Accounting for Target Flexibility and Water Molecules by Docking to Ensembles of Target Structures: The HCV NS5B Palm Site I Inhibitors Case Study

ARTICLE *in* JOURNAL OF CHEMICAL INFORMATION AND MODELING · AUGUST 2013

Impact Factor: 3.74 · DOI: 10.1021/ci400367m · Source: PubMed

8 AUTHORS, INCLUDING:

Nunzio Iraci
Università degli Studi di Perugia
40 PUBLICATIONS **586** CITATIONS
SEE PROFILE

Giuseppe Manfroni
Università degli Studi di Perugia
42 PUBLICATIONS **422** CITATIONS
SEE PROFILE

Stefano Sabatini
Università degli Studi di Perugia
52 PUBLICATIONS **639** CITATIONS
SEE PROFILE

Oriana Tabarrini
Università degli Studi di Perugia
93 PUBLICATIONS **1,209** CITATIONS
SEE PROFILE

JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

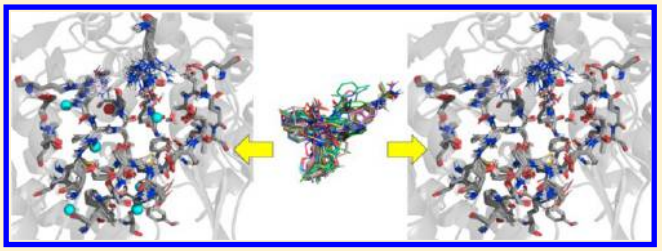
Accounting for Target Flexibility and Water Molecules by Docking to Ensembles of Target Structures: The HCV NS5B Palm Site I Inhibitors Case Study

Maria Letizia Barreca,* Nunzio Iraci,* Giuseppe Manfroni, Rosy Gaetani, Chiara Guercini, Stefano Sabatini, Oriana Tabarrini, and Violetta Cecchetti

Dipartimento di Chimica e Tecnologia del Farmaco, Università degli Studi di Perugia,Via del Liceo 1, 06123 Perugia, Italy

Supporting Information

ABSTRACT: The introduction of new anti-HCV drugs in therapy is an imperative need and is necessary with a view to develop an interferon-free therapy. Thus, the discovery and development of novel small molecule inhibitors of the viral NS5B polymerase represent an exciting area of research for many pharmaceutical companies and academic groups. This study represents a contribution to this field and relies on the identification of the best NS5B model(s) to be used in structure-based computational approaches aimed at identifying novel non-nucleoside inhibitors of one of the protein allosteric sites, namely, palm site I. First, the NS5B inhibitors at palm site I were classified as water-mediated or nonwater-mediated ligands depending on their ability to interact with or displace a specific water molecule. Then, we took advantage of the available X-ray structures of the NS5B/ligand complexes to build different models of protein/water combinations, which were used to investigate the influence on docking studies of solvent sites as well as of the influence of the protein conformations. As the overall trend, we observed improved performance in the docking results of the water-mediated inhibitors by inclusion of explicit water molecules, with an opposite behavior generally happening for the nonwater-mediated inhibitors. The best performing target structures for the two ligand sets were then used for virtual screening simulations of a library containing the known NS5B inhibitors along with related decoys to assess the best performing targets ensembles on the basis of their ability to discriminate active and inactive compounds as well as to generate the correct binding modes. The parallel use of different protein structures/water sets outperformed the use of a single target structure, with the two-protein 3H98/2W-2FVC/7W and 3HKY/NoW-3SKE/NoW models resulting in the best performing ensembles for water-mediated inhibitors and nonwater-mediated inhibitors, respectively. The information gathered from this work confirms the primary role of water molecules and protein flexibility in docking-based studies and can be exploited to aid NS5B-directed HCV drug discovery efforts.



INTRODUCTION

Infection with the hepatitis C virus (HCV), for which no vaccine is available, represents an important health care problem worldwide, being the leading cause of serious chronic liver diseases.[1] HCV is a positive single-stranded RNA virus encoding for structural (E1, E2, C, and p7) and nonstructural proteins (NS2, NS3/4A, NS4B, NS5A, NS5B),[2] which currently are almost all the object of intensive anti-HCV drug discovery programs.

Among the six major genotypes, HCV genotype 1b causes the majority of HCV infections in industrialized countries, and it is the hardest to treat. In particular the standard of care (SOC) treatment (interferon alpha combined with ribavirin) is effective only in about 40−50% of infected patients and is accompanied by numerous side effects and low compliance. Despite intense research efforts, it was only in 2011 that the first two anti-HCV drugs targeting viral proteins (i.e., the NS3/4A protease inhibitors boceprevir and telaprevir) were introduced to the market, and their use in combination with SOC has led to encouraging results for HCV genotype 1 patients. However, the new regimen still has some disadvantages;[3] thus, the pursuit of new anti-HCV drugs remains an exciting and urgent area of interest.

Beside NS3/4A protease, the NS5B RNA-dependent RNA polymerase (RdRp) has been longstandingly recognized as a versatile target to identify antiviral molecules[4] given its essential role in viral replication and substantial structural and functional differences from human enzymes.

NS5B RdRp is a 66 kDa protein of ∼590 amino acids found at the C-terminus of the virus-encoded HCV polyprotein of about 3000 amino acids. Analogously to other polymerases, the NS5B structure resembles a right-hand but it adopts a unique shape, in which the fingers and thumb subdomains interact with RNA and encircle the enzyme active site. NS5B polymerase inhibitors include nucleoside and nucleotide analogues that

Received: June 25, 2013
Published: August 18, 2013

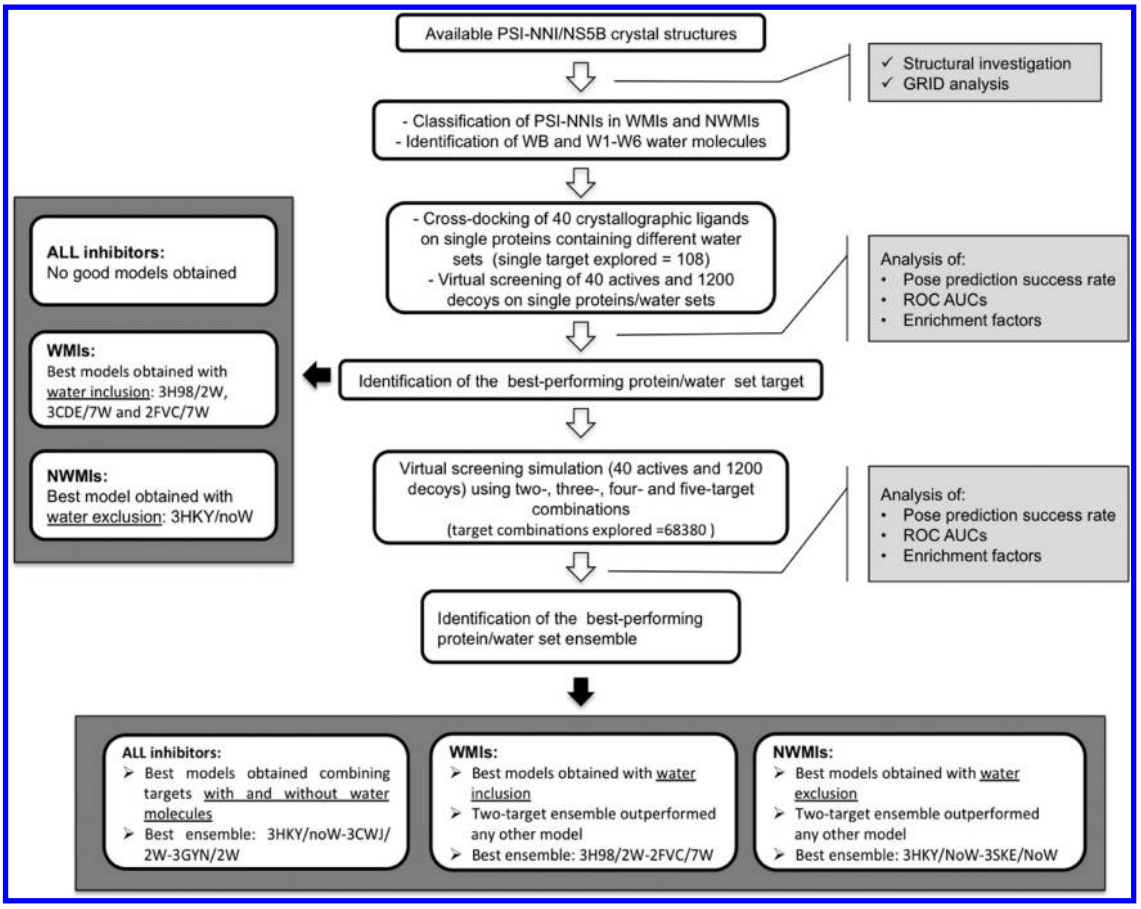


Figure 1. Workflow of our structure-based study. PSI−NNI: palm site I non-nucleoside inhibitor. WMIs: water-mediated inhibitors. NWMIs: nonwater-mediated inhibitors. WB: water bridge.

target the RdRp catalytic site and non-nucleoside inhibitors (NNIs) that target NS5B allosteric sites, provoking protein conformational changes that lower or block enzyme activity.[5]

Despite the tremendous efforts made during the last 15 years by both industrial and academic researchers to discover new anti-NS5B drugs for the treatment of hepatitis C infections,[6−9] the field has suffered from high attrition, with several compounds being pulled from the clinic due to lack of efficacy or unforeseen toxicity, and to date no candidate targeting NS5B has reached the market. Thus, the discovery of new chemical entities able to inhibit NS5B is of great interest and could help the troubled research of anti-HCV drugs.

In this context, our aim has been to focus on the design of small molecule NNIs using rational approaches.[10,11] To date, many crystal structures of NS5B RdRP in complex with different NNIs have been reported in the literature, providing information about location, composition, and conformation of the allosteric binding sites.[6,12] In this work, we have focused on the palm site I (PSI), the most explored NS5B pocket in terms of structurally diverse ligands and crystallographic data.

In particular, we have used structure-based computational techniques to identify the best NS5B model(s) to be used in the discovery of novel PSI−NNIs, paying attention to two important features of the NS5B models: presence and impact of structural water molecules and protein conformation.

Considering that water molecules can be involved in target ligand recognition mainly by mediating hydrogen bonds between the protein and the ligand or by being displaced by the ligand, it is easy to understand why today one of the most important challenges in computer-aided drug discovery is the insight into the role of water molecules in the ligand binding phenomena.[13−33]

Over the past years, several studies reported that inclusion of selected water molecules in the explored targets improves the docking performance and/or accuracy, even though the presence of explicit hydration sites in a protein binding pocket may have different impacts on the outcome of docking-based simulations, mainly depending on the protein conformation and ligand identity.[34−43] Key water positions can be pinpointed by taking advantage of crystallographic information and/or by in silico predictions.[44−59] Many techniques have been developed toward the latter end, but some of them may pay back their quality with computational costs and/or complicate setup, which could make for unsuitable handling of virtual screenings of very large libraries. On the other hand, different methods have been developed to deal with protein flexibility. Among them, the use of ensembles of rigid protein structures has been shown to be useful in accounting also for mutations and the presence of structural water molecules.[34,45,60−68]

In this scenario, our study has a 2-fold objective: (i) evaluation of the inclusion of selected water molecules and different protein conformations on ligand−protein docking simulations at the NS5B polymerase PSI and (ii) identification of the best NS5B template(s) for structure-based discovery of new PSI−NNIs.

The work has been carried out in several steps, summarized in the workflow reported in Figure 1.

RESULTS AND DISCUSSION

Identification of Tightly Bound Water Molecules at the PSI. Up to January 2013, 45 crystal structures of NS5B polymerase (HCV genotype 1b) in complex with PSI−NNIs were published on the RCSB Protein Data Bank (PDB).[69]

Table 1 reports the PDB codes of the different ligand−NS5B complexes and the relative NNI chemotypes, which belong to 10 different chemical families. The structures of the corresponding inhibitors are reported in Figure S1 of the Supporting Information.

Table 1. PSI−NNI/NS5B Crystal Structures (HCV Genotype 1b)

family	NNI chemotypes	PDB ID of PSI−NNI/NS5B complex	references
1	acrylic acids	1YVF, 1Z4U	70,71
2	rhodanines	2AWZ, 2AX0, 2AX1	72
3	benzothiadiazines	2FVC, 2GIQ, 3HHK, 3BSA, 3BSC, 3CDE, 3BR9, 3E51, 3CO9, 3CVK, 3H2L, 3H98, 3GYN, 3IGV	73−82
4	benzothiazines	3CWJ, 3G86, 3H59	83−85
5	benzoisothiazoles dioxide	3D28, 3D5M, 3H5U, 3H5S	86,87
6	proline sulfonamides	2GC8	88
7	acylpyrrolodines	2JC0, 2JC1	89
8	anthranilic acids	2QE2, 2QE5	90
9	benzodiazepines	3GOL, 3CSO, 3GNV, 3GNW, 3HKY	91−94
10	indoles	3SKA, 3SKE, 3SKH, 3U4O, 3U4R, 3UPH, 3UPI, 3TYQ, 3TYV	95−99

Our first goal was to identify which water molecules were bound to the protein tightly enough to be considered as part of the target structure during docking simulations. This study was conducted in several steps, and details of the protocol and parameters used in this work are given in Methods section.

The first step was to thoroughly analyze all the available X-ray structures in order to gain detailed information about ligand−NS5B interactions and bridging water molecules. At this stage, complex 2QE2 was excluded from this analysis because no crystallographic waters were present in the released structure. The 44 crystal structures listed in Table 1 were downloaded from the PDB, processed using the Schrödinger's Protein Preparation Wizard,[100] and then superimposed using as reference protein structure the PDB ID 3TYV, which was solved at high resolution and was co-crystallized with a very potent NNI (compound **45**, Figure S1, Supporting Information).

The set of NS5B residues used for such alignment was defined as all of the amino acids located at a distance ≤5 Å from each ligand. The global set was thus composed by 36 residues: Arg158, Phe193, Tyr195−Pro197, Arg200, Asp225, Thr287, Ser288, Asn291, Thr292, Asn316−Asp319, Ser365−Ser368, Leu384, Arg386, Arg394, Ser407, Gly410, Asn411, Met414, Tyr415, Gln446−Ala450, Phe551, and Gly554−Ser556.

Two alignment types were carried out using the Protein Structure Alignment tool in Maestro.[101] The first alignment was restricted to the backbone atoms, while the second one was expanded to include also the side chains. The structural alignment revealed that the PSI shows similar conformational features in every crystal structure, regardless of the bound ligand, as demonstrated by the root-mean-square deviation (RMSD) values obtained considering just the backbone (average RMSD = 0.35 Å) or the whole residues (average RMSD = 0.64 Å) of the analyzed proteins (Table S1, Supporting Information). Although this rough analysis suggested that PSI presented similar conformations across the multiple crystal structures, we decided to perform a per residue analysis to highlight whether any residue could undergo substantial conformational changes depending on the bound ligand. Recently, Bressanelli et al. assessed the conformational diversity in the 78 genotype 1b NS5B structures available in the PDB up to October 2010. This interesting study was however focused on the influence that ligand binding had on the overall protein conformational/functional changes and not on the binding site itself.[102]

The backbone-aligned complexes represented the starting point for our analysis. Series of RMSD values were calculated for each of the 36 binding site residues previously described, using a different crystallographic conformation at a time to obtain a RMSD matrix. From this analysis, it resulted that although the overall binding site architecture is fairly conserved across the structures, many residues can get into quite different conformations (Figure S2, Supporting Information).

The ligand−protein interactions were then analyzed by taking into account data literature, by visual inspection, and by using the Ligand Interaction Diagram tool of Maestro GUI,[101] which allows for generation of and display of 2D diagrams of the interactions between ligands and their binding sites. It is worth noting that although the literature data report detailed structural analysis of the ligand−protein contacts for many different inhibitors,[12,103,104] to the best of our knowledge, a classification of PSI−NNIs according to the receptor−ligand−water interactions has not been reported so far.

The data literature and in silico analysis showed that the chemical features shared by diverse chemotypes of PSI−NNIs are (i) a hydrophobic group occupying a pocket mainly defined by residues Met414 and Gly410; (ii) a second hydrophobic group deeply buried in a lipophilic pocket shaped by Tyr448, Cys366, Pro197, Tyr415, Arg200, and Leu384; and (iii) a hydrogen bond (HB) acceptor group, which establishes a direct interaction with the backbone amide NH of Tyr448 (Figure 2). In addition, in most of the crystallographic complexes (33 out of 44), the same acceptor group also interacts with a water molecule, which bridges the inhibitor to the Gln446 and/or Gly449 residue. This water molecule will be henceforth indicated as WB (water bridge). The inhibitors interacting with WB will be referred as WMIs (water-mediated inhibitors), while the inhibitors which do not make this water bridge interaction (11 out of 44 structures) will be defined as NWMIs (nonwater-mediated inhibitors) (Table S2, Supporting Information).

As evident from their bound crystal structures, most of the NWMIs are able to displace WB by placing an acceptor group in the position usually occupied by the WB oxygen and establishing a direct HB interaction with Gly449 (3HKY, 2JC1, 2JC0, 2GC8, 3GNW) and/or Gln446 (3UPI, 2GC8). On the contrary, 1YVF, 1Z4U, 2QE5, and 3GOL do not show this additional HB but just a single HB interaction with Tyr448.

Figure 2 illustrates one crystallographic complex as representative of each PSI−NNI chemical family, according to their classification as WMIs or NWMIs, which are chosen taking into account the best compromise between PDB

dx.doi.org/10.1021/ci400367m | J. Chem. Inf. Model. 2014, 54, 481−497

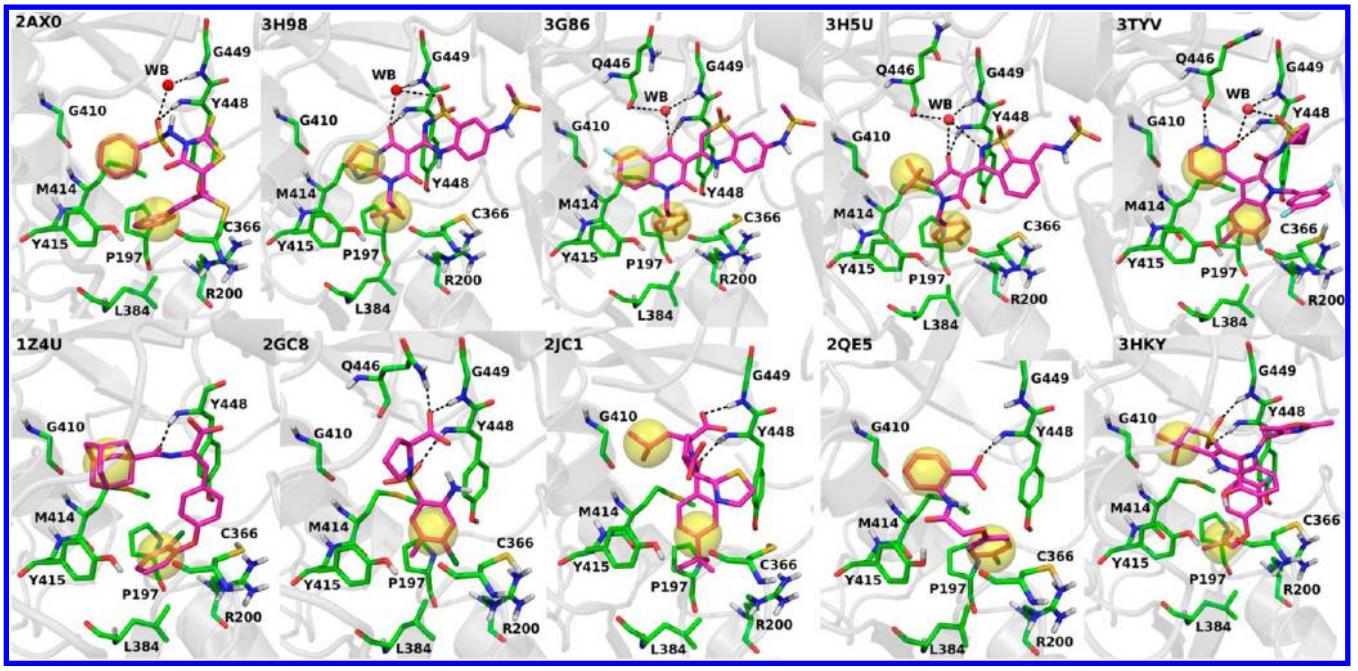


Figure 2. Crystallographic position of WMIs (up) and NWMIs (down) at the HCV polymerase allosteric binding PSI. Each PSI−NNI chemotype is represented by one PDB structure. For WMIs: rhodanines, 2AXO; benzothiadiazines, 3H98; benzothiazines, 3G86; benzoisothiazoles dioxide, 3H5U; and indoles, 3TYV. For NWMIs: acrylic acids, 1Z4U; proline sulfonamides, 2GC8; acylpyrrolodines, 2JC1; anthranilic acids, 2QE5; and benzodiazepines, 3HKY. The NS5B residues involved in interactions with the ligand. Hydrogen bonds are represented as black dashed lines, while the yellow spheres illustrate the two common hydrophobic regions.

Table 2. Statistical and GRID Analysis of Selected Crystal Water Molecules in the PSI−NNI/NS5B Complexes

water molecule	all complexes[a] (44 structures)	mean energy[b]	WMIs complexes[a] (33 structures)	mean energy[b]	NWMIs complexes[a] (11 structures)	mean energy[b]
W1	95.5%	−13.64	100%	−13.70	81.8%	−13.52
W2	90.9%	−11.20	100%	−11.33	63.6%	−10.64
W3	86.4%	−11.47	87.9%	−11.46	81.8%	−11.56
W4	81.8%	−11.99	87.9%	−12.10	63.6%	−11.57
W5	72.7%	−10.56	78.8%	−10.58	54.6%	−10.40
W6	72.7%	−10.54	75.8%	−10.44	63.6%	−10.87
WB	75.0%	−6.72	100%	−6.72	-	-

[a]Frequency of the specified crystal water molecule in the different structural sets, expressed as percentage of presence. [b]Values expressed in kcal/mol and calculated as the average of the interaction binding energies between the GRID water probe, whose position overlapped with the specified crystal water, and the NS5B proteins.

resolution and potency of the co-crystallized inhibitor (Figure S1, Supporting Information).

In the next step, a consensus approach was carried out by evaluating in silico the binding energies of experimentally identified water locations with the final aim to select high affinity water molecules conserved across the NS5B/PSI−NNI complexes.

The X-ray solved water molecules in the vicinity (≤4 Å) of the "superligand", defined by the superposition of all the ligands in the aligned PSI−NNI/NS5B complexes, were selected and extracted from each of the 44 PSI−NNI/NS5B complexes previously described, thus obtaining 44 sets containing a different number of water molecules, ranging from 3 to 30 depending on the analyzed complex (Table S3, Supporting Information). By these criteria, not only the water molecules that interacted with the ligand (e.g., WB) were included but also those that could be important in defining the volume and shape of the allosteric PSI. The 44 waterless NS5B proteins were then imported into GRID, a software meant to estimate the interaction energy between a chemical probe and the investigated region of the protein.[52,105] Because it has been shown that crystallographic water molecules with high specificity exhibit favorable GRID energies, we decided to apply this fast in silico approach to energetically characterize our water sets. Toward this objective, the GRID water probe was used to identify high affinity water locations whose coordinates overlapped with the previously extracted crystallographic water molecules.

A water molecule was considered as a key hydration site when it was highly recurrent (frequency > 50%) across the different crystal structures and recognized by GRID as a tight binder (energy values ≤ −10 kcal/mol). Six water molecules (W1−W6) satisfied the required criteria (Table S4, Supporting Information), and Table 2 summarizes the analysis results considering the whole set of PSI−NNI/NS5B complexes (44 structures), the WMIs set (33 structures), and the NWMIs set (11 structures).

W1−W6 made strong interactions with NS5B PSI residues (Figure 3 and Table S5, Supporting Information) showing at least two hydrogen bonds with nonwaters atoms. In most of the

dx.doi.org/10.1021/ci400367m | J. Chem. Inf. Model. 2014, 54, 481−497

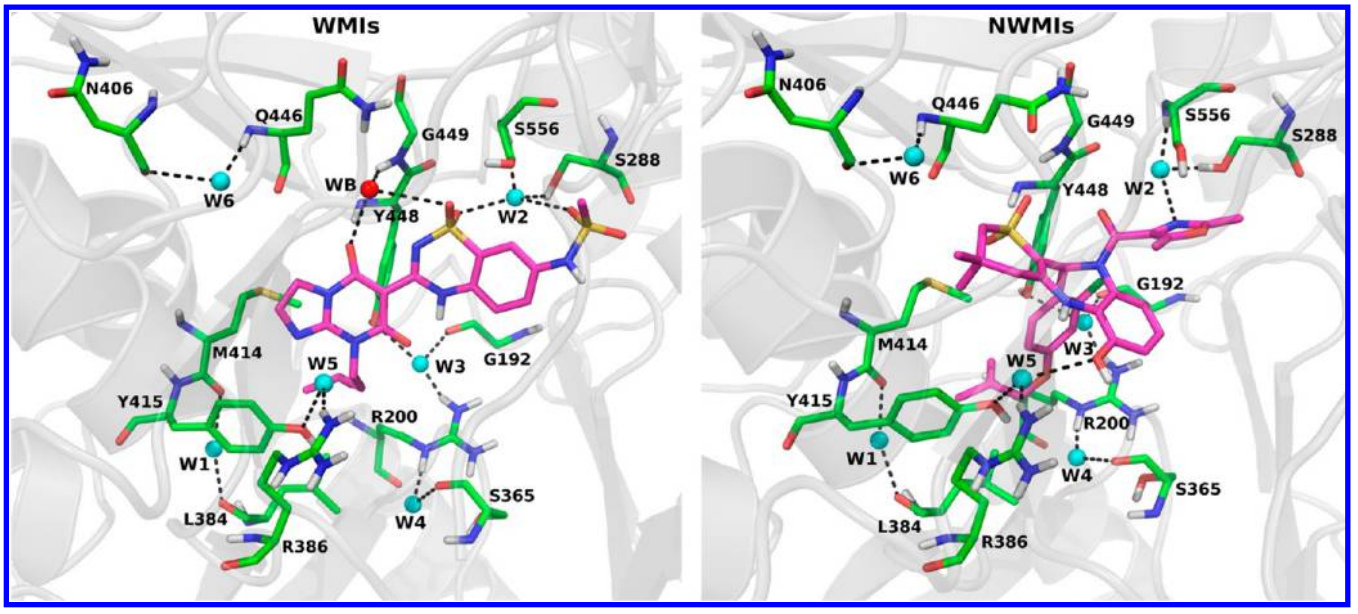


Figure 3. Conserved water molecules (W1−W6, cyan) and WB (red) at the NS5B allosteric PSI of WMI (left, PDB ID 3H98) and NWMI (right, PDB ID 3HKY) complexes. NS5B residues establishing hydrogen bond interactions with the water molecules are shown and colored in green. WB, W2, and W5 interact with both the protein and ligands. Ligand−water molecules interaction through WB, W2, and W5 are highlighted as well.

Table 3. Water Sets Used in Cross-Docking Simulations

water set	description	No. of target proteins	PDB codes
Set-NoW	no water molecules	40	1YVF, 1Z4U, 2FVC, 2GIQ, 3HHK, 3BSA, 3BSC, 3BR9, 3CDE, 3E51, 3CO9, 3CVK, 3H2L, 3H98, 3GYN, 3IGV, 3CWJ, 3G86, 3H59, 3D28, 3D5M, 3H5U, 3H5S, 2GC8, 2JC0, 2JC1, 2QE5, 3GOL, 3CSO, 3GNV, 3GNW, 3HKY, 3SKA, 3SKE, 3SKH, 3U4O, 3U4R, 3UPH, 3UPI, 3TYV
Set-2W	WB, W2	29	1Z4U, 2FVC, 2GIQ, 3BSA, 3BSC, 3BR9, 3CDE, 3E51, 3CO9, 3CVK, 3H2L, 3H98, 3HHK, 3GYN, 3IGV, 3CWJ, 3G86, 3H59, 3D28, 3D5M, 3H5U, 3H5S, 3SKA, 3SKE, 3SKH, 3U4O, 3U4R, 3UPH, 3TYV
Set-6W	W1−W6	22	2FVC, 2GC8, 2GIQ, 2JC0, 2JC1, 3BR9, 3CDE, 3E51, 3G86, 3H2L, 3H5S, 3H5U, 3H59, 3H98, 3HKY, 3SKA, 3SKE, 3TYV, 3U4O, 3U4R, 3UPH, 3UPI
Set-7W	WB, W1−W6	17	2FVC, 2GIQ, 3BR9, 3CDE, 3E51, 3G86, 3H2L, 3H5S, 3H5U, 3H59, 3H98, 3SKA, 3SKE, 3TYV, 3U4O, 3U4R, 3UPH

complexes, W2 established a three-point interaction between the PSI−NNIs and the residues Ser556 and/or Ser288 (Figure 3). On the other hand, in seven complexes (i.e., 3HKY, 3GNW, 2AWZ, 2AX1, 2AX0, 2JC0, and 2JC1), W5 mediated a hydrogen bond interaction between the ligands and the residues Tyr415 and Arg386 (Figure 3). The remaining four identified structural waters (i.e., W1, W3, W4, and W6) do not interact with the ligands directly; however, these molecules fill the PSI space and thus might have an important role in defining volume and shape of this binding site.

The highly conserved nature of the identified structural waters was also supported by the experimentally solved "apo" form polymerase structure (PDB ID 1C2P),[106] which showed the presence of all but one (W5) of the above-mentioned water molecules.

The GRID analysis also confirmed that WB could not be considered as a structural molecule because in all of the structures where this molecule was present (i.e., the WMIs set), it was predicted to have low interaction energy (mean energy value = −6.72 kcal/mol) (Table 2). This result is in agreement with the experimental observation that WB can be displaced by some NWMIs. On the other hand, literature data report that the hydrogen bond networks involving WMIs and WB seem to be important contributing factors to gain high NS5B binding affinity.[75]

Effect of Water in Cross-Docking Calculations. We next evaluated the effect of including WB and the selected water molecules W1−W6 on docking accuracy and performance. Two different tasks were performed: (1) cross-docking experiments, where each target model was evaluated for its ability to reproduce the crystallographic position of the known PSI−NNIs and (2) docking-based virtual screening simulations to assess which target model(s) should be used to identify bioactives within a library.

For this purpose, the complexes where the ligands were covalently bound to the protein (i.e., 2AWZ, 2AX0, 2AX1, and 3TYQ) were excluded from the docking calculations, which were performed using four different water sets (Table 3): (i) Set-NoW: absence of any water molecules, (ii) Set-2W: the two main waters (i.e., WB and W2) seen to mediate the interactions of various ligands with the protein were included, (iii) Set-6W: all of the six identified structural waters (W1−W6) were considered, and (iv) Set-7W: both W1−W6 and WB were kept. It is however noteworthy that the number of target proteins may vary depending on the used set because the presence of water molecules was variable across the complexes (Table S4, Supporting Information). In total, 108 NS5B target proteins structures were obtained. In Set-2W, Set-6W, and Set-7W models, the crystallographic water molecules in the binding site were considered as part of the target structure and thus kept for the docking calculations.

dx.doi.org/10.1021/ci400367m | J. Chem. Inf. Model. 2014, 54, 481−497

Table 4. Summary of Pose Prediction Results of PSI−NNIs Using the Set-NoW[a]

ligands	docking performance[a]	3BR9	3CDE	3CWJ	3E51	3G86	3GNV	3H98	3HKY	3SKE	3UPI
ALL	$RMSD_{avg}$	3.3	3.0	2.8	3.2	2.8	3.6	3.3	3.4	4.1	3.3
	% success	47.5	52.5	52.5	47.5	52.5	45.0	55.0	42.5	35.0	40.0
WMIs	$RMSD_{avg}$	2.5	1.8	2.0	2.4	2.2	4.0	2.8	3.7	4.9	3.5
	% success	62.1	72.4	72.4	65.5	72.4	37.9	72.4	31.0	24.1	31.0
NWMIs	$RMSD_{avg}$	5.4	6.2	4.8	5.5	4.5	2.6	4.7	2.3	2.1	2.5
	% success	9.1	0	0	0	9.1	63.6	9.1	72.7	63.6	63.6

[a]$RMSD_{avg}$ (expressed in Å) is the average RMSD calculated for each protein considering the three sets of ligands, i.e., ALL, WM, and NWM inhibitors. The corresponding success rate (% success) in correctly predicting crystal ligand conformations (i.e., RMSD values ≤2 Å) is also reported.

Each ligand was docked into each protein using the Glide Standard Precision docking (SP),[107] retrieving the best three poses, which in turn were refined and rescored using the Extra Precision mode (XP), finally collecting only the top-ranking solution for each ligand. We decided to use Glide scoring functions because in our own experience[10,108] as well as in literature[109−112] they have been shown to be a good compromise between speed and accuracy for several targets, thus making them appropriate for the use in the first steps of a large scale virtual screening pipeline.

First, each docking model was evaluated for its ability to reproduce the experimental binding modes determined by calculating the RMSD between the XP top-scoring poses and the respective experimental structure, excluding nonpolar hydrogens.

As commonly accepted in the literature, a ligand binding mode was considered to be correctly predicted if the RMSD was ≤2.0 Å. For each simulation, the cross-docking results are presented in matrix form where each cell reports the RMSD of a ligand (in the row) docked into a structure (in the column) (Tables S6−S9, Supporting Information). Moreover, statistical analysis were performed for each simulation to calculate the average RMSD values ($RMSD_{avg}$) for each protein and the "docking performance", expressed as the percentage of correctly predicted ligand conformations (i.e., RMSD < 2 Å) for all 40 ligands as well as for ligands previously classified as WMIs or NWMIs (Tables S10−S13,Supporting Information).

Second, the best performing protein−water sets, chosen considering a success rate of pose prediction ≥70% at 2 Å ligand RMSD threshold and/or detailed analysis of the docking results, were used for virtual screening of a library containing the known co-crystallized NS5B inhibitors and 1200 related decoys (see Methods). Although the activity of the decoys against NS5B was unknown, they were selected on the basis of their dissimilarity in respect to the 40 active PSI−NNIs (see Methods); as such, the set of decoy molecules should not possess the key binding requirements for NS5B (Figure 2) and was thus presumed to be inactive against the viral protein. Once all the virtual screening calculations were performed, the best scoring poses were selected for each ligand and target structure and mining capabilities were evaluated in terms of areas under the curve (AUC) of receiver operating characteristic (ROC) curves[113] for WMIs, NWMIs, and all ligands; the greater the AUC, the better the ability of the model to discriminate active over inactive compounds, with an AUC of less than 0.5 considered random and a maximum possible value of 1. Enrichment factors were also calculated at 2%, 3%, and 4% of the virtual screening progress (see Methods). This metric is commonly employed to evaluate how well the docking procedure is able to identify active inhibitors as high ranking with respect to the inactive ligands.

Set-NoW. The set of proteins used for this study consists of all of the 40 structures. Cross-docking of all ligands in the absence of any water molecule (Tables S5 and Table S10, Supporting Information) gave the best performance for 3H98 protein, which was able to reproduce the 55% of the correct binding conformations with an average RMSD of 3.3 Å (Table 4).

Beside the performance of 3H98, 3CDE, 3CWJ, and 3G86 showed a percentage of reproducibility of 52.5% coupled to an average RMSD of 3.0, 2.8, and 2.8 Å, respectively.

Restricting the analysis to the WMIs, the best results were obtained again for the 3CDE, 3CWJ, 3G86, and 3H98 targets, all obtaining a success rate at a 2 Å ligand RMSD threshold of 72.4% and an average RMSD of 1.8, 2.0, 2.2, and 2.8 Å, respectively. Moreover, 3BR9 and 3E51 showed a percentage of reproducibility of 62.1% and 65.5%, respectively ($RMSD_{avg}$ of 2.5 and 2.4 Å, respectively). In virtual screening, the best performing crystal structures were found to be 3CDE and 3CWJ, with ROC AUC of 0.96 and 0.97 and enrichment factors of 0.76 and 0.72 at 4% of the virtual screening, respectively (Table 5).

Table 5. ROC AUCs and Enrichment Factors (EF) Obtained for the Best Set-NoW Proteins

ligands	ranking efficiency[a]	3CDE	3CWJ	3G86	3H98	3HKY	3SKE
ALL	AUC	0.84	0.91	0.85	0.86	0.86	0.75
	$EF^{2\%}$	0.37	0.42	0.32	0.35	0.27	0.27
	$EF^{3\%}$	0.50	0.50	0.42	0.42	0.37	0.32
	$EF^{4\%}$	0.55	0.57	0.50	0.52	0.45	0.32
WMIs	AUC	0.96	0.97	0.93	0.96	0.82	0.70
	$EF^{2\%}$	0.52	0.59	0.41	0.48	0.31	0.24
	$EF^{3\%}$	0.65	0.62	0.55	0.59	0.34	0.24
	$EF^{4\%}$	0.76	0.72	0.62	0.72	0.38	0.24
NWMIs	AUC	0.52	0.73	0.63	0.58	0.94	0.88
	$EF^{2\%}$	0	0.18	0.18	0.48	0.45	0.45
	$EF^{3\%}$	0	0.27	0.18	0.59	0.64	0.54
	$EF^{4\%}$	0	0.27	0.18	0.62	0.64	0.54

[a]Enrichment factors were calculated at 2% ($EF^{2\%}$), 3% ($EF^{3\%}$), and 4% ($EF^{4\%}$) of the virtual screening progress.

Finally, considering only the NWMIs, the 3HKY protein performed the best, giving a percentage of reproducibility of 72.7%, an average RMSD of 2.3 Å, ROC AUC of 0.94, and enrichment factors of 0.64 at 3% of the library screen.

Set-2W. The first set of water molecules used in the docking simulations was represented by the two water molecules named WB and W2. Because the simultaneous presence of WB and W2 was observed in 29 out of 40 complexes, 29 target proteins were used for our cross-docking studies (Tables S6 and Table

dx.doi.org/10.1021/ci400367m | J. Chem. Inf. Model. 2014, 54, 481−497

Table 6. Summary of Pose Prediction Results of PSI−NNIs Using Set-2W

ligands	docking performance[a]	1Z4U	2GIQ	3BSA	3CDE	3CWJ	3H98
ALL	$RMSD_{avg}$	3.1	2.8	3.3	2.6	3.1	NC[b]
	% success	37.5	55.0	60.0	55.0	57.5	62.5
WMIs	$RMSD_{avg}$	3.1	1.9	2.3	1.6	2.0	1.6
	% success	34.5	75.9	79.3	75.9	75.9	86.2
NWMIs	$RMSD_{avg}$	3.0	5.3	6.2	5.6	6.6	NC[b]
	% success	45.5	0	9.1	0	9.1	0

[a]$RMSD_{avg}$ (expressed in Å) is the average RMSD calculated for each protein considering the three sets of ligands, i.e., ALL, WM, and NWM inhibitors. The corresponding success rate (% success) in correctly predicting crystal ligand conformations (i.e., RMSD values ≤2 Å) is also reported. [b]NC = Not calculated. $RMSD_{avg}$ was calculated only when at least one docking pose per ligand was obtained.

S11, Supporting Information), and the best results obtained are summarized in Table 6.

Considering all ligands, cross-docking with the inclusion of two water molecules gave the best results using 3H98 as the target protein, which showed a docking performance of 62.5%. Moreover, 2GIQ, 3BSA, 3CDE, and 3CWJ gave fairly good results because these proteins demonstrated a percentage of reproducibility greater than 55% and an average RMSD of 2.8, 3.3, 2.6, and 3.1 Å, respectively.

Considering the WMIs, 3H98 protein gave an impressive 86.2% with an average RMSD of 1.6 Å. It is worth noting that for this class of inhibitors the model also gave a great discrimination between actives and decoys having a ROC AUC of 0.99 and enrichment factors of 0.59, 0.79, and 0.93 at 2%, 3%, and 4% of the virtual screening progress, respectively. Other structures that performed well included 2GIQ, 3BSA, 3CDE, and 3CWJ, which were able to reproduce more than 75% of correct binding poses with an average RMSD of 1.9, 2.3, 1.6, and 2.0 Å, respectively. These models also provided good performances in library screening simulation (Table 7).

Table 7. ROC AUCs and Enrichment Factors (EF) Obtained for the Best Set-2W Proteins

ligands	ranking efficiency[a]	2GIQ	3BSA	3CDE	3CWJ	3H98	3BR9
ALL	AUC	0.88	0.85	0.86	0.90	0.89	0.89
	$EF^{2\%}$	0.40	0.42	0.40	0.42	0.43	0.45
	$EF^{3\%}$	0.45	0.55	0.55	0.55	0.61	0.57
	$EF^{4\%}$	0.60	0.60	0.60	0.65	0.74	0.62
WMIs	AUC	0.99	0.95	0.98	0.99	0.99	0.96
	$EF^{2\%}$	0.55	0.59	0.55	0.59	0.59	0.62
	$EF^{3\%}$	0.62	0.69	0.72	0.76	0.79	0.76
	$EF^{4\%}$	0.79	0.76	0.76	0.86	0.93	0.76
NWMIs	AUC	0.59	0.60	0.55	0.65	0.59	0.68
	$EF^{2\%}$	0.09	0.18	0.18	0.18	0.20	0.27
	$EF^{3\%}$	0.09	0.18	0.18	0.18	0.30	0.27
	$EF^{4\%}$	0.18	0.18	0.27	0.18	0.30	0.27

[a]Enrichment factors were calculated at 2% ($EF^{2\%}$), 3% ($EF^{3\%}$) and 4% ($EF^{4\%}$) of the virtual screening progress.

Finally, if we analyze the results obtained considering only the NWMIs, as we would expect considering the presence of the water WB that should be displaced by this class of inhibitors, no target was able to nicely reproduce the experimental ligand conformations. Anyway, the best results were obtained for 1Z4U, which showed a reproducibility of 45.5% coupled with an $RMSD_{avg}$ of 3.0 Å.

Set-6W. The water set included in these docking simulations was defined by the six molecules named W1−W6, which are present in 22 out of 40 complexes. This means that 22 structures have been used for this cross-docking study (Tables S7 and S12, Supporting Information), and a summary of the most interesting results is shown in Table 8.

Table 8. Summary of Pose Prediction Results of PSI−NNIs using Set-6W

ligands	docking performance[a]	3BR9	3CDE	3G86	3HKY
ALL	$RMSD_{avg}$	2.8	2.7	2.7	NC[b]
	% success	60.0	57.5	62.5	45.0
WMIs	$RMSD_{avg}$	2.1	1.7	1.8	3.8
	% success	72.4	75.9	75.9	34.5
NWMIs	$RMSD_{avg}$	4.4	5.1	5.1	NC[b]
	% success	36.4	9.1	27.3	72.7

[a]$RMSD_{avg}$ (expressed in Å) is the average RMSD calculated for each protein considering the three sets of ligands, i.e., all, WM, and NWM inhibitors. The corresponding percentage of correctly predicted ligand conformations, (i.e., RMSD values ≤2 Å) is also reported. [b]NC = Not calculated. $RMSD_{avg}$ was calculated only when at least one docking pose per ligand was obtained.

Cross-docking simulations with the inclusion of six water molecules gave the best results, considering all ligands, when using the 3G86 protein target (average RMSD was 2.7 Å, and the docking poses with an RMSD < 2 Å represented the 62.5% of the total). The 3BR9 and 3CDE proteins also gave quite good results because they were able to correctly pose more than 55% of the ligands with $RMSD_{avg}$ of 2.8 and 2.7 Å, respectively. When analyzing the value obtained considering only the WMIs, the best results were obtained for 3CDE and 3G86, which both produced a docking performance of 75.9% and an average RMSD of 1.7 and 1.8 Å, respectively. Furthermore, 3BR9 demonstrated a reproducibility of 72.4% coupled with an $RMSD_{avg}$ of 2.1 Å. In virtual screening, again the best performing structure resulted to be 3G86, showing a 0.98 AUC and enrichment factors of 0.59, 0.79, and 0.83 at 2%, 3%, and 4% of the virtual screening progress, respectively (Table 9).

Considering only NWMIs, once more the best results were obtained for the 3HKY protein, which showed a success rate at a 2 Å ligand RMSD threshold of 72.7% and good performance with a ROC AUC of 0.89. Anyway, this structure only showed an enrichment factor of 0.4 at 4% of the library screening progress.

Set-7W. W1−W6 and WB were simultaneously present in 17 out of 40 complexes; 17 protein templates were thus used in this study (Table S8 and Table S13 in Supporting Information). Cross-docking simulation with the inclusion of seven water molecules gave the best results relative to all

dx.doi.org/10.1021/ci400367m | J. Chem. Inf. Model. 2014, 54, 481−497

Table 9. ROC AUCs and Enrichment Factors (EF) Obtained for the Best Set-6W Proteins

ligands	ranking efficiency[a]	3BR9	3CDE	3G86	3HKY	3SKE
ALL	AUC	0.86	0.80	0.87	0.77	0.75
	$EF^{2\%}$	0.45	0.40	0.42	0.36	0.30
	$EF^{3\%}$	0.57	0.47	0.57	0.41	0.37
	$EF^{4\%}$	0.60	0.55	0.62	0.46	0.40
WMIs	AUC	0.95	0.93	0.98	0.72	0.73
	$EF^{2\%}$	0.62	0.55	0.59	0.38	0.31
	$EF^{3\%}$	0.79	0.65	0.79	0.45	0.41
	$EF^{4\%}$	0.83	0.76	0.83	0.48	0.48
NWMIs	AUC	0.59	0.45	0.55	0.89	0.76
	$EF^{2\%}$	0.00	0.00	0.09	0.30	0.27
	$EF^{3\%}$	0.00	0.00	0.18	0.40	0.27
	$EF^{4\%}$	0.00	0.00	0.18	0.40	0.36

[a]Enrichment factors were calculated at 2% ($EF^{2\%}$), 3% ($EF^{3\%}$) and 4% ($EF^{4\%}$) of the virtual screening progress.

ligands for 2FVC, and 3CDE and 3H98 proteins (Table 10), which were all able to reproduce the 60% of the correct binding pose with average $RMSD_{avg}$ of 2.9 Å and 3.0 Å, respectively.

Moreover, 2GIQ, 3BR9, 3G86, and 3H59 models showed a docking performance greater than 50% coupled to an average RMSD of 3.2 Å.

The 3CDE model resulted to be the best target protein when considering the WMIs only; this template showed a good docking performance of 82.8% with an average $RMSD_{avg}$ of 1.6 Å. Interestingly, results were also obtained with 2FVC, 2GIQ, 3BR9, 3G86, 3H59, and 3H98 proteins, which were all able to reproduce more than 70% of the correct binding poses with $RMSD_{avg}$ values ranging from 1.7 to 2.3 Å. The best model in virtual screening was 2FVC, which gave good discrimination having ROC AUC of 0.98 and enrichment factors of 0.69, 0.79, and 0.86 at 2%, 3%, and 4% of the virtual screening progress, respectively (Table 11).

As expected, the docking results obtained considering only the NWMIs were not satisfactory because no protein was able to significantly reproduce the correct binding poses.

Highlights from the Docking-Based Simulations. Taken together, the obtained results showed a different impact of the selected water molecules on the docking performance depending on the analyzed ligand set.

As overall trend, we observed improved performance in the docking results of WMIs by inclusion of explicit water molecules, with the opposite generally happening for the NWMIs. This was especially true when dealing with WB that, in the used docking protocol, cannot be displaced by NWMIs claiming the space occupied by this water molecule.

In particular, for the WMIs a success rate at 2 Å ligand RMSD threshold of 86.2% and 82.8% was obtained when using 3H98/2W and 3CDE/7W templates, respectively. Of note, the percent of correctly predicted poses for the WMIs was coupled with impressive RMSD values of 1.8 and 1.6 Å. In virtual screening, the best performing structures resulted to be 3H98/2W and 2FVC/7W, showing ROC AUCs of 0.98 and 0.99, respectively. Furthermore, the late model showed the best enrichment at 2% of the database screen, while the opposite was observed at 4% (3H98/2W: $EF^{2\%}$= 0.59 and $EF^{4\%}$= 0.93; 2FVC/7W: $EF^{2\%}$= 0.69 and $EF^{4\%}$= 0.86). Finally, the two models performed equally at 3% of the library screening.

A close comparison among the different water−/protein sets highlighted that the presence of explicit water molecules always improved the docking performance of WMIs. In this class of inhibitors, WB and W2 have an essential role because the ligands were guided toward their experimental binding poses by establishing hydrogen bond interaction with these two water molecules. Moreover, the formation of water-mediated hydrogen bond interaction provided higher docking scores for many of the ligands, allowing an easier ranking of active over inactive ones. On the other hand, in several targets, Set-7W docking simulations showed superior performance over Set-2W both in terms of average RMSD values and success rate. Thus, in these proteins, also the non-interacting water molecules probably facilitated the ligand to achieve a pose closer to the experimental one by presumably better defining the steric shape of the binding pocket.

On the contrary, the best docking performance of NWMIs in terms of pose prediction and enrichment was obtained when using a target protein lacking explicit water molecules at the PSI. In fact, although both 3HKY/NoW and 3HKY/6W models gave a success rate of pose prediction at 2 Å ligand RMSD threshold of 72.7%, AUCs and enrichment factors analysis clearly showed that the model with no W1−W6 displayed improved ability in discriminating active from inactive ligands (3HKY/NoW: AUC = 0.94 and $EF^{4\%}$ = 0.64; 3HKY/6W: AUC = 0.89 and $EF^{4\%}$ = 0.40). Moreover, as the overall trend, we observed that the Set-NoW models performed better than the Set-6W models for this class of inhibitors. Thus, when dealing with NWMIs, the six water molecules W1−W6 probably impose steric constraints that hamper the scaffold(s) to correctly fit in the allosteric site.

When considering all of the 40 ligands, the best overall performance was shown by the target 3G86/6W, which gave a percentage of correctly predicted poses of 62.5% coupled with RMSD values of 2.7 Å. Anyway, this result was obviously biased by the numeric disparity between the WMIs and NWMIs set. Moreover, the analysis of AUCs and enrichment factors for the

Table 10. Summary of Pose Prediction Results of PSI−NNIs Using the Set-7W

ligands	docking performance[a]	2FVC	2GIQ	3BR9	3CDE	3G86	3H59	3H98
ALL	$RMSD_{avg}$	2.9	NC[b]	3.1	3.0	3.3	3.2	3.0
	% success	60.0	52.5	52.5	60.0	55.0	55.0	60.0
WMIs	$RMSD_{avg}$	1.8	2.2	1.7	1.6	2.1	2.3	1.8
	% success	79.3	72.4	72.4	82.8	72.4	75.9	79.3
NWMIs	$RMSD_{avg}$	5.7	NC[b]	6.7	7.2	6.6	5.7	6.2
	% success	9.1	0	0	0	9.1	0	9.1

[a]$RMSD_{avg}$ (expressed in Å) is the average RMSD calculated for each protein considering the three sets of ligands, i.e., ALL, WM, and NWM inhibitors. The corresponding percentage of correctly predicted ligand conformations, (i.e., RMSD values ≤2 Å) is also reported. [b]NC = Not calculated. $RMSD_{avg}$ was calculated only when at least one docking pose per ligand was obtained.

dx.doi.org/10.1021/ci400367m | J. Chem. Inf. Model. 2014, 54, 481−497

Table 11. ROC AUCs and Enrichment Factors (EF) Obtained for the Best Set-7W Proteins

ligands	ranking efficiency[a]	2FVC	2GIQ	3BR9	3CDE	3G86	3H59	3H98
ALL	AUC	0.87	0.87	0.85	0.82	0.84	0.81	0.88
	$EF^{2\%}$	0.50	0.76	0.45	0.45	0.50	0.37	0.42
	$EF^{3\%}$	0.57	0.59	0.55	0.60	0.57	0.45	0.45
	$EF^{4\%}$	0.62	0.59	0.57	0.60	0.62	0.47	0.57
WMIs	AUC	0.97	0.97	0.98	0.94	0.98	0.93	0.98
	$EF^{2\%}$	0.69	0.62	0.62	0.62	0.69	0.52	0.59
	$EF^{3\%}$	0.79	0.76	0.72	0.79	0.76	0.62	0.59
	$EF^{4\%}$	0.86	0.76	0.79	0.83	0.83	0.65	0.79
NWMIs	AUC	0.60	0.57	0.48	0.48	0.44	0.46	0.54
	$EF^{2\%}$	0.00	0.10	0.00	0.00	0.09	0.00	0..00
	$EF^{3\%}$	0.00	0.10	0.09	0.09	0.18	0.09	0.09
	$EF^{4\%}$	0.00	0.10	0.09	0.09	0.27	0.18	0.09

[a]Enrichment factors were calculated at 2% ($EF^{2\%}$), 3% ($EF^{3\%}$), and 4% ($EF^{4\%}$) of the virtual screening progress.

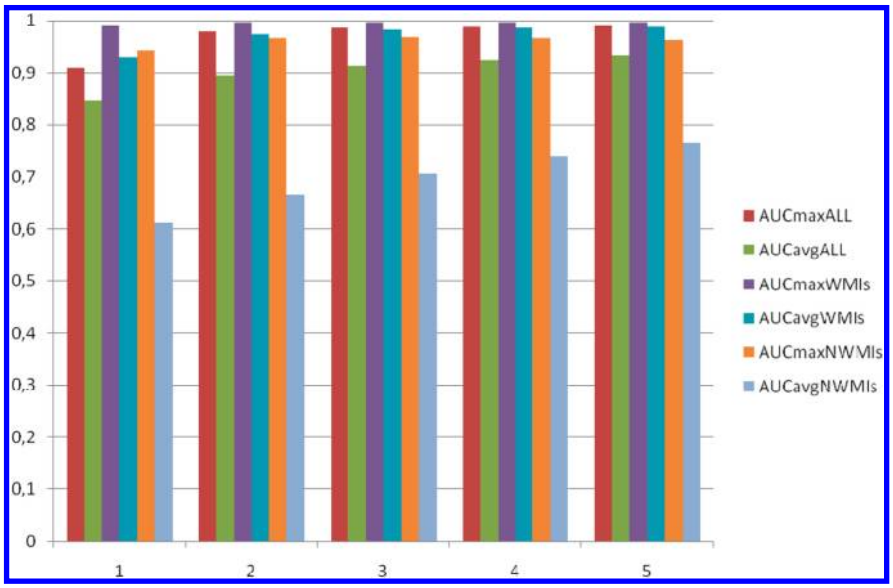


Figure 4. Mining performance obtained using ensembles of one, two, three, four, and five target structures. AUCmax is maximum AUC obtained for each set of targets, while AUCavg is the average of the AUCs for each target set.

whole set of ligands clearly showed that no one single model was able to provide satisfying overall performance in virtual screening.

Docking results also gave some clues about the influence of conformational flexibility of the PSI. Indeed, even though structural alignment revealed that the studied binding site showed a similar architecture across the multiple complexes analyzed, the analysis of the cross-docking results for the Set-NoW (Table 4 and Table S10, Supporting Information) clearly and as expected showed two straight facts. (i) WMIs and NWMIs were preferentially recognized by different conformations. (ii) Even small movements of some residue side chains could influence the correct binding mode for specific chemical classes of PSI−NNIs. For instance, residues like Ser556 and Leu384 participate closer to the ligand binding phenomena, and their conformational changes (Figure S2, Supporting Information) resulted in a deep influence on the docking results. In particular, according to the inhibitor and the protein analyzed, the side chain of Ser556 may adopt a conformation that did not allow the formation of the required hydrogen bond interaction(s) between the ligand and the receptor and in some cases could even create steric hindrance (Figure S3, Supporting Information). Leu384 seems to be another important residue to binding site flexibility and thus to docking performance. This amino acid is important in defining one of the two hydrophobic pocket occupied by PSI−NNIs, and in some docking simulations, its side chain conformation probably hampered the regeneration of the experimental PSI−NNI binding pose due to steric clashes with the ligand (Figure S3, Supporting Information). Of note, residue Arg158, which is the residue whose conformation varies the most across the crystal structure (Figure S2, Supporting Information), does not show direct interaction with the PSI−NNIs in any of the complex structures, and thus, its flexibility did not show direct influence on the docking results.

Evaluation of Target Ensembles. As already mentioned, docking compounds to ensembles of targets may greatly improve the mining capabilities of virtual screening procedures, enabling them to account for protein flexibility, mutations, or, as in this case, the presence of water molecules in the binding site. Anyway, when dealing with large numbers of compounds, it is important to select the appropriate receptors to use.[114] The selection of an appropriate ensemble of target structures may indeed enable the development of virtual screening protocols to apply in the early stage identification of bioactive molecules. Taking advantage of the impressive speed that the modern

dx.doi.org/10.1021/ci400367m | J. Chem. Inf. Model. 2014, 54, 481−497

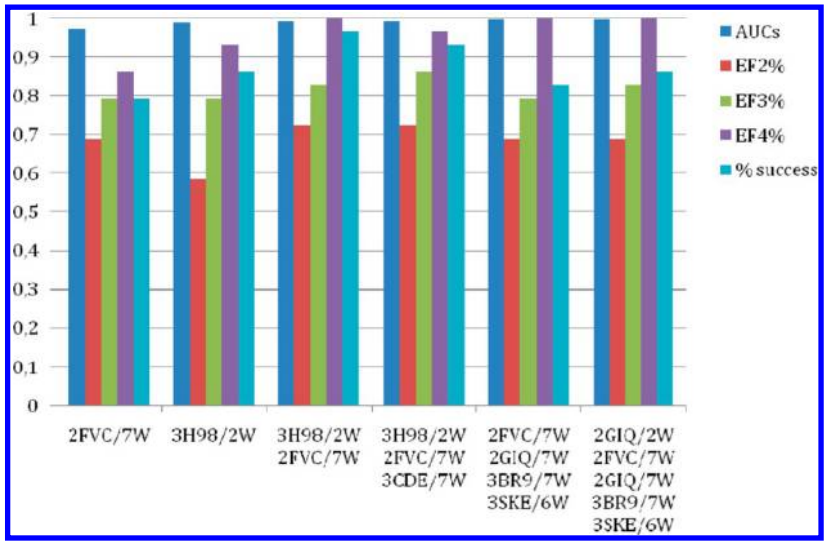


Figure 5. AUCs and enrichment factors for the best performing one-, two-, three-, four-, and five-target ensemble combinations in mining WMIs. % success is scaled by a 0.01 factor for plotting convenience.

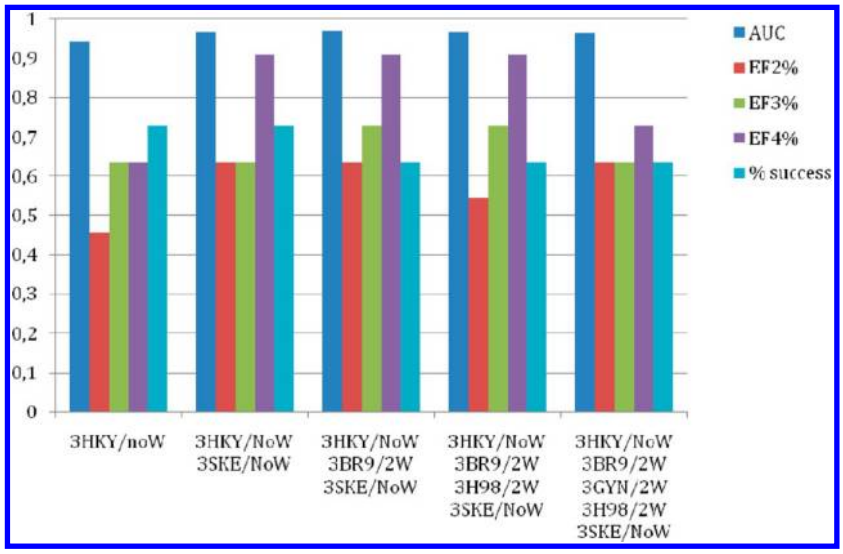


Figure 6. AUCs and enrichment factors for the best performing one-, two-, three-, four-, and five-target ensemble combinations in retrieving NWMIs. % success is scaled by a 0.01 factor for plotting convenience.

docking algorithms are embedded with and the advancements in computational resources, the number of candidates from millions of compounds might be efficiently narrowed down in order to select candidates to further analyze with more rigorous techniques.

To evaluate the NS5B target ensembles to be used in the docking-based virtual screening procedure, 1200 decoy molecules enriched with the 40 known PSI−NNIs (30 decoys for each active) were docked against ensembles of up to five proteins. This study was carried out on the 25 best performing single proteins−water sets, chosen considering a success rate in the pose predictions ≥70% at a 2 Å ligand RMSD threshold and/or detailed analysis of the docking results (see Methods and Tables S10−S13, Supporting Information).

Once all the docking calculations were performed, mining capabilities were evaluated in terms of ROC AUC for the 68,405 target combinations. For each ligand, only the best scoring pose was selected regardless of the target protein. The scored poses were then ranked, and overall AUCs, WMIs AUCs, and NWMIs AUCs were calculated. To assess the ensemble performance, enrichment factors were calculated at 2%, 3%, and 4% of the virtual screening progress as another metric to identify the best performing target combinations. Furthermore, because it is all but obvious that the top-ranked pose corresponds to the real bound pose,[112] we took advantage of the knowledge of the experimental position of the 40 ligands to calculate for each ensemble the success rate of pose prediction at a 2 Å ligand RMSD threshold.

As expected, the use of more than one structure resulted in an improvement of the mining performance (Figure 4). Generally, the more target structures used the better was the average performance, particularly when dealing with ALL and NWM ligands.

Anyway, the performance boost for ALL, WM, and NWM ligands was considerable going from one target structure to two, while a further increase in the number of structures led to weaker performance improvements or in some cases even to a slight performance worsening, thus suggesting the use of two or

dx.doi.org/10.1021/ci400367m | J. Chem. Inf. Model. 2014, 54, 481−497

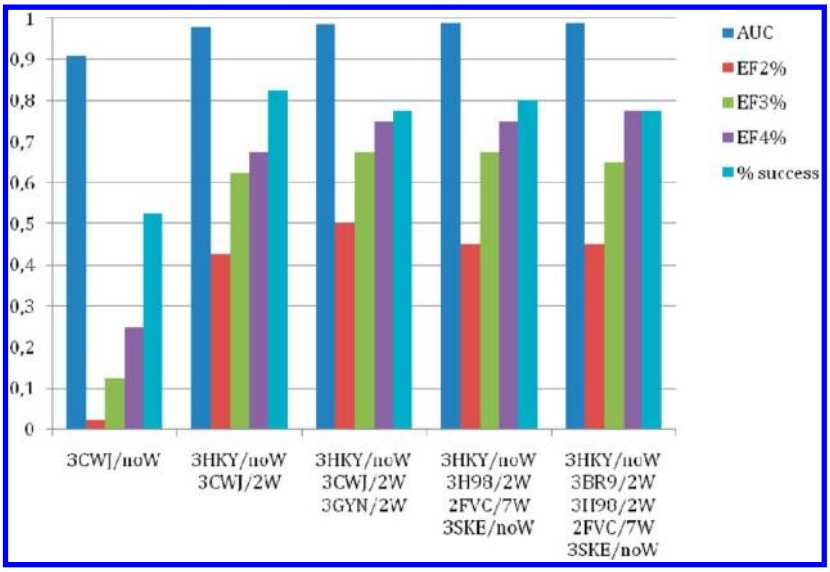


Figure 7. AUCs and enrichment factors for the best performing one-, two-, three-, four-, and five-target ensemble combinations in mining ALL ligands. % success is scaled by a 0.01 factor for plotting convenience.

three structures for early stage big scale virtual screening studies. Moreover, the AUCs analysis once again confirmed the importance of the targets selection because maximum AUCs always outperformed over average AUCs, even when the latter ones were obtained by using more target structures.

The importance of the water sets inclusion is evident looking at the best performing combinations of protein structures−water sets used to mine WM and NWM ligands. Indeed, the ensembles that perform the best for WM ligands always included waters in the target structures, while the ones that perform the best for NWM ligands included at least two structures with no water molecules.

We have then deeply analyzed the ROC AUCs and the enrichment factors for the five best performing (in terms of ROC AUC) one-, two-, three-, four-, and five-target ensemble combinations for each set of ligands, i.e., ALL, WM, and NWM inhibitors.

Considering WMIs, the increase in target structures in the ensemble did not result in a massive growth of the AUCs (Figure 5 and Table S14, Supporting Information).

Anyway, using a two-target ensemble 3H98/2W−2FVC/7W instead of a single structure, the virtual screening procedure improved the success rate up to 0.97% and mined all of the ligand compounds in the first 4% of the screening advancement and more of the half of them in the first 2% ($EF^{2\%}$ = 0.72).

As in the WMIs case, AUCs for NWM ligands mining did not improve significantly with the increase in target structures (Figure 6 and Table S15, Supporting Information), but once again, a substantial increase in the enrichment factors at 2% and 4% was observed when using the two-target ensemble 3HKY/NoW−3SKE/NoW. Furthermore, the rate of success was maintained at 73% in the two-member ensemble, while a further increase in the number of target structures did bring a worse value.

It is worth noting that, although both WM and NWM ligands were efficiently retrieved by two-target ensembles, increasing in the number of structures tended to increase the retrieving of NWM and WM inhibitors simultaneously (Tables S14 and S15, Supporting Information)

Indeed, considering all PSI−NNIs, an increase of the number of structures improved the mining of ligands regardless of them being WM or NWM, with the best performances obtained using ensembles composed of three, four, or five target structures and different water sets (Figure 7).

Target 3HKY with no water molecules was always present in the best performing structure ensembles, and this result was not surprisingly considering that it was the best template for NWM ligands retrieving. Analogously, 2FVC/7W and 3H98/2W were also included in the best three- and four-structure ensembles, and they were already found to be the best single targets in WMIs retrieving. Although the two-target ensembles found to be the best in search of WMIs or NWMIs showed very good performance in retrieving their own ligand classes, the 3HKY/NoW−3BR9/2W−3SKE/NoW target combination could probably result in a very good compromise between computing speed and identification rate of WM- and NWM-PSI−NNIs.

3. CONCLUSIONS

In this work, analysis of cross-docking simulations and virtual screening studies using different water sets in the PSI of HCV NS5B target structures allowed several general conclusions to be drawn.

In a focused structure-based drug design scenario, where the type of ligand (i.e., WMIs or NWMIs in this project) to deal with is known a priori, an insight into the role played by water molecules in the binding site and the use of different protein conformations might permit us to restrict the number of targets for virtual screening experiments. We indeed found that two-structure ensembles outperformed a single target structure in search of both WMIs and NWMIs and resulted in an optimal compromise between actives retrieving and screening speed.

In particular, the two-protein ensemble 3H98/2W−2FVC/7W resulted to be the best target template for the discovery of WM ligands, with the included water molecules playing a key role in bridging the ligand and the protein, besides a plausible role in defining the steric shape of the PSI. On the other hand, the 3HKY/NoW−3SKE/NoW ensemble, with no water included in the target structures, gave optimal results for the mining of NWMIs. When looking at the overall performance in

dx.doi.org/10.1021/ci400367m | J. Chem. Inf. Model. 2014, 54, 481−497

active mining, regardless of them being WM or NWM, good performances are observed with ensembles containing not less than three target structures and employing the water sets NoW and 2W simultaneously.

All together, these results highlight once more that the impact of water molecules in docking-based experiments is highly dependent on several factors like target identity and conformation and ligand chemotype.

Finally, the information gathered from this study may be used to boost the HCV drug discovery programs having as the ultimate objective the rational identification of new NNIs binding at the allosteric PSI.

METHODS

GRID Analysis. The GRID program[52,105] was used to calculate the energies of interaction between the GRID water probe (called "OH2") and PSI residues. To perform the energy calculation, a grid is first defined, and the probe is then moved along the grid points. The energy is evaluated on the basis of a possible interaction with the amino acids of the protein. To define the box size, the 44 X-ray PSI−NNI/NS5B complexes were aligned to the reference structure PDB ID 3TYV using the Schrödinger Maestro interface,[101] and the centroid of the superimposed ligands ("the superligand") was generated. The coordinates of the centroid were *x*: 9.99, *y*: 43.98, *z*: 50.20, and they were used to define the grid center for the GRID calculations performed on each of the studied proteins. The cubic grid had a side length of 25 Å and a grid spacing of 0.33 Å. GRID-generated water molecules were then overlapped to crystallographic water molecules to identify key hydratation sites.

Protein and Ligand Preparation. The crystal structures of NS5B RdRP in complex with PSI−NNIs were retrieved from the RCSB Protein Data Bank[69] and used for our modeling studies.

First, each complex was prepared using Schrödinger's Protein Preparation Wizard[100] to obtain satisfactory starting structures for docking studies. This facility is designed to ensure chemical correctness and to optimize a protein structure for further analysis. In particular, all water molecules except the selected ones were deleted, hydrogen atoms were added, and bond orders were assigned to amino acid residues and ligand. Epik was then used to predict ionization and tautomeric states for the ligands using a pH of 7 ± 1, and the results were double checked using MoKa,[115−117] a software able to accurately compute the pK_a values as well as the most stable tautomer in the aqueous medium for the analyzed compound. According to this analysis, the ligands were considered in the tautomeric forms reported in Table S1 of the Supporting Information, and their ionizable groups were treated as follows: (1) Acrylic acids, proline sulfonamides, acylpyrrolodines, anthranilic acids and indole 2-carboxyl acids were modeled as carboxylates; (2) The enolic group of benzothiadiazine, benzothiazines, and benzoisothiazoles dioxide was modeled as enolate; (3) The nitrogen atom of the acylsulfonamide moiety in the indoles was modeled as negatively charged.

In a next step, optimization of the hydrogen bond network was obtained by reorienting the hydroxyl and thiol groups, amide groups of Asn and Gln, His rings, and water molecules (when present). In addition, the ionization and tautomeric states of His, Asp, Glu, Arg, and Lys were adjusted to match a pH of 7.4. Then, the complexes were refined by running a restrained minimization (OPLS2005 force field),[118] which was stopped when RMSD of heavy atoms reached 0.30 Å, the specified limit by default.

PSI−NNIs were extracted from the prepared corresponding complexes with NS5B, and their coordinates were transformed in order to exclude any biasing toward the experimental NS5B-bound conformation. The compounds were then submitted to a Polak−Ribiere conjugate gradient minimization [0.0005 kJ/(Å mol) convergence].

Docking Experiments. The docking studies were performed without explicit solvent (Set-NoW) and with explicit solvent by using three different sets of water molecules (Set-2W, Set-6W, and Set-7W). Docking simulations were performed using the Glide program.[107]

The prepared structures were used to generate the receptor grids, and no scaling was done for the van der Waals radii of nonpolar receptor atoms. The box was defined by aligning the 40 prepared proteins to the reference structure PDB ID 3TYV using the Schrödinger Maestro interface[101] and calculating coordinates of the centroid of the superimposed ligands ("the superligand"); thus, these coordinates (*x*: 9.99, *y*: 43.98, *z*: 50.20) were used to center the docking grid for all the proteins. The docking space was defined as a 34 Å^3 cubic box, while the diameter\midpoint of docked ligands was required to remain within a smaller nested 14 Å^3 cubic box. In our study, no constrains were used in any of the receptor grids. After grid preparation, previously prepared PSI−NNIs were flexibly docked in a stepwise manner using Glide SP (Standard Precision) and XP (Extra Precision),[107] where all protein structures were treated as rigid. Docking experiments were performed using a 0.80 factor to scale the VdW radii of the nonpolar ligand atoms with a partial atomic charge less than 0.15.

Three poses for each ligand were retrieved from the SP docking. These poses were then refined and rescored using Glide XP, retaining the best scoring bound conformation for each ligand. The obtained average XP score for active ligands was −7.14 (standard deviation = 1.66, higher score obtained = −11.40), while for decoys it was −5.06 (standard deviation = 1.05, higher score obtained = −10.12).

RMSDs with respect to the ligand crystallographic position were computed by all atoms except nonpolar hydrogens using the rmsdcalc utility from Schrödinger.

AUC and Enrichment Factor Calculations. A set of 1200 (30 decoys of each active ligand) decoy compounds was prepared using DecoyFinder.[119] In particular, 30 decoys of each active ligand were selected on the basis of the following constraints. In respect to the known NS5B ligands, we used a Tanimoto coefficient < 0.8, number of hydrogen bonds acceptors and donors ± 2, logP ± 2, MW ± 50, rotatable bonds ± 2, while in respect to the other decoys a Tanimoto coefficient < 0.9 was used. Decoys were mined from the everything-usual subset of the ZINC database version 12.[120,121]

The final virtual screening database, consisting of 1240 compounds, was docked against the best 26 protein−water sets using the previously described docking setting. For each ligand, only the best scoring pose was retained and ranked, regardless of the target protein, and overall AUCs, WMIs AUCs, and NWMIs AUCs, as well as the screening performance in function of the screening progress for WMIs, NWMIs, and all ligands, were calculated for each selected protein−water set and for each possible target ensemble of up to five protein−water structures. The ROC AUC, together with the enrichment factors at 2%, 3%, and 4% of the virtual screening were

dx.doi.org/10.1021/ci400367m | J. Chem. Inf. Model. 2014, 54, 481−497

calculated using the "Enrichment Plotter" node in KNIME.[122] The maximum value for both AUCs and enrichment factors is set to a value of 1.

Figures 2 and 3 were prepared using the software PyMOL.[123]

■ ASSOCIATED CONTENT

Supporting Information

Structures of PSI−NNIs (Figure S1), RMSD values for the aligned proteins (Table S1), RMS fluctuations of NS5B PSI residues (Figure S2), classification of WMIs and NWMIs (Table S2), number of selected water molecules in the complexes (Table S3), GRID energy values (Table S4), NS5B residues interacting with W1−W6 and WB (Table S5), pose predictions results for water sets NoW, 2W, 6W, and 7W (Tables S6−S9), summary of docking performance using the different water sets (Tables S10−S13), conformational flexibility of residues L384 and S556 (Figure S3), and ROC AUCs and enrichment factors for the best performing target ensembles (Tables S13−S15). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*Phone: +39-075-5855157. Fax: +39-075-5855115. E-mail: lbarreca@unipg.it (M.L.B.)

*Phone: +39-075-5855157. Fax: +39-075-5855115. E-mail: nunzio.iraci@gmail.com (N.I.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

R.G. is funded through a scholarship from the POR UMBRIA FSE 2007-2013.

■ ABBREVIATIONS

NNIs: non-nucleoside inhibitors; RdRp: RNA-dependent RNA polymerase; PSI: palm site I; WB: water bridge; WMI: water-mediated inhibitor; NWMI: nonwater-mediated inhibitor. W1−W6: water molecules 1−6; ROC: receiver operating characteristic

■ REFERENCES

(1) Tran, T. T. Overview of epidemiology, diagnosis, and disease progression associated with hepatitis C. *Am. J. Managed Care* **2012**, *18*, S335−339.

(2) Kim, C. W.; Chang, K. M. Hepatitis C virus: Virology and life cycle. *Clin. Mol. Hepatol.* **2013**, *19*, 17−25.

(3) Maasoumy, B.; Port, K.; Markova, A. A.; Serrano, B. C.; Rogalska-Taranta, M.; Sollik, L.; Mix, C.; Kirschner, J.; Manns, M. P.; Wedemeyer, H.; Cornberg, M. Eligibility and safety of triple therapy for hepatitis C: Lessons learned from the first experience in a real world setting. *PLoS One* **2013**, *8*, e55285.

(4) Waheed, Y.; Bhatti, A.; Ashraf, M. RNA dependent RNA polymerase of HCV: A potential target for the development of antiviral drugs. *Infect. Genet. Evol.* **2013**, *14*, 247−257.

(5) Sofia, M. J.; Chang, W.; Furman, P. A.; Mosley, R. T.; Ross, B. S. Nucleoside, nucleotide, and non-nucleoside inhibitors of hepatitis C virus NS5B RNA-dependent RNA-polymerase. *J. Med. Chem.* **2012**, *55*, 2481−2531.

(6) Barreca, M. L.; Iraci, N.; Manfroni, G.; Cecchetti, V. Allosteric inhibition of the hepatitis C virus NS5B polymerase: in silico strategies for drug discovery and development. *Future Med. Chem.* **2011**, *3*, 1027−1055.

(7) 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.

(8) 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.

(9) Haudecoeur, R.; Peuchmaur, M.; Ahmed-Belkacem, A.; Pawlotsky, J. M.; Boumendjel, A. Structure-activity relationships in the development of allosteric hepatitis C virus RNA-dependent RNA polymerase inhibitors: Ten years of research. *Med. Res. Rev.* **2012**, *0*, 1−51.

(10) Barreca, M. L.; Manfroni, G.; Leyssen, P.; Winquist, J.; Kaushik-Basu, N.; Paeshuyse, J.; Krishnan, R.; Iraci, N.; Sabatini, S.; Tabarrini, O.; Basu, A.; Danielson, U. H.; Neyts, J.; Cecchetti, V. Structure-based discovery of pyrazolobenzothiazine derivatives as inhibitors of hepatitis C virus replication. *J. Med. Chem.* **2013**, *56*, 2270−2282.

(11) Manfroni, G.; Meschini, F.; Barreca, M. L.; Leyssen, P.; Samuele, A.; Iraci, N.; Sabatini, S.; Massari, S.; Maga, G.; Neyts, J.; Cecchetti, V. Pyridobenzothiazole derivatives as new chemotype targeting the HCV NS5B polymerase. *Bioorg. Med. Chem.* **2012**, *20*, 866−876.

(12) Talele, T. T. Multiple allosteric pockets of HCV NS5B polymerase and its inhibitors: A structure-based insight. *Curr. Bioact. Compd.* **2008**, *4*, 86−109.

(13) Cozzini, P.; Fornabaio, M.; Marabotti, A.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Free energy of ligand binding to protein: evaluation of the contribution of water molecules by computational methods. *Curr. Med. Chem.* **2004**, *11*, 3093−3118.

(14) Wang, L.; Berne, B. J.; Friesner, R. A. Ligand binding to protein-binding pockets with wet and dry regions. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 1326−1330.

(15) Hummer, G. Molecular binding: Under water's influence. *Nat. Chem.* **2010**, *2*, 906−907.

(16) Ladbury, J. E. Just add water! The effect of water on the specificity of protein−ligand binding sites and its potential application to drug design. *Chem. Biol.* **1996**, *3*, 973−980.

(17) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Energetics of displacing water molecules from protein binding sites: Consequences for ligand optimization. *J. Am. Chem. Soc.* **2009**, *131*, 15403−15411.

(18) Mancera, R. L. Molecular modeling of hydration in drug design. *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 275−280.

(19) de Beer, S. B.; Vermeulen, N. P.; Oostenbrink, C. The role of water molecules in computational drug design. *Curr. Top Med. Chem.* **2010**, *10*, 55−66.

(20) Wong, S. E.; Lightstone, F. C. Accounting for water molecules in drug design. *Expert Opin. Drug Discovery* **2011**, *6*, 65−74.

(21) Amadasi, A.; Surface, J. A.; Spyrakis, F.; Cozzini, P.; Mozzarelli, A.; Kellogg, G. E. Robust classification of "relevant" water molecules in putative protein binding sites. *J. Med. Chem.* **2008**, *51*, 1063−1067.

(22) Amadasi, A.; Spyrakis, F.; Cozzini, P.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Mapping the energetics of water−protein and water−ligand interactions with the "natural" HINT forcefield: Predictive tools for characterizing the roles of water in biomolecules. *J. Mol. Biol.* **2006**, *358*, 289−309.

(23) Lu, Y.; Wang, R.; Yang, C. Y.; Wang, S. Analysis of ligand-bound water molecules in high-resolution crystal structures of protein−ligand complexes. *J. Chem. Inf. Model.* **2007**, *47*, 668−675.

(24) Yang, Y.; Lightstone, F. C.; Wong, S. E. Approaches to efficiently estimate solvation and explicit water energetics in ligand binding: The use of WaterMap. *Expert Opin. Drug Discovery* **2013**, *8*, 277−287.

(25) Poornima, C. S.; Dean, P. M. Hydration in drug design. 1. Multiple hydrogen-bonding features of water molecules in mediating protein−ligand interactions. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 500−512.

(26) Poornima, C. S.; Dean, P. M. Hydration in drug design. 2. Influence of local site surface shape on water binding. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 513−520.

(27) Poornima, C. S.; Dean, P. M. Hydration in drug design. 3. Conserved water molecules at the ligand-binding sites of homologous proteins. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 521−531.
(28) Barillari, C.; Taylor, J.; Viner, R.; Essex, J. W. Classification of water molecules in protein binding sites. *J. Am. Chem. Soc.* **2007**, *129*, 2577−2587.
(29) Garcia-Sosa, A. T. Hydration properties of ligands and drugs in protein binding sites: Tightly-bound, bridging water molecules and their effects and consequences on molecular design strategies. *J. Chem. Inf. Model.* **2013**, *53*, 1388−1405.
(30) Stegmann, C. M.; Seeliger, D.; Sheldrick, G. M.; de Groot, B. L.; Wahl, M. C. The thermodynamic influence of trapped water molecules on a protein−ligand interaction. *Angew. Chem., Int. Ed. Engl.* **2009**, *48*, 5207−5210.
(31) Genheden, S.; Mikulskis, P.; Hu, L.; Kongsted, J.; Soderhjelm, P.; Ryde, U. Accurate predictions of nonpolar solvation free energies require explicit consideration of binding-site hydration. *J. Am. Chem. Soc.* **2011**, *133*, 13081−13092.
(32) Huggins, D. J.; Tidor, B. Systematic placement of structural water molecules for improved scoring of protein−ligand interactions. *Protein Eng., Des. Sel.* **2011**, *24*, 777−789.
(33) Limongelli, V.; Marinelli, L.; Cosconati, S.; La Motta, C.; Sartini, S.; Mugnaini, L.; Da Settimo, F.; Novellino, E.; Parrinello, M. Sampling protein motion and solvent effect during ligand binding. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 1467−1472.
(34) Corbeil, C. R.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 3. Impact of input ligand conformation, protein flexibility, and water molecules on the accuracy of docking programs. *J. Chem. Inf. Model.* **2009**, *49*, 997−1009.
(35) Lu, S. Y.; Jiang, Y. J.; Lv, J.; Zou, J. W.; Wu, T. X. Role of bridging water molecules in GSK3beta-inhibitor complexes: insights from QM/MM, MD, and molecular docking studies. *J. Comput. Chem.* **2011**, *32*, 1907−1918.
(36) Kumar, A.; Zhang, K. Y. Investigation on the effect of key water molecules on docking performance in CSARdock exercise. *J. Chem. Inf. Model.* **2013**, *53*, 1880−1892.
(37) Thilagavathi, R.; Mancera, R. L. Ligand−protein cross-docking with water molecules. *J. Chem. Inf. Model.* **2010**, *50*, 415−421.
(38) Roberts, B. C.; Mancera, R. L. Ligand−protein docking with water molecules. *J. Chem. Inf. Model.* **2008**, *48*, 397−408.
(39) de Graaf, C.; Pospisil, P.; Pos, W.; Folkers, G.; Vermeulen, N. P. Binding mode prediction of cytochrome p450 and thymidine kinase protein−ligand complexes by consideration of water and rescoring in automated docking. *J. Med. Chem.* **2005**, *48*, 2308−2318.
(40) Yang, J. M.; Chen, C. C. GEMDOCK: A generic evolutionary method for molecular docking. *Proteins* **2004**, *55*, 288−304.
(41) Santos, R.; Hritz, J.; Oostenbrink, C. Role of water in molecular docking simulations of cytochrome P450 2D6. *J. Chem. Inf. Model.* **2010**, *50*, 146−154.
(42) Huang, N.; Shoichet, B. K. Exploiting ordered waters in molecular docking. *J. Med. Chem.* **2008**, *51*, 4862−4865.
(43) Villacanas, O.; Madurga, S.; Giralt, E.; Belda, I. Explicit treatment of water molecules in protein−ligand docking. *Curr. Comput.-Aided Drug Des.* **2009**, *5*, 145−154.
(44) Rossato, G.; Ernst, B.; Vedani, A.; Smiesko, M. AcquaAlta: A directional approach to the solvation of ligand−protein complexes. *J. Chem. Inf. Model.* **2011**, *51*, 1867−1881.
(45) Ross, G. A.; Morris, G. M.; Biggin, P. C. Rapid and accurate prediction and scoring of water molecules in protein binding sites. *PLoS One* **2012**, *7*, e32036.
(46) Raymer, M. L.; Sanschagrin, P. C.; Punch, W. F.; Venkataraman, S.; Goodman, E. D.; Kuhn, L. A. Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. *J. Mol. Biol.* **1997**, *265*, 445−464.
(47) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Prediction of the water content in protein binding sites. *J. Phys. Chem. B.* **2009**, *113*, 13337−13346.
(48) Mason, J. S.; Bortolato, A.; Congreve, M.; Marshall, F. H. New insights from structural biology into the druggability of G protein-coupled receptors. *Trends Pharmacol. Sci.* **2012**, *33*, 249−260.
(49) Imai, T.; Oda, K.; Kovalenko, A.; Hirata, F.; Kidera, A. Ligand mapping on protein surfaces by the 3D-RISM theory: toward computational fragment-based drug design. *J. Am. Chem. Soc.* **2009**, *131*, 12430−12440.
(50) Imai, T.; Hiraoka, R.; Kovalenko, A.; Hirata, F. Locating missing water molecules in protein cavities by the three-dimensional reference interaction site model theory of molecular solvation. *Proteins* **2007**, *66*, 804−813.
(51) Henchman, R. H.; McCammon, J. A. Extracting hydration sites around proteins from explicit water simulations. *J. Comput. Chem.* **2002**, *23*, 861−869.
(52) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.
(53) Garcia-Sosa, A. T.; Mancera, R. L.; Dean, P. M. WaterScore: A novel method for distinguishing between bound and displaceable water molecules in the crystal structure of the binding site of protein−ligand complexes. *J. Mol. Model.* **2003**, *9*, 172−182.
(54) Forli, S.; Olson, A. J. A force field with discrete displaceable waters and desolvation entropy for hydrated ligand docking. *J. Med. Chem.* **2012**, *55*, 623−638.
(55) Ehrlich, L.; Reczko, M.; Bohr, H.; Wade, R. C. Prediction of protein hydration sites from sequence by modular neural networks. *Protein Eng.* **1998**, *11*, 11−19.
(56) Bottoms, C. A.; White, T. A.; Tanner, J. J. Exploring structurally conserved solvent sites in protein families. *Proteins* **2006**, *64*, 404−421.
(57) Beuming, T.; Che, Y.; Abel, R.; Kim, B.; Shanmugasundaram, V.; Sherman, W. Thermodynamic analysis of water molecules at the surface of proteins and applications to binding site prediction and characterization. *Proteins* **2012**, *80*, 871−883.
(58) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.* **2008**, *130*, 2817−2831.
(59) *SZMAP*, version 1.1.0; OpenEye Scientific Software, Inc.: Santa Fe, NM, 2012. www.eyesopen.com.
(60) Elokely, K. M.; Doerksen, R. J. Docking challenge: Protein sampling and molecular docking performance. *J. Chem. Inf. Model.* **2013**, *53*, 1934−1945.
(61) Huang, S. Y.; Zou, X. Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking. *Proteins* **2007**, *66*, 399−421.
(62) Huang, K.; Lu, W.; Anderson, S.; Laskowski, M., Jr.; James, M. N. Water molecules participate in proteinase-inhibitor interactions: Crystal structures of Leu18, Ala18, and Gly18 variants of turkey ovomucoid inhibitor third domain complexed with Streptomyces griseus proteinase B. *Protein Sci.* **1995**, *4*, 1985−1997.
(63) Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins* **2002**, *46*, 34−40.
(64) Corbeil, C. R.; Englebienne, P.; Yannopoulos, C. G.; Chan, L.; Das, S. K.; Bilimoria, D.; L'Heureux, L.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 2. Development and application of fitted 1.5 to the virtual screening of potential HCV polymerase inhibitors. *J. Chem. Inf. Model.* **2008**, *48*, 902−909.
(65) Knegtel, R. M.; Kuntz, I. D.; Oshiro, C. M. Molecular docking to ensembles of protein structures. *J. Mol. Biol.* **1997**, *266*, 424−440.
(66) Nichols, S. E.; Domaoal, R. A.; Thakur, V. V.; Tirado-Rives, J.; Anderson, K. S.; Jorgensen, W. L. Discovery of wild-type and Y181C mutant non-nucleoside HIV-1 reverse transcriptase inhibitors using virtual screening with multiple protein structures. *J. Chem. Inf. Model.* **2009**, *49*, 1272−1279.
(67) Garcia-Sosa, A. T.; Mancera, R. L. Free energy calculations of mutations involving a tightly bound water molecule and ligand substitutions in a ligand−protein complex. *Mol. Inf.* **2010**, *29*, 589−600.

(68) Vinh, N. B.; Simpson, J. S.; Scammells, P. J.; Chalmers, D. K. Virtual screening using a conformationally flexible target protein: Models for ligand binding to p38alpha MAPK. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 409−423.
(69) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(70) Pfefferkorn, J. A.; Nugent, R.; Gross, R. J.; Greene, M.; Mitchell, M. A.; Reding, M. T.; Funk, L. A.; Anderson, R.; Wells, P. A.; Shelly, J. A.; Anstadt, R.; Finzel, B. C.; Harris, M. S.; Kilkuskie, R. E.; Kopta, L. A.; Schwende, F. J. Inhibitors of HCV NS5B polymerase. Part 2: Evaluation of the northern region of (2Z)-2-benzoylamino-3-(4-phenoxy-phenyl)-acrylic acid. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2812−2818.
(71) Pfefferkorn, J. A.; Greene, M. L.; Nugent, R. A.; Gross, R. J.; Mitchell, M. A.; Finzel, B. C.; Harris, M. S.; Wells, P. A.; Shelly, J. A.; Anstadt, R. A.; Kilkuskie, R. E.; Kopta, L. A.; Schwende, F. J. Inhibitors of HCV NS5B polymerase. Part 1: Evaluation of the southern region of (2Z)-2-(benzoylamino)-3-(5-phenyl-2-furyl)acrylic acid. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2481−2486.
(72) Powers, J. P.; Piper, D. E.; Li, Y.; Mayorga, V.; Anzola, J.; Chen, J. M.; Jaen, J. C.; Lee, G.; Liu, J.; Peterson, M. G.; Tonn, G. R.; Ye, Q.; Walker, N. P.; Wang, Z. SAR and mode of action of novel non-nucleoside inhibitors of hepatitis C NS5b RNA polymerase. *J. Med. Chem.* **2006**, *49*, 1034−1046.
(73) Tedesco, R.; Shaw, A. N.; Bambal, R.; Chai, D.; Concha, N. O.; Darcy, M. G.; Dhanak, D.; Fitch, D. M.; Gates, A.; Gerhardt, W. G.; Halegoua, D. L.; Han, C.; Hofmann, G. A.; Johnston, V. K.; Kaura, A. C.; Liu, N.; Keenan, R. M.; Lin-Goerke, J.; Sarisky, R. T.; Wiggall, K. J.; Zimmerman, M. N.; Duffy, K. J. 3-(1,1-Dioxo-2H-(1,2,4)-benzothiadiazin-3-yl)-4-hydroxy-2(1H)-quinolinones, potent inhibitors of hepatitis C virus RNA-dependent RNA polymerase. *J. Med. Chem.* **2006**, *49*, 971−983.
(74) Le Pogam, S.; Kang, H.; Harris, S. F.; Leveque, V.; Giannetti, A. M.; Ali, S.; Jiang, W. R.; Rajyaguru, S.; Tavares, G.; Oshiro, C.; Hendricks, T.; Klumpp, K.; Symons, J.; Browner, M. F.; Cammack, N.; Najera, I. Selection and characterization of replicon variants dually resistant to thumb- and palm-binding nonnucleoside polymerase inhibitors of the hepatitis C virus. *J. Virol.* **2006**, *80*, 6146−6154.
(75) Zhou, Y.; Webber, S. E.; Murphy, D. E.; Li, L. S.; Dragovich, P. S.; Tran, C. V.; Sun, Z.; Ruebsam, F.; Shah, A. M.; Tsan, M.; Showalter, R. E.; Patel, R.; Li, B.; Zhao, Q.; Han, Q.; Hermann, T.; Kissinger, C. R.; Lebrun, L.; Sergeeva, M. V.; Kirkovsky, L. Novel HCV NS5B polymerase inhibitors derived from 4-(1′,1′-dioxo-1′,4′-dihydro-1′lambda6-benzo[1′,2′,4′]thiadiazin-3′-yl)-5-hydroxy −2H-pyridazin-3-ones. Part 1: Exploration of 7′-substitution of benzothiadiazine. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1413−1418.
(76) Shaw, A. N.; Tedesco, R.; Bambal, R.; Chai, D.; Concha, N. O.; Darcy, M. G.; Dhanak, D.; Duffy, K. J.; Fitch, D. M.; Gates, A.; Johnston, V. K.; Keenan, R. M.; Lin-Goerke, J.; Liu, N.; Sarisky, R. T.; Wiggall, K. J.; Zimmerman, M. N. Substituted benzothiadizine inhibitors of Hepatitis C virus polymerase. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 4350−4353.
(77) Li, L. S.; Zhou, Y.; Murphy, D. E.; Stankovic, N.; Zhao, J.; Dragovich, P. S.; Bertolini, T.; Sun, Z.; Ayida, B.; Tran, C. V.; Ruebsam, F.; Webber, S. E.; Shah, A. M.; Tsan, M.; Showalter, R. E.; Patel, R.; Lebrun, L. A.; Bartkowski, D. M.; Nolan, T. G.; Norris, D. A.; Kamran, R.; Brooks, J.; Sergeeva, M. V.; Kirkovsky, L.; Zhao, Q.; Kissinger, C. R. Novel HCV NS5B polymerase inhibitors derived from 4-(1′,1′-dioxo-1′,4′-dihydro-1′lambda(6)-benzo[1′,2′,4′]thiadiazin-3′-yl)-5-hydro xy-2H-pyridazin-3-ones. Part 3: Further optimization of the 2-, 6-, and 7′-substituents and initial pharmacokinetic assessments. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 3446−3455.
(78) Dragovich, P. S.; Blazel, J. K.; Ellis, D. A.; Han, Q.; Kamran, R.; Kissinger, C. R.; LeBrun, L. A.; Li, L. S.; Murphy, D. E.; Noble, M.; Patel, R. A.; Ruebsam, F.; Sergeeva, M. V.; Shah, A. M.; Showalter, R. E.; Tran, C. V.; Tsan, M.; Webber, S. E.; Kirkovsky, L.; Zhou, Y. Novel HCV NS5B polymerase inhibitors derived from 4-(1′,1′-dioxo-1′,4′-dihydro-1′lambda(6)-benzo[1′,2′,4′]thiadiazin-3′-yl)-5-hydro xy-2H-pyridazin-3-ones. Part 5: Exploration of pyridazinones containing 6-amino-substituents. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5635−5639.
(79) Ruebsam, F.; Murphy, D. E.; Tran, C. V.; Li, L. S.; Zhao, J.; Dragovich, P. S.; McGuire, H. M.; Xiang, A. X.; Sun, Z.; Ayida, B. K.; Blazel, J. K.; Kim, S. H.; Zhou, Y.; Han, Q.; Kissinger, C. R.; Webber, S. E.; Showalter, R. E.; Shah, A. M.; Tsan, M.; Patel, R. A.; Thompson, P. A.; Lebrun, L. A.; Hou, H. J.; Kamran, R.; Sergeeva, M. V.; Bartkowski, D. M.; Nolan, T. G.; Norris, D. A.; Khandurina, J.; Brooks, J.; Okamoto, E.; Kirkovsky, L. Discovery of tricyclic 5,6-dihydro-1H-pyridin-2-ones as novel, potent, and orally bioavailable inhibitors of HCV NS5B polymerase. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 6404−6412.
(80) Wang, G.; Lei, H.; Wang, X.; Das, D.; Hong, J.; Mackinnon, C. H.; Coulter, T. S.; Montalbetti, C. A.; Mears, R.; Gai, X.; Bailey, S. E.; Ruhrmund, D.; Hooi, L.; Misialek, S.; Rajagopalan, P. T.; Cheng, R. K.; Barker, J. J.; Felicetti, B.; Schonfeld, D. L.; Stoycheva, A.; Buckman, B. O.; Kossen, K.; Seiwert, S. D.; Beigelman, L. HCV NS5B polymerase inhibitors 2: Synthesis and in vitro activity of (1,1-dioxo-2H-[1,2,4]benzothiadiazin-3-yl) azolo[1,5-a]pyridine and azolo[1,5-a]-pyrimidine derivatives. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 4480−4483.
(81) Ellis, D. A.; Blazel, J. K.; Tran, C. V.; Ruebsam, F.; Murphy, D. E.; Li, L. S.; Zhao, J.; Zhou, Y.; McGuire, H. M.; Xiang, A. X.; Webber, S. E.; Zhao, Q.; Han, Q.; Kissinger, C. R.; Lardy, M.; Gobbi, A.; Showalter, R. E.; Shah, A. M.; Tsan, M.; Patel, R. A.; LeBrun, L. A.; Kamran, R.; Bartkowski, D. M.; Nolan, T. G.; Norris, D. A.; Sergeeva, M. V.; Kirkovsky, L. 5,5′- and 6,6′-Dialkyl-5,6-dihydro-1H-pyridin-2-ones as potent inhibitors of HCV NS5B polymerase. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 6047−6052.
(82) Ruebsam, F.; Webber, S. E.; Tran, M. T.; Tran, C. V.; Murphy, D. E.; Zhao, J.; Dragovich, P. S.; Kim, S. H.; Li, L. S.; Zhou, Y.; Han, Q.; Kissinger, C. R.; Showalter, R. E.; Lardy, M.; Shah, A. M.; Tsan, M.; Patel, R.; Lebrun, L. A.; Kamran, R.; Sergeeva, M. V.; Bartkowski, D. M.; Nolan, T. G.; Norris, D. A.; Kirkovsky, L. Pyrrolo[1,2-b]pyridazin-2-ones as potent inhibitors of HCV NS5B polymerase. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 3616−3621.
(83) Ellis, D. A.; Blazel, J. K.; Webber, S. E.; Tran, C. V.; Dragovich, P. S.; Sun, Z.; Ruebsam, F.; McGuire, H. M.; Xiang, A. X.; Zhao, J.; Li, L. S.; Zhou, Y.; Han, Q.; Kissinger, C. R.; Showalter, R. E.; Lardy, M.; Shah, A. M.; Tsan, M.; Patel, R.; LeBrun, L. A.; Kamran, R.; Bartkowski, D. M.; Nolan, T. G.; Norris, D. A.; Sergeeva, M. V.; Kirkovsky, L. 4-(1,1-Dioxo-1,4-dihydro-1lambda6-benzo[1,4]thiazin-3-yl)-5-hydroxy-2H-pyridazin- 3-ones as potent inhibitors of HCV NS5B polymerase. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4628−4632.
(84) de Vicente, J.; Hendricks, R. T.; Smith, D. B.; Fell, J. B.; Fischer, J.; Spencer, S. R.; Stengel, P. J.; Mohr, P.; Robinson, J. E.; Blake, J. F.; Hilgenkamp, R. K.; Yee, C.; Adjabeng, G.; Elworthy, T. R.; Tracy, J.; Chin, E.; Li, J.; Wang, B.; Bamberg, J. T.; Stephenson, R.; Oshiro, C.; Harris, S. F.; Ghate, M.; Leveque, V.; Najera, I.; Le Pogam, S.; Rajyaguru, S.; Ao-Ieong, G.; Alexandrova, L.; Larrabee, S.; Brandl, M.; Briggs, A.; Sukhtankar, S.; Farrell, R.; Xu, B. Non-nucleoside inhibitors of HCV polymerase NS5B. Part 2: Synthesis and structure−activity relationships of benzothiazine-substituted quinolinediones. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3642−3646.
(85) de Vicente, J.; Hendricks, R. T.; Smith, D. B.; Fell, J. B.; Fischer, J.; Spencer, S. R.; Stengel, P. J.; Mohr, P.; Robinson, J. E.; Blake, J. F.; Hilgenkamp, R. K.; Yee, C.; Zhao, J.; Elworthy, T. R.; Tracy, J.; Chin, E.; Li, J.; Lui, A.; Wang, B.; Oshiro, C.; Harris, S. F.; Ghate, M.; Leveque, V. J.; Najera, I.; Le Pogam, S.; Rajyaguru, S.; Ao-Ieong, G.; Alexandrova, L.; Fitch, B.; Brandl, M.; Masjedizadeh, M.; Wu, S. Y.; de Keczer, S.; Voronin, T. Non-nucleoside inhibitors of HCV polymerase NS5B. Part 3: synthesis and optimization studies of benzothiazine-substituted tetramic acids. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 5648−5651.
(86) Kim, S. H.; Tran, M. T.; Ruebsam, F.; Xiang, A. X.; Ayida, B.; McGuire, H.; Ellis, D.; Blazel, J.; Tran, C. V.; Murphy, D. E.; Webber, S. E.; Zhou, Y.; Shah, A. M.; Tsan, M.; Showalter, R. E.; Patel, R.; Gobbi, A.; LeBrun, L. A.; Bartkowski, D. M.; Nolan, T. G.; Norris, D. A.; Sergeeva, M. V.; Kirkovsky, L.; Zhao, Q.; Han, Q.; Kissinger, C. R. Structure-based design, synthesis, and biological evaluation of 1,1-

dioxoisothiazole and benzo[b]thiophene-1,1-dioxide derivatives as novel inhibitors of hepatitis C virus NS5B polymerase. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4181−4185.
(87) de Vicente, J.; Hendricks, R. T.; Smith, D. B.; Fell, J. B.; Fischer, J.; Spencer, S. R.; Stengel, P. J.; Mohr, P.; Robinson, J. E.; Blake, J. F.; Hilgenkamp, R. K.; Yee, C.; Adjabeng, G.; Elworthy, T. R.; Li, J.; Wang, B.; Bamberg, J. T.; Harris, S. F.; Wong, A.; Leveque, V. J.; Najera, I.; Le Pogam, S.; Rajyaguru, S.; Ao-Ieong, G.; Alexandrova, L.; Larrabee, S.; Brandl, M.; Briggs, A.; Sukhtankar, S.; Farrell, R. Non-nucleoside inhibitors of HCV polymerase NS5B. Part 4: structure-based design, synthesis, and biological evaluation of benzo[d]-isothiazole-1,1-dioxides. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 5652−5656.
(88) Gopalsamy, A.; Chopra, R.; Lim, K.; Ciszewski, G.; Shi, M.; Curran, K. J.; Sukits, S. F.; Svenson, K.; Bard, J.; Ellingboe, J. W.; Agarwal, A.; Krishnamurthy, G.; Howe, A. Y.; Orlowski, M.; Feld, B.; O'Connell, J.; Mansour, T. S. Discovery of proline sulfonamides as potent and selective hepatitis C virus NS5b polymerase inhibitors. Evidence for a new NS5b polymerase binding site. *J. Med. Chem.* **2006**, *49*, 3052−3055.
(89) Slater, M. J.; Amphlett, E. M.; Andrews, D. M.; Bravi, G.; Burton, G.; Cheasty, A. G.; Corfield, J. A.; Ellis, M. R.; Fenwick, R. H.; Fernandes, S.; Guidetti, R.; Haigh, D.; Hartley, C. D.; Howes, P. D.; Jackson, D. L.; Jarvest, R. L.; Lovegrove, V. L.; Medhurst, K. J.; Parry, N. R.; Price, H.; Shah, P.; Singh, O. M.; Stocker, R.; Thommes, P.; Wilkinson, C.; Wonacott, A. Optimization of novel acyl pyrrolidine inhibitors of hepatitis C virus RNA-dependent RNA polymerase leading to a development candidate. *J. Med. Chem.* **2007**, *50*, 897−900.
(90) Nittoli, T.; Curran, K.; Insaf, S.; DiGrandi, M.; Orlowski, M.; Chopra, R.; Agarwal, A.; Howe, A. Y.; Prashad, A.; Floyd, M. B.; Johnson, B.; Sutherland, A.; Wheless, K.; Feld, B.; O'Connell, J.; Mansour, T. S.; Bloom, J. Identification of anthranilic acid derivatives as a novel class of allosteric inhibitors of hepatitis C NS5B polymerase. *J. Med. Chem.* **2007**, *50*, 2108−2116.
(91) McGowan, D.; Nyanguile, O.; Cummings, M. D.; Vendeville, S.; Vandyck, K.; Van den Broeck, W.; Boutton, C. W.; De Bondt, H.; Quirynen, L.; Amssoms, K.; Bonfanti, J. F.; Last, S.; Rombauts, K.; Tahri, A.; Hu, L.; Delouvroy, F.; Vermeiren, K.; Vandercruyssen, G.; Van der Helm, L.; Cleiren, E.; Mostmans, W.; Lory, P.; Pille, G.; Van Emelen, K.; Fanning, G.; Pauwels, F.; Lin, T. I.; Simmen, K.; Raboisson, P. 1,5-Benzodiazepine inhibitors of HCV NS5B polymerase. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 2492−2496.
(92) Nyanguile, O.; Pauwels, F.; Van den Broeck, W.; Boutton, C. W.; Quirynen, L.; Ivens, T.; van der Helm, L.; Vandercruyssen, G.; Mostmans, W.; Delouvroy, F.; Dehertogh, P.; Cummings, M. D.; Bonfanti, J. F.; Simmen, K. A.; Raboisson, P. 1,5-Benzodiazepines, a novel class of hepatitis C virus polymerase nonnucleoside inhibitors. *Antimicrob. Agents Chemother.* **2008**, *52*, 4420−4431.
(93) Vandyck, K.; Cummings, M. D.; Nyanguile, O.; Boutton, C. W.; Vendeville, S.; McGowan, D.; Devogelaere, B.; Amssoms, K.; Last, S.; Rombauts, K.; Tahri, A.; Lory, P.; Hu, L.; Beauchamp, D. A.; Simmen, K.; Raboisson, P. Structure-based design of a benzodiazepine scaffold yields a potent allosteric inhibitor of hepatitis C NS5B RNA polymerase. *J. Med. Chem.* **2009**, *52*, 4099−4102.
(94) Nyanguile, O.; Devogelaere, B.; Vijgen, L.; Van den Broeck, W.; Pauwels, F.; Cummings, M. D.; De Bondt, H. L.; Vos, A. M.; Berke, J. M.; Lenz, O.; Vandercruyssen, G.; Vermeiren, K.; Mostmans, W.; Dehertogh, P.; Delouvroy, F.; Vendeville, S.; VanDyck, K.; Dockx, K.; Cleiren, E.; Raboisson, P.; Simmen, K. A.; Fanning, G. C. 1a/1b Subtype profiling of nonnucleoside polymerase inhibitors of hepatitis C virus. *J. Virol.* **2010**, *84*, 2923−2934.
(95) Anilkumar, G. N.; Lesburg, C. A.; Selyutin, O.; Rosenblum, S. B.; Zeng, Q.; Jiang, Y.; Chan, T. Y.; Pu, H.; Vaccaro, H.; Wang, L.; Bennett, F.; Chen, K. X.; Duca, J.; Gavalas, S.; Huang, Y.; Pinto, P.; Sannigrahi, M.; Velazquez, F.; Venkatraman, S.; Vibulbhan, B.; Agrawal, S.; Butkiewicz, N.; Feld, B.; Ferrari, E.; He, Z.; Jiang, C. K.; Palermo, R. E.; McMonagle, P.; Huang, H. C.; Shih, N. Y.; Njoroge, G.; Kozlowski, J. A. I. Novel HCV NS5B polymerase inhibitors: Discovery of indole 2-carboxylic acids with C3-heterocycles. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 5336−5341.
(96) Anilkumar, G. N.; Selyutin, O.; Rosenblum, S. B.; Zeng, Q.; Jiang, Y.; Chan, T. Y.; Pu, H.; Wang, L.; Bennett, F.; Chen, K. X.; Lesburg, C. A.; Duca, J.; Gavalas, S.; Huang, Y.; Pinto, P.; Sannigrahi, M.; Velazquez, F.; Venkatraman, S.; Vibulbhan, B.; Agrawal, S.; Ferrari, E.; Jiang, C. K.; Huang, H. C.; Shih, N. Y.; George Njoroge, F.; Kozlowski, J. A., II Novel HCV NS5B polymerase inhibitors: Discovery of indole C2 acyl sulfonamides. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 713−717.
(97) Velazquez, F.; Venkatraman, S.; Lesburg, C. A.; Duca, J.; Rosenblum, S. B.; Kozlowski, J. A.; Njoroge, F. G. Synthesis of new 4,5-dihydrofuranoindoles and their evaluation as HCV NS5B polymerase inhibitors. *Org. Lett.* **2012**, *14*, 556−559.
(98) Chen, K. X.; Lesburg, C. A.; Vibulbhan, B.; Yang, W.; Chan, T. Y.; Venkatraman, S.; Velazquez, F.; Zeng, Q.; Bennett, F.; Anilkumar, G. N.; Duca, J.; Jiang, Y.; Pinto, P.; Wang, L.; Huang, Y.; Selyutin, O.; Gavalas, S.; Pu, H.; Agrawal, S.; Feld, B.; Huang, H. C.; Li, C.; Cheng, K. C.; Shih, N. Y.; Kozlowski, J. A.; Rosenblum, S. B.; Njoroge, F. G. A novel class of highly potent irreversible hepatitis C virus NS5B polymerase inhibitors. *J. Med. Chem.* **2012**, *55*, 2089−2101.
(99) Chen, K. X.; Vibulbhan, B.; Yang, W.; Sannigrahi, M.; Velazquez, F.; Chan, T. Y.; Venkatraman, S.; Anilkumar, G. N.; Zeng, Q.; Bennet, F.; Jiang, Y.; Lesburg, C. A.; Duca, J.; Pinto, P.; Gavalas, S.; Huang, Y.; Wu, W.; Selyutin, O.; Agrawal, S.; Feld, B.; Huang, H. C.; Li, C.; Cheng, K. C.; Shih, N. Y.; Kozlowski, J. A.; Rosenblum, S. B.; Njoroge, F. G. Structure−activity relationship (SAR) development and discovery of potent indole-based inhibitors of the hepatitis C virus (HCV) NS5B polymerase. *J. Med. Chem.* **2012**, *55*, 754−765.
(100) *Schrödinger Suite 2012: Protein Preparation Wizard; Epik*, version 2.3; Schrödinger, LLC: New York, 2012. *Impact*, version 5.8; Schrödinger, LLC: New York, 2012. *Prime*, version 3.1; Schrödinger, LLC: New York, 2012.
(101) *Maestro*, version 9.3; Schrödinger, LLC: New York, 2012.
(102) Caillet-Saguy, C.; Simister, P. C.; Bressanelli, S. An objective assessment of conformational variability in complexes of hepatitis C virus polymerase with non-nucleoside inhibitors. *J. Mol. Biol.* **2011**, *414*, 370−384.
(103) Ismail, M. A.; Abou El Ella, D. A.; Abouzid, K. A.; Mahmoud, A. H. Integrated structure-based activity prediction model of benzothiadiazines on various genotypes of HCV NS5b polymerase (1a, 1b and 4) and its application in the discovery of new derivatives. *Bioorg. Med. Chem.* **2012**, *20*, 2455−2478.
(104) Mahmoud, A. H.; Abou El Ella, D. A.; Ismail, M. A.; Abouzid, K. A. A highly selective structure-based virtual screening model of Palm I allosteric inhibitors of HCV Ns5b polymerase enzyme and its application in the discovery and optimization of new analogues. *Eur. J. Med. Chem.* **2012**, *57*, 468−482.
(105) *GRID*, version 22c; Molecular Discovery Ltd.: Pinner, Middlesex, U.K., 2012. http://www.moldiscovery.com/soft_grid.php.
(106) 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.
(107) *Schrödinger Suite 2012: Glide*, version 5.8; Schrödinger LLC: New York, 2012.
(108) Barreca, M. L.; De Luca, L.; Iraci, N.; Rao, A.; Ferro, S.; Maga, G.; Chimirri, A. Structure-based pharmacophore identification of new chemical scaffolds as non-nucleoside reverse transcriptase inhibitors. *J. Chem. Inf. Model.* **2007**, *47*, 557−62.
(109) Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins* **2004**, *57*, 225−242.
(110) Perola, E.; Walters, W. P.; Charifson, P. S. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* **2004**, *56*, 235−249.
(111) Bauer, M. R.; Ibrahim, T. M.; Vogel, S. M.; Boeckler, F. M. Evaluation and optimization of virtual screening workflows with

DEKOIS 2.0: A public library of challenging docking benchmark sets. *J. Chem. Inf. Model.* **2013**, *53*, 1447−1462.
(112) Damm-Ganamet, K. L.; Smith, R. D.; Dunbar, J. B., Jr.; Stuckey, J. A.; Carlson, H. A. CSAR Benchmark Exercise 2011−2012: Evaluation of results from docking and relative ranking of blinded congeneric series. *J. Chem. Inf. Model.* **2013**, *53*, 1853−1870.
(113) Triballeau, N.; Acher, F.; Brabet, I.; Pin, J. P.; Bertrand, H. O. Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J. Med. Chem.* **2005**, *48*, 2534−2547.
(114) Rao, S.; Sanschagrin, P. C.; Greenwood, J. R.; Repasky, M. P.; Sherman, W.; Farid, R. Improving database enrichment through ensemble docking. *J. Comput.- Aided Mol. Des.* **2008**, *22*, 621−627.
(115) *MoKa*, version 2.0.1; Molecular Discovery Ltd.: Pinner, Middlesex, U.K., 2012. http://www.moldiscovery.com/soft_moka.php.
(116) Milletti, F.; Storchi, L.; Sforna, G.; Cruciani, G. New and Original p*K*a prediction method using Grid Molecular Interaction Fields. *J. Chem. Inf. Mod.* **2007**, *47*, 2172−2181.
(117) Milletti, F.; Storchi, L.; Sforna, G.; Cross, S.; Cruciani, G. Tautomer enumeration and stability prediction for virtual screening on large chemical databases. *J. Chem. Inf. Mod.* **2009**, *49*, 68−75.
(118) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *45*, 11225−11236.
(119) Cereto-Massague, A.; Guasch, L.; Valls, C.; Mulero, M.; Pujadas, G.; Garcia-Vallve, S. DecoyFinder: An easy-to-use python GUI application for building target-specific decoy sets. *Bioinformatics* **2012**, *28*, 1661−1662.
(120) Irwin, J. J.; Shoichet, B. K. ZINC−A free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177−182.
(121) Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC: A free tool to discover chemistry for biology. *J. Chem. Inf. Model.* **2012**, *52*, 1757−1768.
(122) Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Sieb, C.; Thiel, K.; Wiswedel, B. KNIME: The Konstanz Information Miner. In *Data Analysis, Machine Learning and Applications*; Preisach, C., Burkhardt, H., Schmidt-Thieme, L., Decker, R., Eds.; Springer: Berlin, 2008; pp 319−326.
(123) *The PyMOL Molecular Graphics System*, version 1.5.0.4; Schrödinger, LLC: New York, 2010.