors Have A Unique Binding Mode. Similar to other polymerases, NS5B conserves the right-hand shape with the three conventional subdomains; palm, fingers, and thumb.[35,36] Virus entry introduces a positive-stranded RNA genome in to the hepatocyte where it is replicated in to a negative strand. The latter are then copied in to more positive strands for packaging into the newly synthesized virions. Interestingly, the HCV NS5B can also initiate a de novo polymerase reaction without the need for a primer strand.[37] This reaction usually employs a single nucleotide as a primer, with more preference toward guanosine-5′-triphosphates (**1**) as the activating substrate. Therefore, our structural model resembles the initiation complex for all nucleotides. Nearly 100 crystal structures have been deposited into the PDB for NS5B. Most of these structures describe the binding modes of non-nucleoside-inhibitors within the different allosteric sites of the protein.[35,38] Very few structures explain the binding interaction between RNA molecules and the active site, with no description at all for any nucleotide inhibitor bound to the enzyme. It is noteworthy to emphasize that all the structural models that we described herein are mainly focused on the fine binding details that take place during the initiation and elongation reactions of the NS5B polymerase. This reaction requires a unique orientation of all the chemical ingredients and elements involved in this reaction including the two metal ions, the triphosphate terminal of the nucleotide, the three aspartate residues D220, D318, and D319 and any necessary water molecules (see below). After careful inspection of all NS5B structures deposited in the literature, the only structure that we could find having the above-mentioned constituents to successfully model this complicated binding mode was the one we used in this study, namely 1GX6.

Sequence alignment of representative NS5B structures of different genotypes is shown in Figure 1. As can be clearly seen, the active site is highly conserved. An inhibitor that blocks the active site of the viral polymerase, with the ability to accommodate the limited variations within the catalytic domain among the different genotypes, has the potential of exhibiting a potent and selective pan-genotype inhibition.

The catalytic active site is located within the palm subdomain (see Figure 6). Three preserved aspartate residues D220, D318, and D319 comprise a conserved motif, which typically

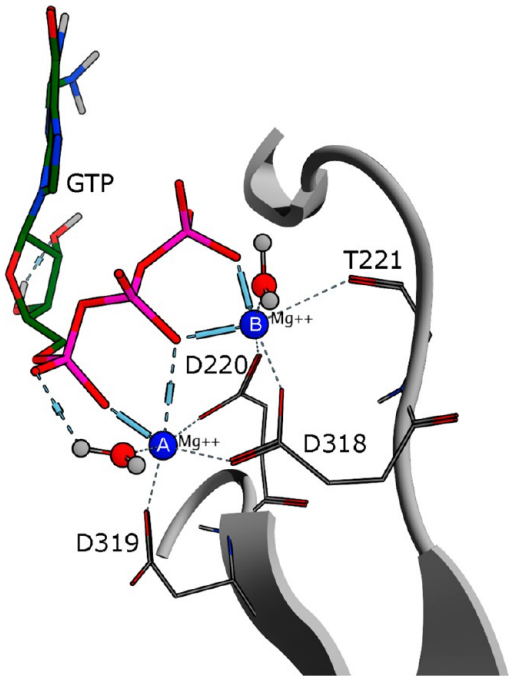


Figure 6. Coordination of the two magnesium ions. Three preserved aspartate residues D220, D318, and D319 comprise a conserved motif, which bind to the two Mg^{2+} ions. The first magnesium ion (ion A) is coordinated by the side chains of the three aspartate residues, a water molecule, the α-phosphate group, and the oxygen connecting this phosphate to the other two phosphate moieties of the nucleotide. On the other hand, the second magnesium ion (ion B) interacts with the side chains of D220 and D318, two phosphate groups, and a water molecule.

dx.doi.org/10.1021/ci4003969 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

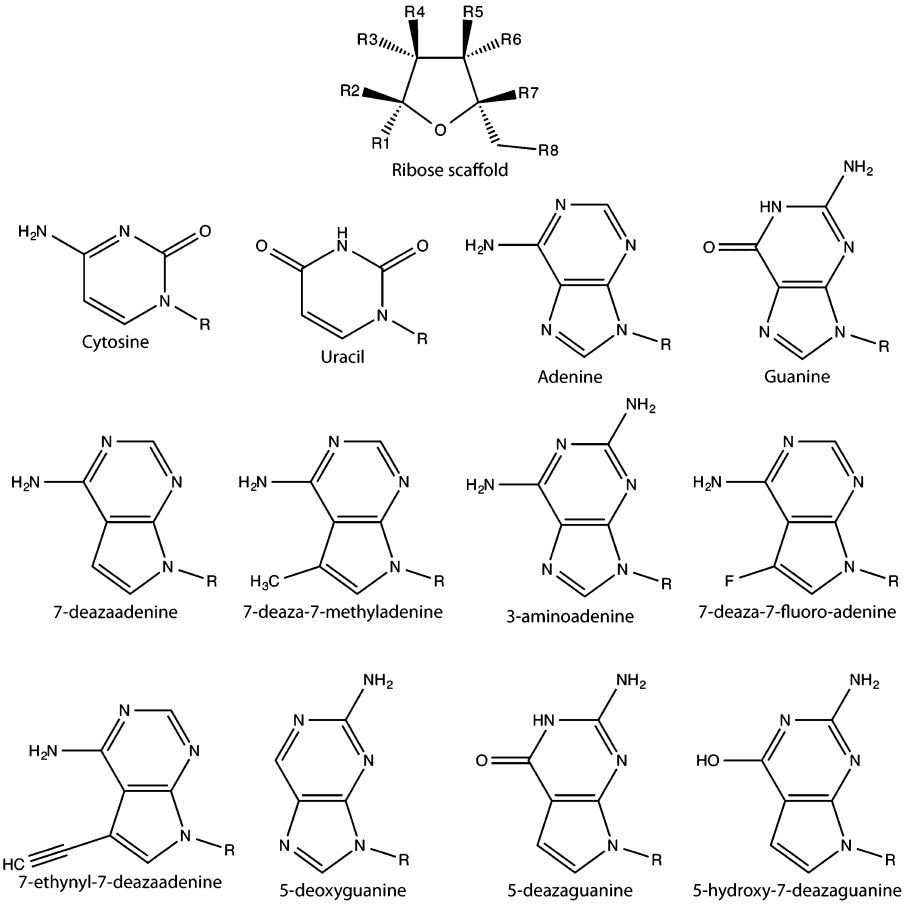


Figure 7. General structure of the tested 41 nucleotides. All tested structures are derivatives of a natural nucleotide which with a sugar scaffold with eight variable points. The substitutions for these points are listed in Table 3 for the 41 nucleotide structures.

synchronizes the nucleotide transfer reaction in other polymerases. The conserved aspartate residues bind to two divalent metal ions (Mn^{2+} or Mg^{2+}). This coordination is critical in attacking the 3′-hydroxyl group of the α-phosphate of the substrate nucleotide triphosphate. However, Mn^{2+} and Mg^{2+} are not equal ions and the choice of which divalent ion to include in our final model was a critical step. In our model we decided to replace the Mn^{2+} ions initially located in the 1GX6 crystal structure with Mg^{2+} for two reasons. First, with no exception, all the experimental studies we found in the literature and from which we obtained values for the binding affinities unanimously use Mg^{2+} in their HCV polymerase assays and not Mn^{2+}.[39−46] Second, in an important early study by Charles Rice's group to investigate HCV viral replication,[47] they tested the requirements of the RNA replication process for Mn^{2+} and Mg^{2+} ions. They found that the production of the full-length replicon RNA increased with an Mg^{2+} concentration up to 5 mM, while an Mn^{2+} concentration greater than 1 mM inhibited the synthesis of the full-length replicon RNA. Therefore, and in order to be in line with these important aspects of RNA polymerase assays and to be able to correlate our findings with the observed experimental binding affinities, we were obliged to use Mg^{2+} instead of Mn^{2+}.

In our model the scissile phosphate is placed between the two Mg^{2+} ions. Metal A (the nucleophilic magnesium ion) is coordinated by the side chains of the three aspartate residues, a water molecule, the α-phosphate group, and the oxygen connecting this phosphate to the other two phosphate moieties of the nucleotide. This connecting oxygen will be cleaved during the nucleotide transfer reaction. The arrangement of the α-phosphate along with the magnesium coordinating residues suggest that the nucleotide is bound in its active form to contact the 3′-hydroxyl group of the primer. On the other hand metal B interacts with the side chains of D220 and D318, two phosphate groups, and a water molecule. In our model, the base is in the correct orientation to form the Watson−Crick hydrogen bonds with the template pairing nucleotide. I160, S282, and G283 seem to be critical in synchronizing the base-paring mechanism by stacking against the base of the substrate nucleotide, positioning it in the right conformation with the template base. On the basis of this conformation of the substrate nucleotide, we built all our models for other nucleotide analogues for the different genotypes.

4.2. Overview of Selected Compounds. The importance of the NS5B enzyme in the life cycle of HCV combined with the availability of many crystal structures both free and bound to several inhibitors made it a promising target for developing directly acting antiviral agents. The aim has been to design novel, potent, and specific interferon/ribavirin-free drugs. As mentioned above, our focus in this study is placed on active site inhibitors (NIs). Analysis for non-nucleotide inhibitors (NNIs) of NS5B will be published elsewhere. Here, we made an effort to include different types of nucleotide inhibitors that cover a wide range of structural diversity and binding affinities (see Figure 7 and Table 3). Below, we discuss the structural variability and activities of the selected compounds that were used in building the models depicted below.

dx.doi.org/10.1021/ci4003969 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 3. Substitutions for the 41 Nucleotide Structures Studied

no.	name	r1	R2	R3	R4	R5	R6	R7	R8[a]
1	2′-C-methylcytidine (NM-107)	cytosine	H	CH_3	OH	OH	H	H	TP
2	2′-C-methylguanosine	guanine	H	CH_3	OH	OH	H	H	TP
3	2′-C-methyladenosine	adenine	H	CH_3	OH	OH	H	H	TP
4	2′-*O*-methylcytidine	cytosine	H	H	OCH_3	OH	H	H	TP
5	2′-*O*-methylguanosine	guanine	H	H	OCH_3	OH	H	H	TP
6	2′-deoxy-2′-fluoro-2′-cytidine	cytosine	H	H	F	OH	H	H	TP
7	2′-deoxy-2′-fluoro-2′-C-methylcytidine	cytosine	H	CH_3	F	OH	H	H	TP
8	2′-deoxy-2′-fluoro-2′-C-methyluridine	uracil	H	CH_3	F	OH	H	H	TP
9	3′-deoxycytidine	cytosine	H	H	OH	H	H	H	TP
10	3′-deoxyguanosine	guanine	H	H	OH	H	H	H	TP
11	4′-azidocytidine (R1479)	cytosine	H	H	OH	OH	H	N3	TP
12	4′-azidoadenosine	adenine	H	H	OH	OH	H	N3	TP
13	3′-deoxy-3′-fluoro-adenosine	adenine	H	H	OH	F	H	H	TP
14	3′-deoxy-3′-fluoro-guanosine	guanine	H	H	OH	F	H	H	TP
15	2′-deoxy-2′-fluoro-adenosine	adenine	H	H	F	OH	H	H	TP
16	2′-deoxycytidine	cytosine	H	H	H	H	OH	H	TP
17	2′-*O*-methyladenosine	adenine	H	H	OCH_3	H	OH	H	TP
18	arabino-guanosine	guanine	H	OH	H	OH	H	H	TP
19	arabino-adenosine	adenine	H	OH	H	OH	H	H	TP
20	2′-C-methyl-2′-deoxy-2′-fluoroguanosine	guanine	H	CH_3	F	OH	H	H	TP
21	2′-deoxy-2′-fluoro-arabinocytidine (RO-8013)	cytosine	H	F	H	OH	H	H	TP
22	4′-azido-arabinocytidine (RO-9187)	cytosine	H	OH	H	OH	H	N3	TP
23	4″-azido-uridine	uracil	H	H	OH	OH	H	N3	TP
24	arabino-cytidine	cytosine	H	OH	H	OH	H	H	TP
25	3′-fluoro-3′-deoxycytidine	cytosine	H	H	OH	F	H	H	TP
26	3′-azido-3′-deoxycytidine	cytosine	H	OH	H	H	H	N3	TP
27	2′-*O*-methyl-7-deazaguanosine	7-deazaguanine	H	H	OCH_3	OH	H	H	TP
28	3′-deoxy-7-deazaadenosine	7-deazaadenine	H	H	OH	H	H	H	TP
29	3′-deoxy-7-deazaguanosine	7-deazaguanine	H	H	OH	H	H	H	TP
30	2′-methyl-7-deaza-7-fluoro-adenosine	7-deaza-7-fluoro-adenine	H	CH_3	OH	OH	H	H	TP
31	2′-C-methyl-3-hydroxyadenosine	guanine	H	CH_3	OH	OH	H	H	TP
32	2′-C-methyl-3-aminoadenosine	3-aminoadenine	H	CH_3	OH	OH	H	H	TP
33	2′-C-methyl-5-deoxyguanonsine	5-deoxyguanine	H	CH_3	OH	OH	H	H	TP
34	2′-C-methyl-7-deazaguanosine	7-deazaguanine	H	CH_3	OH	OH	H	H	TP
35	2′-*O*-methyl-7-deazadenosine	7-deazaadenine	H	CH_3	OH	OH	H	H	TP
36	2′-C-methyl-7-deazadenosine	adenine	H	CH_3	OH	OH	H	H	TP
37	2′-C-methyl-5-hydroxy-7-deazaguanosine	5-hydroxy-7-deazaguanine	H	CH_3	OH	OH	H	H	TP
38	2′-C-methyl-7-deaza-7-methylguanosine	7-deaza-7-methyladenine	H	CH_3	OH	OH	H	H	TP
39	arabino-7-deaza-guanosine	7-deazaguanine	H	OH	H	OH	H	H	TP
40	2′-deoxy-2′-fluoro-2′-C-methylguanosine	adenine	H	CH_3	F	OH	H	H	TP
41	2′-F-2′-C-methyl-7-ethynyl-7-deazaadenosine	7-ethynyl-7-deazaadenine	H	H	F	OH	H	H	TP

[a]R8 is replaced by H in the nucleoside analogue and is replaced by triphosphate (TP) in the active nucleotide.

So far, all NIs described in the literature are analogues of the natural nucleotides.[14] They act as chain terminators and can be classified as either obligate or nonobligate terminators. Obligate chain terminators lack the 3′-hydroxyl group on the sugar and, hence, cannot be extended as no possible attachment to the coming 5′-phosphate is available. Examples of such compounds include compounds **9**, **10**, **16**, and **26** (see Figure 7 and Table 3).[39−41,48] The affinity of these 3′-deoxy-modified nucleotides (IC_{50}) is comparable to those of the natural enzyme substrates in vitro. Conversely, their efficacy in subgenomic replicon assays (EC_{50}s) was very poor compared to the in vitro studies. This discrepancy was attributed to the rate-limiting step of phosphorylating these compounds to their triphosphate active form by the host cell kinases. Since the simulations performed in this study directly correspond to the in vitro studies (the compounds modeled are in their triphosphate form and bound to the active site), for all the compounds tested, we compared only the IC_{50} values that we found in the literature to the outcomes of our structural models (see Results).

Nonobligate NIs possess a 3′-hydroxyl side chain at the sugar location; therefore, hypothetically, they can be incorporated in the resultant positive RNA strand. In spite of this, as is shown below, these compounds typically have modifications to their structures rendering them incapable of forming the phosphodiester bond with the incoming nucleotide. This is mainly due to steric clashes with the essential components of the elongation/translocation process. Accordingly, many of these compounds stop the elongation of the RNA primer and indirectly terminate the process. Nonobligate NIs can be classified into two major subgroups, namely, sugar-modified or base-sugar-modified NIs.

Sugar-modified compounds entail many structural variations and comprise most of the compounds found in the literature (including the obligate chain terminator described above). The

dx.doi.org/10.1021/ci4003969 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

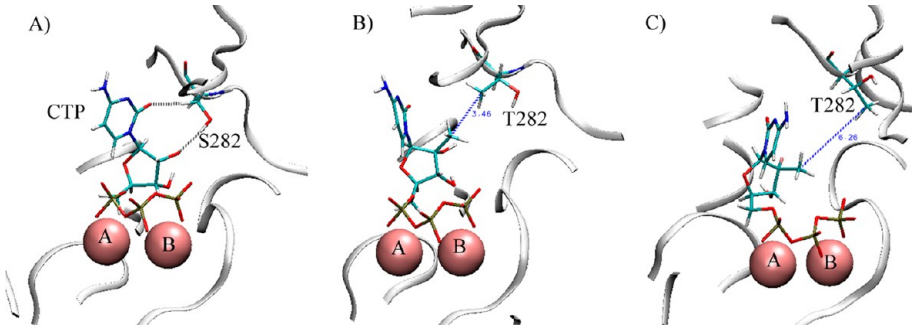


Figure 8. Hydrogen bonding and interaction with the S282T mutation. (A) A bound Cytidine triphosphate (CTP), (B) 2′-C-methylcytidine (NM-107) (**1**), and (C) 2′-deoxy-2′-fluoro-2′-C-methylcytidine (**7**). Structures shown here are the average structures as calculated from the last 10 ns of the MD simulations for the three compounds using the Ptraj utility of AMBER 10.

first group, and most successful so far, are the 2′-fluoronucleotides, namely, compounds **6**, **7**, **8**, **15**, and **20** (see Figure 7 and Table 3).[38−40] The leading HCV NI is GS-7977, an active conformation of two diastereoisomers of PSI-7851. GS-7977 is a prodrug for PSI-7409 (**8**) (see Figure 7 and Table 3).[43] The prodrug was developed to help in phosphorylating the nucleoside (2′-deoxy-2′-fluoro-2′-uridine) to the active triphosphate form. Other nucleotide analogues from this group include the potent compound PSI-6130 (**7**) (see Figure 7 and Table 3), which is currently in phase IIb clinical trials. This group encompasses also 2′-deoxy-2′-fluoro-2′-cytidine (**6**) that is active against NS5B as well as other polymerases. The second set of sugar-modified compounds are 2′-*β*-methylated analogues. The most intensively studied structure in this set is 2′-C-methylcytidine (**1**), which has an acceptable IC_{50} of 3.8 μM with no activity toward DNA-virus polymerases or HIV polymerases. However, this compound has been found to have a low bioavailability profile. Consequently, other alternatives such as 2′-C-methylguanosine (**2**) (see Figure 7 and Table 3) and 2′-C-methyladenosine (**3**) have been explored and showed an enhanced bioavailability for this set of compounds, although showing only moderate activity. The third group of sugar-modified NIs includes 2′-*O*-methyl analogues. Examples of such analogues comprise 2′-*O*-methylcytodine (**4**) (see Figure 7 and Table 3) and 2′-*O*-methyl-guanosine (**5**).[40] The activity of these compounds is similar to that of 3′-deoxy-modified nucleotides. That is, although they possess adequate binding affinities to the polymerase, their cellular activities are very weak due to difficulties in turning their respective nucleosides into the triphosphates. The last group of sugar-modified NIs is the 4′-azid derivatives. Compounds from this class compete with the natural substrate of the enzyme and bind with high potency to the active site. Examples include compounds **11** and **12** (see Figure 7 and Table 3).[40]

Finally, the last group of NIs that we studied involves compounds that have modifications to both the base and sugar moieties. These compounds are more chemically stable compared to sugar-modified analogues. They are more resistant to being phosphorylated by purine nucleoside phosphorylase or deaminated by adenosine deaminase. They are also bioavailable and have better in vivo pharmacokinetic profiles. This group includes compounds **27** to **41** (see Figure 7 and Table 3).[43,44,49,50]

Our modeling procedure confirmed the suitability of all these structures as NS5B inhibitors. All these nucleotide structures were able to serve as substrates for the active site with a high degree of space-filling and complementarity (See Figures 6 and 8). However, all of the studied molecules embodied structural elements that would preclude a further elongation of the RNA strand. Figure 8, for example, illustrates these facts by showing the binding mode of a number of selected inhibitors. Evidently, any modification that is close to the 3′-position is a common feature for most of the active compounds. As discussed above, the hydroxyl group at this position is essential for the elongation reaction and phosphodiester bond formation.

4.3. Effects of Mutations on NIs. The sequence alignment shown in Figure 1 revealed the conserved nature of the active site. In our analysis for the six representative NS5B structures (see Results), only four different mutations occurred within the active site (see Figure 2). A similar, but more thorough procedure, identified these mutations and classified them in terms of genotype recurrence.[51] The most important NS5B mutation is the S282T alternation,[15,52] which usually takes place in genotype 4a. S282 and I160 coordinate stacking between the substrate nucleotide and the pairing template nucleotide. The serine residue is very close to the 2′-hydroxyl of the substrate nucleotide and stabilizes the orientation of the base of the substrate nucleotide through two hydrogen bonds (see Figure 8). Mutating this residue into threonine directly reduces the efficacy of 2′-modified nucleotide derivatives by disturbing the two hydrogen bonds creating a steric clash between the methyl group of the threonine residue and any bulky substituent at this 2′-position. For example, the active triphosphate form of 2′-C-methylcydine (**1**) (see Figure 7 and Table 3) demonstrated a 36-fold lower affinity for the S282T mutant NS5B genotype than the wild type.[52] This also explains the observed resistance against the INX-189 NS5B inhibitor, which has the same sugar modification as 2′-C-methylcydine.[53] The distance between the center of mass of the two methyl groups is less than 3.5 Å (see Figure 8). We did not observe this steric clash in the case of the corresponding 2′-*O*-methylcytosine analogue **5** (see Figure 7 and Table 3) for two reasons (see Figure 8). First, the orientation of the nucleotide methyl group is directly opposite to that of the threonine residue. This means that the two substitutions do not strongly interact with each other compared to the 2′-C-methylcytidine case. Another interesting observation has been made for compound **7** (see Figure 7 and Table 3), which has two substitutions at the 2′-position: a methyl and fluorine moiety. This compound fits well within the active site (see Figure 8) and there are no steric clashes with either of T282 or T287. The distance between the two methyl groups is more than 6 Å (see Figure 8). The fluorine substitution has no

dx.doi.org/10.1021/ci4003969 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

apparent unfavorable conflict with any of the protein residues and enhances the binding affinity of the compound. Other compounds that have alternations at positions different from the 2′-hydrogen and 2′-hydroxyl have no reported resistance associated with S282 mutations. This is obvious since these moieties do not interact directly with this region of the enzyme, however, as shown below, they may be affected by other mutations.

The second observed NS5B mutation is C316N. This alternation commonly takes place in genotype 1b (see Figure 1). According to our model, this mutation is predicted to affect nucleotide analogues that have changes close to the triphosphate tail. Another mutation is G283R, which typically occurs in genotype 1a, and has no direct effect on the substrate nucleotide. Instead, G283 is found to be in a position to stack against the base of the template pairing nucleotide on the other side of S282 and I160. Mutating G283 to arginine can affect base-modified nucleotides.

Finally, according to our model, the reported S96T mutation lies far from the active site. This indicates that introducing a threonine residue at this location will indirectly affect the binding of the coming nucleotide inhibitor. In the wild type structure and in the presence of a template structure, Ser96 usually forms hydrogen bonds with Arg168, stabilizing it to interact with the phosphate of the template.[54] We hypothesize that the threonine residue reduces the available space for the coming nucleotide inhibitor by affecting the orientation of the base of the template pairing nucleotide. This will reduce the preference of the enzyme to bind nucleotide inhibitors and favor the natural nucleotides instead.

Our binding energy data correlate well with the aforementioned interactions (see Table 2) which quantitatively explains the effects of mutations on the binding of the different nucleotide analogues. This is clear from the energy term fingerprints produced through MM-PBSA energy calculations. The effect of the S282T mutation, for example, is apparent in genotype 4a, which is to be expected (see binding energy terms in Figure 5). The contribution of T282 to the binding energy is lower by at least 2 kcal/mol for the 2′-C-methyl nucleotides. The rest of the residues have almost the same energy contribution to the binding energy.

5. CONCLUSION

Over 180 million people are living with chronic HCV infection,[1] a disease that does not always present symptoms until there is a serious liver damage.[7] The virus exists as at least six major different genotypes with each genotype having numerous subtypes.[8,9] A vital need, which still poses a challenging problem, is to develop a potent therapy that would simultaneously inhibit the different genotypes with minimal side effects.

In our current work, we have focused on a detailed understanding the structural basis for nucleotide inhibitors that occupy the active site of the NS5B polymerase. We selected forty-one distinctive nucleotide analogues that cover a wide range of chemical diversity. We then employed 287 different MD simulations to model these structures within the active site of six different genotypes (1a, 1b, 2a, 3a, 4a, 5a, and 6a). All bound structures were ranked using the MM-PBSA method and their interactions were broken down into their atomic level.[16] Our modeling protocol was able to rank and reproduce the binding affinities of nucleotide−mimetic structures with great accuracy. By studying the different genotypes, we were able to explain their differential sensitivity toward various nucleoside inhibitors as well as the effects of the common mutations associated with drug resistance against NS5B nucleotide inhibitors. To the best of our knowledge, this is the first attempt to build such a comprehensive model for the NS5B problem. This study presents a broad investigation focused on understanding the mode of action of these compounds, identifies the key structural elements necessary for their activity, and provides the tools for ranking and developing novel inhibitors of NS5B activity.

■ ASSOCIATED CONTENT

Supporting Information

One figure with three panels describing the same binding energy decomposition analysis shown in Figure 5 for three additional genotypes (2a, 3a, and 5a). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Corresponding Authors Kbarakat@ualberta.ca.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

All of the molecular dynamics simulations and docking experiments were produced using the WestGrid and SHARCNET super computer facilities. The authors would like to thank Dr. Anwar Anwar-Mohamed, Karl Fischer, and Philip Winter for their assistance. Funding for this research was provided through grants from Canada Excellence Research Chairs (CERC), Canadian Institutes of Health Research (CIHR), Alberta Innovates-Health Solutions (AIHS), and the Canadian Association of Gastroenterology (CAG).

■ REFERENCES

(1) Karaman, M. W.; Herrgard, S.; Treiber, D. K.; Gallant, P.; Atteridge, C. E.; Campbell, B. T.; Chan, K. W.; Ciceri, P.; Davis, M. I.; Edeen, P. T.; Faraoni, R.; Floyd, M.; Hunt, J. P.; Lockhart, D. J.; Milanov, Z. V.; Morrison, M. J.; Pallares, G.; Patel, H. K.; Pritchard, S.; Wodicka, L. M.; Zarrinkar, P. P. A quantitative analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **2008**, *26*, 127−132.

(2) Lehman, E. M.; Wilson, M. L. Epidemic hepatitis C virus infection in Egypt: estimates of past incidence and future morbidity and mortality. *J. Viral. Hepat.* **2009**, *16*, 650−658.

(3) Chak, E.; Talal, A. H.; Sherman, K. E.; Schiff, E. R.; Saab, S. Hepatitis C virus infection in USA: an estimate of true prevalence. *Liver Int.* **2011**, *31*, 1090−1101.

(4) He, Y.; Zhang, J.; Zhong, L.; Chen, X.; Liu, H. M.; Wan, L. K.; Wang, H.; Li, H.; Tian, L.; Hu, J. L.; Luo, P.; Wang, L.; Chen, Y.; Liu, T.; Liu, S. L.; Lu, W. B. Prevalence of and risk factors for hepatitis C virus infection among blood donors in Chengdu, China. *J. Med. Virol.* **2011**, *83*, 616−621.

(5) Grakoui, A.; Wychowski, C.; Lin, C.; Feinstone, S. M.; Rice, C. M. Expression and identification of hepatitis C virus polyprotein cleavage products. *J. Virol.* **1993**, *67*, 1385−1395.

(6) Lesburg, C. A.; Cable, M. B.; Ferrari, E.; Hong, Z.; Mannarino, A. F.; Weber, P. C. Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nat. Struct. Biol.* **1999**, *6*, 937−943.

(7) Ohno, T.; Lau, J. Y. The "gold-standard," accuracy, and the current concepts: hepatitis C virus genotype and viremia. *Hepatology* **1996**, *24*, 1312−1315.

(8) Simmonds, P.; Bukh, J.; Combet, C.; Deleage, G.; Enomoto, N.; Feinstone, S.; Halfon, P.; Inchauspe, G.; Kuiken, C.; Maertens, G.;

Mizokami, M.; Murphy, D. G.; Okamoto, H.; Pawlotsky, J. M.; Penin, F.; Sablon, E.; Shin, I. T.; Stuyver, L. J.; Thiel, H. J.; Viazov, S.; Weiner, A. J.; Widell, A. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* **2005**, *42*, 962−973.

(9) Kim, A. Y.; Timm, J. Resistance mechanisms in HCV: from evolution to intervention. *Exp. Rev. Anti-Infect. Ther.* **2008**, *6*, 463−478.

(10) Fried, M. W.; Shiffman, M. L.; Reddy, K. R.; Smith, C.; Marinos, G.; Goncales, F. L., Jr.; Haussinger, D.; Diago, M.; Carosi, G.; Dhumeaux, D.; Craxi, A.; Lin, A.; Hoffman, J.; Yu, J. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N. Engl. J. Med.* **2002**, *347*, 975−982.

(11) Choo, Q. L.; Kuo, G.; Weiner, A. J.; Overby, L. R.; Bradley, D. W.; Houghton, M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* **1989**, *244*, 359−362.

(12) Cisneros-Garza, L. E. [Current advances in hepatitis C treatment]. *Rev. Invest. Salud Publica* **2011**, *53* (Suppl1), S52−60.

(13) Patil, V. M.; Gupta, S. P.; Samanta, S.; Masand, N. Current perspective of HCV NS5B inhibitors: a review. *Curr. Med. Chem.* **2011**, *18*, 5564−5597.

(14) Pockros, P. J. Nucleoside/nucleotide analogue polymerase inhibitors in development. *Clin. Liver Dis.* **2013**, *17*, 105−110.

(15) Kronenberger, B.; Zeuzem, S. New developments in HCV therapy. *J. Viral Hepat.* **2012**, *19* (Suppl 1), 48−51.

(16) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., 3rd Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889−897.

(17) Chang, M. W.; Ayeni, C.; Breuer, S.; Torbett, B. E. Virtual screening for HIV protease inhibitors: a comparison of AutoDock 4 and Vina. *PloS One* **2010**, *5*, e11955.

(18) Vistoli, G.; Pedretti, A.; Mazzolari, A.; Testa, B. Homology modeling and metabolism prediction of human carboxylesterase-2 using docking analyses by GriDock: a parallelized tool based on AutoDock 4.0. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 771−787.

(19) Barakat, K.; Tuszynski, J. Relaxed complex scheme suggests novel inhibitors for the lyase activity of DNA polymerase beta. *J. Mol. Graphics Modell.* **2011**, *29*, 702−716.

(20) Greenman, C.; Stephens, P.; Smith, R.; Dalgliesh, G. L.; Hunter, C.; Bignell, G.; Davies, H.; Teague, J.; Butler, A.; Stevens, C.; Edkins, S.; O'Meara, S.; Vastrik, I.; Schmidt, E. E.; Avis, T.; Barthorpe, S.; Bhamra, G.; Buck, G.; Choudhury, B.; Clements, J.; Cole, J.; Dicks, E.; Forbes, S.; Gray, K.; Halliday, K.; Harrison, R.; Hills, K.; Hinton, J.; Jenkinson, A.; Jones, D.; Menzies, A.; Mironenko, T.; Perry, J.; Raine, K.; Richardson, D.; Shepherd, R.; Small, A.; Tofts, C.; Varian, J.; Webb, T.; West, S.; Widaa, S.; Yates, A.; Cahill, D. P.; Louis, D. N.; Goldstraw, P.; Nicholson, A. G.; Brasseur, F.; Looijenga, L.; Weber, B. L.; Chiew, Y. E.; DeFazio, A.; Greaves, M. F.; Green, A. R.; Campbell, P.; Birney, E.; Easton, D. F.; Chenevix-Trench, G.; Tan, M. H.; Khoo, S. K.; Teh, B. T.; Yuen, S. T.; Leung, S. Y.; Wooster, R.; Futreal, P. A.; Stratton, M. R. Patterns of somatic mutation in human cancer genomes. *Nature* **2007**, *446*, 153−158.

(21) Chemical Computing Group. *Molecular Operating Environment (MOE)*. http://www.chemcomp.com (accessed October 12, 2013)

(22) Kalé, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K. NAMD2: Greater Scalability for Parallel Molecular Dynamics. *J. Comput. Phys.* **1999**, *151*, 283−312.

(23) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712−725.

(24) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

(25) Barakat, K.; Mane, J.; Friesen, D.; Tuszynski, J. Ensemble-based virtual screening reveals dual-inhibitors for the p53-MDM2/MDMX interactions. *J. Mol. Graphics Modell.* **2010**, *28*, 555−568.

(26) Barakat, K. H.; Jordheim, L. P.; Perez-Pineiro, R.; Wishart, D.; Dumontet, C.; Tuszynski, J. A. Virtual screening and biological evaluation of inhibitors targeting the XPA-ERCC1 interaction. *PloS One* **2012**, *7*, e51329.

(27) Jordheim, L. P.; Barakat, K. H.; Heinrich-Balard, L.; Matera, E. L.; Cros-Perrial, E.; Bouledrak, K.; El Sabeh, R.; Perez-Pineiro, R.; Wishart, D. S.; Cohen, R.; Tuszynski, J.; Dumontet, C. Small molecule inhibitors of ERCC1-XPF protein-protein interaction synergize alkylating agents in cancer cells. *Mol. Pharmacol.* **2013**, *84*, 12−24.

(28) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **2007**, *35*, W522−525.

(29) Genheden, S.; Ryde, U. How to obtain statistically converged MM/GBSA results. *J. Comput. Chem.* **2010**, *31*, 837−846.

(30) Legrand-Abravanel, F.; Nicot, F.; Izopet, J. New NS5B polymerase inhibitors for hepatitis C. *Expert Opin. Invest. Drugs* **2010**, *19*, 963−975.

(31) Barakat, K.; Issack, B. B.; Stepanova, M.; Tuszynski, J. Effects of temperature on the p53-DNA binding interactions and their dynamical behavior: comparing the wild type to the R248Q mutant. *PloS One* **2011**, *6*, e27651.

(32) Bennett, M. J.; Barakat, K.; Huzil, J. T.; Tuszynski, J.; Schriemer, D. C. Discovery and characterization of the laulimalide-microtubule binding mode by mass shift perturbation mapping. *Chem. Biol.* **2010**, *17*, 725−734.

(33) Fogolari, F.; Moroni, E.; Wojciechowski, M.; Baginski, M.; Ragona, L.; Molinari, H. MM/PBSA analysis of molecular dynamics simulations of bovine beta-lactoglobulin: free energy gradients in conformational transitions? *Proteins* **2005**, *59*, 91−103.

(34) Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. Validation and use of the MM-PBSA approach for drug discovery. *J. Med. Chem.* **2005**, *48*, 4040−4048.

(35) Ando, I.; Adachi, T.; Ogura, N.; Toyonaga, Y.; Sugimoto, K.; Abe, H.; Kamada, M.; Noguchi, T. Preclinical characterization of JTK-853, a novel nonnucleoside inhibitor of the hepatitis C virus RNA-dependent RNA polymerase. *Antimicrob. Agents Chemother.* **2012**, *56*, 4250−4256.

(36) Scrima, N.; Caillet-Saguy, C.; Ventura, M.; Harrus, D.; Astier-Gin, T.; Bressanelli, S. Two crucial early steps in RNA synthesis by the hepatitis C virus polymerase involve a dual role of residue 405. *J. Virol.* **2012**, *86*, 7107−7117.

(37) Tobi, D.; Bahar, I. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 18908−18913.

(38) Kumar, D. V.; Rai, R.; Brameld, K. A.; Somoza, J. R.; Rajagopalan, R.; Janc, J. W.; Xia, Y. M.; Ton, T. L.; Shaghafi, M. B.; Hu, H.; Lehoux, I.; To, N.; Young, W. B.; Green, M. J. Quinolones as HCV NS5B polymerase inhibitors. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 82−87.

(39) Ali, S.; Leveque, V.; Le Pogam, S.; Ma, H.; Philipp, F.; Inocencio, N.; Smith, M.; Alker, A.; Kang, H.; Najera, I.; Klumpp, K.; Symons, J.; Cammack, N.; Jiang, W. R. Selected replicon variants with low-level in vitro resistance to the hepatitis C virus NS5B polymerase inhibitor PSI-6130 lack cross-resistance with R1479. *Antimicrob. Agents Chemother.* **2008**, *52*, 4356−4369.

(40) Mayhoub, A. S. Hepatitis C RNA-dependent RNA polymerase inhibitors: a review of structure-activity and resistance relationships; different scaffolds and mutations. *Bioorg. Med. Chem.* **2012**, *20*, 3150−3161.

(41) Klumpp, K.; Leveque, V.; Le Pogam, S.; Ma, H.; Jiang, W. R.; Kang, H.; Granycome, C.; Singer, M.; Laxton, C.; Hang, J. Q.; Sarma, K.; Smith, D. B.; Heindl, D.; Hobbs, C. J.; Merrett, J. H.; Symons, J.; Cammack, N.; Martin, J. A.; Devos, R.; Najera, I. The novel nucleoside analog R1479 (4′-azidocytidine) is a potent inhibitor of NS5B-

dependent RNA synthesis and hepatitis C virus replication in cell culture. *J. Biol. Chem.* **2006**, *281*, 3793−3799.
(42) Prakash, T. P.; Prhavc, M.; Eldrup, A. B.; Cook, P. D.; Carroll, S. S.; Olsen, D. B.; Stahlhut, M. W.; Tomassini, J. E.; MacCoss, M.; Galloway, S. M.; Hilliard, C.; Bhat, B. Synthesis and evaluation of S-acyl-2-thioethyl esters of modified nucleoside 5′-monophosphates as inhibitors of hepatitis C virus RNA replication. *J. Med. Chem.* **2005**, *48*, 1199−1210.
(43) Madela, K.; McGuigan, C. Progress in the development of anti-hepatitis C virus nucleoside and nucleotide prodrugs. *Future Med. Chem.* **2012**, *4*, 625−650.
(44) Eldrup, A. B.; Allerson, C. R.; Bennett, C. F.; Bera, S.; Bhat, B.; Bhat, N.; Bosserman, M. R.; Brooks, J.; Burlein, C.; Carroll, S. S.; Cook, P. D.; Getty, K. L.; MacCoss, M.; McMasters, D. R.; Olsen, D. B.; Prakash, T. P.; Prhavc, M.; Song, Q.; Tomassini, J. E.; Xia, J. Structure-activity relationship of purine ribonucleosides for inhibition of hepatitis C virus RNA-dependent RNA polymerase. *J. Med. Chem.* **2004**, *47*, 2283−2295.
(45) Carroll, S. S.; Tomassini, J. E.; Bosserman, M.; Getty, K.; Stahlhut, M. W.; Eldrup, A. B.; Bhat, B.; Hall, D.; Simcoe, A. L.; LaFemina, R.; Rutkowski, C. A.; Wolanski, B.; Yang, Z.; Migliaccio, G.; De Francesco, R.; Kuo, L. C.; MacCoss, M.; Olsen, D. B. Inhibition of hepatitis C virus RNA replication by 2′-modified nucleoside analogs. *J. Biol. Chem.* **2003**, *278*, 11979−11984.
(46) Lam, A. M.; Murakami, E.; Espiritu, C.; Steuer, H. M.; Niu, C.; Keilman, M.; Bao, H.; Zennou, V.; Bourne, N.; Julander, J. G.; Morrey, J. D.; Smee, D. F.; Frick, D. N.; Heck, J. A.; Wang, P.; Nagarathnam, D.; Ross, B. S.; Sofia, M. J.; Otto, M. J.; Furman, P. A. 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. *Antimicrob. Agents Chemother.* **2010**, *54*, 3187−3196.
(47) Hardy, R. W.; Marcotrigiano, J.; Blight, K. J.; Majors, J. E.; Rice, C. M. Hepatitis C virus RNA synthesis in a cell-free system isolated from replicon-containing hepatoma cells. *J. Virol.* **2003**, *77*, 2029−2037.
(48) Klumpp, K.; Kalayanov, G.; Ma, H.; Le Pogam, S.; Leveque, V.; Jiang, W. R.; Inocencio, N.; De Witte, A.; Rajyaguru, S.; Tai, E.; Chanda, S.; Irwin, M. R.; Sund, C.; Winqist, A.; Maltseva, T.; Eriksson, S.; Usova, E.; Smith, M.; Alker, A.; Najera, I.; Cammack, N.; Martin, J. A.; Johansson, N. G.; Smith, D. B. 2′-deoxy-4′-azido nucleoside analogs are highly potent inhibitors of hepatitis C virus replication despite the lack of 2′-alpha-hydroxyl groups. *J. Biol. Chem.* **2008**, *283*, 2167−2175.
(49) Olsen, D. B.; Eldrup, A. B.; Bartholomew, L.; Bhat, B.; Bosserman, M. R.; Ceccacci, A.; Colwell, L. F.; Fay, J. F.; Flores, O. A.; Getty, K. L.; Grobler, J. A.; LaFemina, R. L.; Markel, E. J.; Migliaccio, G.; Prhavc, M.; Stahlhut, M. W.; Tomassini, J. E.; MacCoss, M.; Hazuda, D. J.; Carroll, S. S. A 7-deaza-adenosine analog is a potent and selective inhibitor of hepatitis C virus replication with excellent pharmacokinetic properties. *Antimicrob. Agents Chemother.* **2004**, *48*, 3944−3953.
(50) Eldrup, A. B.; Prhavc, M.; Brooks, J.; Bhat, B.; Prakash, T. P.; Song, Q.; Bera, S.; Bhat, N.; Dande, P.; Cook, P. D.; Bennett, C. F.; Carroll, S. S.; Ball, R. G.; Bosserman, M.; Burlein, C.; Colwell, L. F.; Fay, J. F.; Flores, O. A.; Getty, K.; LaFemina, R. L.; Leone, J.; MacCoss, M.; McMasters, D. R.; Tomassini, J. E.; Von Langen, D.; Wolanski, B.; Olsen, D. B. Structure-activity relationship of heterobase-modified 2′-C-methyl ribonucleosides as inhibitors of hepatitis C virus RNA replication. *J. Med. Chem.* **2004**, *47*, 5284−5297.
(51) Waheed, Y.; Saeed, U.; Anjum, S.; Afzal, M. S.; Ashraf, M. Development of Global Consensus Sequence and Analysis of Highly Conserved Domains of the HCV NS5B Protein. *Hepat. Mon.* **2012**, *12*, e6142.
(52) Dutartre, H.; Bussetta, C.; Boretto, J.; Canard, B. General catalytic deficiency of hepatitis C virus RNA polymerase with an S282T mutation and mutually exclusive resistance towards 2′-modified nucleotide analogues. *Antimicrob. Agents Chemother.* **2006**, *50*, 4161−4169.
(53) Membreno, F. E.; Lawitz, E. J. The HCV NS5B nucleoside and non-nucleoside inhibitors. *Clin. Liver Dis.* **2011**, *15*, 611−626.
(54) Mosley, R. T.; Edwards, T. E.; Murakami, E.; Lam, A. M.; Grice, R. L.; Du, J.; Sofia, M. J.; Furman, P. A.; Otto, M. J. Structure of hepatitis C virus polymerase in complex with primer-template RNA. *J. Virol.* **2012**, *86*, 6503−6511.
(55) Furman, P. A.; Murakami, E.; Niu, C.; Lam, A. M.; Espiritu, C.; Bansal, S.; Bao, H.; Tolstykh, T.; Micolochick Steuer, H.; Keilman, M.; Zennou, V.; Bourne, N.; Veselenak, R. L.; Chang, W.; Ross, B. S.; Du, J.; Otto, M. J.; Sofia, M. J. Activity and the metabolic activation pathway of the potent and selective hepatitis C virus pronucleotide inhibitor PSI-353661. *Antiviral Res.* **2011**, *91*, 120−132.
(56) Shim, J. H.; Hong, Z.; Wu, J. Z. Recent patents on nucleoside and nucleotide inhibitors for HCV. *Recent Pat. Anti-Infect. Drug Discov.* **2006**, *1*, 323−331.