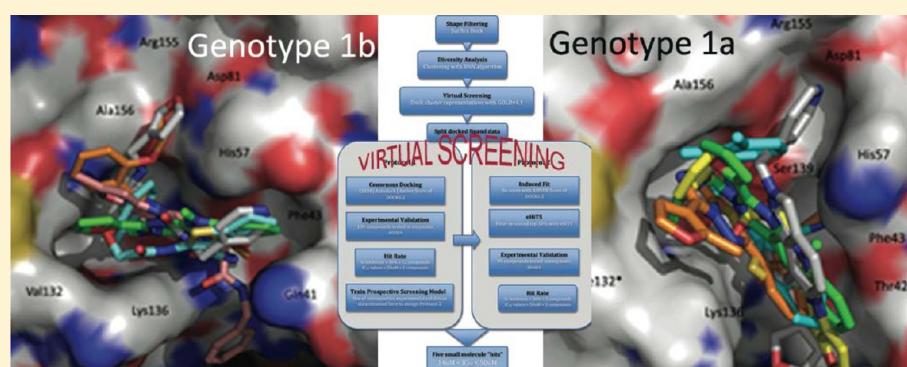


Identification of Non-Macrocyclic Small Molecule Inhibitors against the NS3/4A Serine Protease of Hepatitis C Virus through in Silico Screening

Rima Chaudhuri,^{†,‡,§,||} Hyun Lee,^{†,§} Lena Truong,[†] Jaime Torres,[†] Kavankumar Patel,[†] and Michael E. Johnson^{*,†}

[†]Center for Pharmaceutical Biotechnology, Department of Medicinal Chemistry and Pharmacognosy, University of Illinois at Chicago, 900 S. Ashland Ave., M/C 870, Chicago, Illinois 60607, United States

[‡]Department of Bioengineering, Bioinformatics Program, University of Illinois at Chicago, 851 S. Morgan St., 218 SEO, M/C 063, Chicago, Illinois 60607, United States



ABSTRACT: Drug discovery and design for inhibition of the Hepatitis C Virus (HCV) NS3/4A serine protease is a major challenge. The broad, shallow, and generally featureless nature of the active site makes it a difficult target for “hit” selection especially using standard docking programs. There are several macrocyclic NS3/4A protease inhibitors that have been approved or are in clinical trials to treat chronic HCV (alone or as combination therapy), but most of the current therapies for HCV infection have untoward side effects, indicating a continuing medical need for the discovery of novel therapeutics with improved efficacy. In this study, we designed and implemented a two-tiered and progressive docking regime that successfully identified five non-macrocyclic small molecules that show inhibitory activity in the low micromolar range. Of these, four compounds show varying inhibition against HCV subgenotypes 1b, 1a, 2a, and 4d. The top inhibitor (3) has an IC_{50} value of 15 μ M against both subgenotypes 1b and 2a of the NS3/4A protease enzyme. Another inhibitor, 1, inhibits all four subgenotypes with moderate activity, showing highest activity for genotype 2a (24 μ M). The five inhibitors presented in this study could be valuable candidates for future hit to lead optimization. Additionally, enzyme–inhibitor interaction models presented herein provide key information regarding structural differences between the active sites of the NS3/4A protease of the HCV subgenotype 1a and 1b that might explain the variable inhibitory activity between subgenotypes of the small molecule inhibitors identified here.

INTRODUCTION

Hepatitis C is an infectious disease that can remain asymptomatic and undetected for decades in humans, ultimately resulting in chronic liver failure or liver cancer.^{1,2} Although the percentage of people affected by acute infection has declined considerably (>80%) over the past years,³ the dramatic proportion of patients that progress to chronic infection demonstrates a need for continued therapeutic development. The causative agent of Hepatitis C infection, the Hepatitis C virion, is a small enveloped virus with a positive single-stranded RNA-genome of the family Flaviviridae.^{4,5} The presence of the Hepatitis C Virus (HCV) in the liver triggers the human immune system, which leads to inflammation of the liver cells.⁶ Long-term infection can turn the liver cirrhotic, leading to liver failure, portal hypertension, and hepatocellular

carcinoma, causing eventual death.^{7,8} HCV is usually transmitted through exposure to infected blood. More than 170 million people worldwide are infected with hepatitis C.^{9,10} HCV has a wide variety of genotypes (currently with six major genotypes), with genotype 1 being most prevalent in United States, Europe, and Japan,^{11,12} and mutates rapidly because of a high error rate from the viral RNA-dependent RNA polymerase.

HCV encodes two proteases, the NS2 cysteine protease and the NS3/4A serine protease. The NS3 protease is responsible for proteolytic processing of the entire downstream region of

Received: April 4, 2012

Published: June 14, 2012

the Hepatitis C polyprotein to release the mature NS3, NS4a, NS4b, NSSa, and NSSb proteins.¹³

HCV protease inhibitors function by blocking the release of nonstructural proteins from the polyprotein, thereby preventing the formation of the HCV replicase complex. The N-terminal domain of the NS3 protease houses the catalytic motif of a chymotrypsin-like serine protease,¹⁴ with residues Ser139, His57, and Asp81 forming the catalytic triad.¹³ For proper function, NS3 requires NS4a as a cofactor and displays sensitivity to divalent metal ions.¹⁵ Introducing mutations to the NS3 protease region of the HCV genome abolishes infection, demonstrating that the NS3 serine protease is indeed essential for viral function and is hence an important target for antiviral drug development to combat HCV infection.¹³

The standard treatment has been the use of peginterferon α -2a (Pegasys) or peginterferon α -2b (Pegintron) in combination with Ribavirin.^{16–18} With combination therapy of peginterferon and ribavirin, ~45% of patients with genotype 1 infection and ~80% of patients with genotype 2 or 3 infection can achieve a sustained virologic response (SVR) with undetectable HCV RNA for 6 months post therapy. There is no current established treatment for patients with genotype 4, 5, or 6.^{19,20} The recent approval of the protease inhibitors VX950 (Telaprevir)^{21–23} and SCH503034 (Boceprevir),^{24,25} with others including ITMN-191 (Danoprevir)²⁶ and TMC 435350^{27–29} in the approval pipeline has opened promise of alternative treatment that may be more effective. Recently, both Telaprevir and Boceprevir were approved for the treatment against genotype 1 in combination with peginterferone and ribavirin. While therapeutic effectiveness is enhanced with pegylated alpha-interferon,^{21,25} there are several warnings associated with this form of treatment. Alpha-interferons, including PEG-INTRON, might induce fatality or life-threatening neuropsychiatric, autoimmune, ischemic, and infectious disorders.^{30–36} Thus, there is a continuing unmet medical need for the discovery and design of novel drugs for the treatment of HCV infection.

The NS3/4A protease active site is shallow, broad, and relatively featureless. The lack of well-defined deep pockets makes small molecule design quite difficult, with a high rate of false positives generated when using in silico methods. Thus far most inhibitors with binding affinity in the nanomolar range that have been reported are large macrocyclic peptidic inhibitors. The NS3/4A protease is a particularly interesting target with the active site sitting at the interface of two interacting domains, the helicase domain and the protease domain³⁰ (Figure 1), which makes drug design particularly challenging because the drug-binding target site is a protein–protein interacting surface. In the absence of any inhibitor, the C-terminus of the helicase domain occupies the protease active site. The side chain of Glu628 (numbering based on the full length structure of NS3) in this C-terminal region occupies the same space as the P2 to P4 linker in the class of inhibitors (MK-7009) introduced by Liverton et al., 2008³⁷ and McCauley et al., 2010.³⁸ The P2 pocket is the protease–helicase interface pocket. Recent research from Tibotec, Inc., has shown that the protein–protein interaction site binds to the inhibitor via induced fit in a multi-step binding process.^{29,38}

The rapidly mutating residues in the NS3/4A protease present another challenge for drug design and development by rendering the mutated enzyme resistant to current therapeutics. The amino acids that are prone to mutation in the binding site and enable drug resistance against Telaprevir and Boceprevir

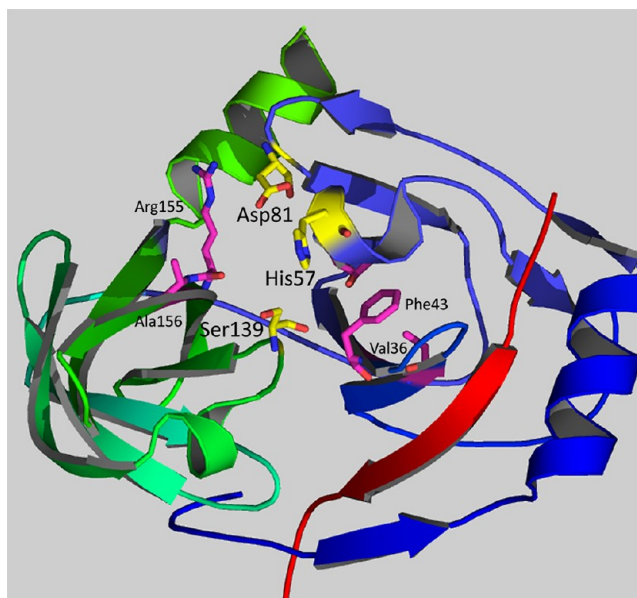


Figure 1. The active site of HCV NS3/4A serine protease is shown in a ribbon diagram. The protease domain is colored in blue, and the helicase domain is colored in green. The cofactor 4A (beta strand) is shown in red. The active site residues, Ser139, His57, and Asp81, sit on the protein–protein interaction surface and are shown in stick figures in yellow. The amino acids prone to mutation in the binding site enabling drug resistance against Telaprevir and Boceprevir are shown in stick figures in magenta (Val36, Phe43, Arg155, and Ala156).

are shown in stick figures in magenta (Val36, Phe43, Arg155, and Ala156) in Figure 1.

In this study, we implemented a two-tiered virtual screening protocol to identify novel small molecule scaffolds for the inhibition of the NS3/4A protease of HCV. A total of 196 compounds predicted to be active from the computational screening were purchased and tested in enzymatic assays, of which 17 compounds showed an inhibition rate of >50% at 50 μ M concentration. In the end, five small molecules that exhibit consistent IC_{50} values <100 μ M were considered to be hits (Table 1). We suggest that these five low micromolar non-macrocyclic compounds found from a first-pass computational screening effort may provide a foundation for future hit to lead optimization for the development of improved small molecule inhibitors for NS3/4A serine protease. In addition, we characterized the inhibitory data for these five molecules against genotype 1a, 1b, 2a, and 4d (Table 1). Interestingly, we also found that even within genotype 1, there are differences in how the subgenotype enzymes respond to the inhibitors. Four out of the five hits tested in subtype 1b were inactive against subtype 1a.

MATERIALS AND METHODS

Computational Methods. Two sequential in silico screening protocols were implemented in this study. Figure 3 summarizes the screening steps that we implemented to identify novel inhibitors of the HCV NS3/4A serine protease. The rationale for each of the two protocols and detailed execution of each step are described below.

At the time this work was initiated, there was no small molecule inhibitor data available for the NS3/4A protease because almost all then known inhibitors were macrocyclic structures. This introduced difficulties in training the computa-

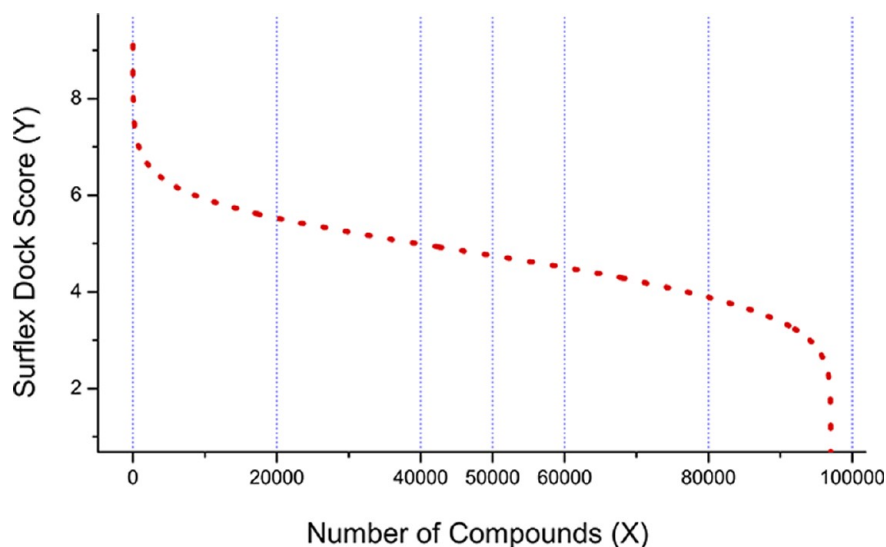


Figure 2. This plot of the incremental number of compounds versus the Surflex scoring function shows the progressive decrease in score; 50,000 compounds out of original 96,000 (Vernalis lead-like data set) from the ZINC database were retained for further analysis. The Y-axis indicates that compounds that scored below 5 probably do not fit well in the active site and can be discarded after the first sieving step.

tional models and scoring functions to attain reasonable results. Thus, we designed and utilized Protocol 1 (**S4-Pr1**) to derive a limited pool of small molecules enriched with enzyme inhibition data (IC_{50}) that could act as a training set enabling us to optimize our computational models and workflow for the ensuing detailed virtual screenings for this target. In Protocol 1, a small subset of compounds was docked using multiple algorithms and through a consensus ranking procedure, 100 molecules were selected for experimental testing. Compounds initially identified in this Protocol 1 (**S4-Pr1**), were then purchased and tested to establish a foundation for training the subsequent computational strategies. Protocol 2 (**S4-Pr2**) was then subsequently formulated and tested on the basis of the inhibitory data of the initial 100 compounds. Chronologically, **S4-Pr2** was designed and implemented after **S4-Pr1**. In particular, the choice of algorithms implemented in **S4-Pr2-2** were validated using the hits retrieved from Protocol 1 for training and optimizing the workflow. Methods used in Protocol 2 were validated on the basis of the inhibitory data of these 100 compounds. Subsequently, we purchased and tested an additional 96 compounds that were predicted to be active by Protocol 2.

In **Step 1**, we performed a shape filtering test, a fast and flexible ligand search to eliminate compounds that did not sterically fit in the active site. The Vernalis lead (96,990) compounds from the ZINC database version 7³⁹ were docked using version 2.3 of Surflex-Dock^{40,41} using the Surflex scoring function for a rapid first screen. The ligand-removed or apo forms of the proteins (PDB id: 4A92 for genotype 1b and PDB id: 2GVF for genotype 1a) were used as the starting target structures for all docking studies (screening and binding site comparison studies, respectively). In Figure 2, the plot of compounds vs Surflex scoring function shows a relatively few compounds with good fit, followed by a slow tail-off in docking score; the top 50,000 compounds out of the starting 96,000 from the ZINC database were retained for clustering in step 2.

In **Step 2**, we selected only scaffold representative compounds to avoid redundancy and to expedite the computational time for the more rigorous screening algorithms in the following steps. Clustering of 50,000 compounds from

Step 1 was performed using the Reciprocal Nearest Neighbor (RNN) method in Sybyl 8.2 (Tripos International, St. Louis, MO) with a test radius of 0.6 and neighbor exclusion radius of 0.85. This reduced the ligand set to 39,000 scaffold representative compounds.

In **Step 3**, this resulting set of 39,000 compounds was docked using the genetic algorithm of GOLDv4.1 (Cambridge, UK)^{42,43} and the GoldScore scoring function with default virtual screening settings. The number of islands was 5, with a population size of 100, number of operations of 100,000, a niche size of 2, and a selection pressure of 1.1. Early termination was allowed if the top 3 solutions were within 1.5 Å of each other. The binding site was defined as 10 Å around the cognate ligand in the X-ray crystal structure. This step was designed to obtain the initial docked poses for the ligands in the protease active site (required for the following rounds of AMBER Score based rescoring that minimizes an already docked ligand-bound protein pose before and after a short MD run to incorporate protein flexibility and entropic contributions) and as another filtering step to reduce the hits to a smaller set of top scoring compounds for training and testing.

To “bootstrap” the process, we chose a subset of 5,000 top scoring compounds from the Gold docking process to select an initial set of compounds for testing. Specifically, in **Step 4**, the ligand set of 39,000 compounds was split into two parts: in Protocol 1 (**S4-Pr1**) we worked with the subset of 5,000 top scoring compounds from the Gold docking, and in Protocol 2 (**S4-Pr2**), we screened the remaining 34,000 compounds.

In **S4-Pr1-1**, these 5,000 compounds were redocked using three available docking programs with the goal of choosing ligands that ranked consistently well in all three docking algorithms. We chose state of the art flexible docking algorithms for this consensus rank-based selection process, specifically including the AMBER Score module of DOCK in order to accommodate receptor flexibility and induced fit because the Induced Fit docking module of GLIDE is relatively computationally expensive particularly for virtual high-throughput screening purposes. We used standard GLIDEv5.0^{44,45} (SP scoring function) from Schrodinger, LLC., Autodock4.2,⁴⁶ and the AMBER Score⁴⁷ module of Dock6.2 to obtain a consensus

top ranking hit list in step **S4-Pr1-1**. The docking from GLIDE and Autodock were conducted using default settings on the GOLDV4.1 generated docked conformations of the ligands. In GLIDE, the receptor was prepared using the protein preparation wizard, with all waters removed. Standard precision docking was used for all Glide docking runs, with default settings for all other parameters and no constraints or similarity scoring applied. Default values were accepted for van der Waals scaling. In Autodock, the ligand set and the receptor were prepared using Autodock Tools with default parameters. The AMBER Score module of DOCK was run for 100 minimization cycles and 50 ps of molecular dynamics (MD) for each of the 5,000 molecules. The ligands were parametrized using the AMBER molecular mechanics force field called from within DOCK6.2. A prepared script runs the antechamber program to determine the semi-empirical charges, such as AM1-BCC,⁴⁸ for each ligand. It runs the tleap⁴⁹ program to create parameter and coordinate files for each ligand using the GAFF^{50,51} force field and atom types.

From the results of the three docking algorithms, a master list was generated with the compounds that ranked consistently well in all three methods. In step **S4-Pr1-2**, the top 20% of compounds from each of the docking runs were combined and the common compounds were chosen and reranked on the basis of their highest common rank. In step **S4-Pr1-3**, the top 100 compounds from this consensus list were purchased for experimental testing in enzymatic assays. The percent inhibition at 50 μ M and IC₅₀ values were determined for these 100 compounds.

Results from the **S4-Pr1** model-based predictions showed that the ranks assigned by AMBER Score correlated best with the experimental affinity data at 13% compared to the rather poor performance by GoldScore, GlideScore, and Autodock generated rankings. Hence, in the first step of Protocol 2, **S4-Pr2-1**, the remaining 34,000 compounds from Step 4 were first rescored using the AMBER Score module of Dock6.2 from their docked conformation as generated from GOLDv4.1 (from Step 3) to allow some extent of protein flexibility to be accommodated through induced fit that seemed critical for obtaining a decent ranking correlation between the experimental and predicted data. The minimization and MD were run for a longer period this time (500 minimization cycles and 200 ps of MD).

Recently, 18 available scoring functions were assessed for their accuracy to predict binding affinities and to rank-order compounds by their affinity to cocrystallized proteins.⁵² The eHiTS Score (SymBioSys, Inc., Toronto, Ontario)⁵³ came out at the top in this comprehensive scoring evaluation article. eHiTS 2009 is thus a powerful tool for docking and scoring potentially active molecules in a way that is well correlated with their binding affinity. The next step in our screening workflow was the selection of a scoring function that could improve the rank correlation coefficients between predicted active compounds and their experimental binding affinity data. Validating the power of eHiTS to reliably direct rational drug discovery efforts and prioritizing compounds for experimental testing using the data of the 100 compounds from Protocol 1 was the obvious next step. The 'bindener' flag of eHiTS Score was used because the goal was to rank order active compounds and not focus on the binding mode (RMSD) at this point. This option calls a slightly modified scoring scheme that produces improved scores and hence binding affinity correlations. The Pearson's correlation coefficient obtained between the ranks of the

predicted active compounds (algorithm-based score derived) and the actual compound ranks (IC₅₀ value derived) of the training set of 100 compounds are as follows: Rescoring the 100 compounds using eHiTS improved the correlation between the ranks of the compounds to 22% from 9% in the original GoldScore-based docking (from **Step3**), the AMBER Score of DOCK6.2-based rescoring was at 13%, and the GLIDES.0-based rank correlation was 11%. Hence, in keeping with our goal of selecting the algorithm that can best maximize our probability of obtaining high affinity binders, we chose eHiTS for the final step of Protocol 2.

In step **S4-Pr2-2**, the top ranked 15,000 compounds from step **S4-Pr2-1** were rescored using eHiTS.⁵³ A cumulative hit list of 300 compounds was generated on the basis of the top ranked hits from the eHiTS rescoring run and then manual inspection was used to eliminate several potentially reactive compounds in step **S4-Pr2-3** to select 96 compounds. At the end of step **S4-Pr2-4**, 96 compounds were selected for purchase and experimental evaluation and IC₅₀ determination.

In summary, the purpose of Protocol 1 (a consensus docking based strategy to maximize our confidence in a predicted true positive "hit") was to derive a pool of compounds enriched with inhibitory data, which could be used to establish a foundation for training the subsequent computational strategies. It was known from a literature survey that binding of compounds to the NS3/4A protease involved a multi-step induced fit process; this prompted us to include AMBER Score (which allows protein–ligand flexibility through minimization and molecular dynamics of the complex) as one of our screening methods. Consequently, the AMBER Score-based scoring function also correlated the ranks of compounds best with the experimental affinity data obtained from the pool of compounds tested from Protocol 1. This led us to utilize AMBER Score as the first screening step for Protocol 2. The logic behind introducing a second screening step in Protocol 2 was to utilize a scoring function that could improve the prediction of binding affinities and rank-order compounds by their affinity to cocrystallized proteins most accurately. eHiTS 2009 was validated as a powerful tool that could potentially achieve this. eHiTS could reliably direct rational drug discovery efforts and prioritize compounds through virtual screening for experimental testing and was hence the preferred algorithm for our second filtering step in Protocol 2.

All pocket druggability tests were performed using the SiteMap^{54–56} program of Schrodinger Version 8.5. The binding site region atoms (defined by the position of the docked ligand in the enzyme structures) plus an additional 8 Å buffer was included for evaluating the binding site. The OPLS-2005 force field and its more restrictive definition of hydrophobicity were used for all calculations.

Experimental Testing of Virtually Screened Compounds. Plasmid construction and purification of full length HCV NS3/4A genotype 1b (HCV polyprotein residues 1027–1657) were done as previously described.⁵⁷ Compounds predicted by virtual screening were purchased from Chembridge, Asinex, IBScreen, ChemDiv, Life Chemicals, and Enamine and were prepared as 10 mM stocks in 100% DMSO and stored in desiccated condition at –30 °C. All compounds were diluted to 50 μ M final concentration and distributed to 384-well black low-volume microplates (Corning, Inc.). The NS3/4A (10 nM) enzyme was added to the wells and incubated for 10 min at room temperature in assay buffer (50 mM Tris, pH 7.6, 0.5% Chaps, 20% glycerol, 2 mM GSH,

and 0.01 mg/mL BSA). The reaction was initiated by adding 1 μM (final) of substrate. Fluorescence intensity (492/520 nm, excitation/emission) was monitored continuously for at least 6 min with a POLARstar OPTIMA microplate reader (BMG LABTECH). All assays were done in triplicate, and compounds with over 50% inhibition at 50 μM were selected as hit compounds.

IC₅₀ Value Determination by Dose Response Curve.

Hit compounds that showed over 50% inhibition at 50 μM were selected for further evaluation. IC₅₀ values were measured in triplicate in the same final concentration of enzyme and substrate as the single concentration assay described above in assay buffer (50 mM Tris, pH 7.6, 0.5% Chaps, 15% glycerol, 2 mM GSH, and 0.01 mg/mL BSA). A series of compound concentrations (0–200 μM final concentration) in 100% DMSO were prepared in a 384-well plate. Twenty microliters of enzyme solution was distributed to wells, and 0.5 μL of varying concentration of compounds were added and incubated for 5 min. The enzyme reaction was initiated by adding 5 μL of the substrate, and its activity was continuously monitored for at least 6 min. The IC₅₀ values were calculated by fitting with the three parameter Hill equation, $y = V_{\text{max}}((x^n)/(IC_{50}^n + x^n))$, with OriginPro 8.5 (OriginLab, Inc.) where y is percent inhibition, x is inhibitor concentration, n is the slope of the concentration–response curve (Hill slope), and V_{max} is the maximal inhibition from two to four independent assays.

Reversibility of Inhibition. Reversibility of hit compounds was determined by dilution methods. The full length NS3/4A enzyme complex was prepared as 25-fold (250 nM) of the assay concentration and was incubated with screened compounds at 25-fold of each IC₅₀ concentration for 30 min at room temperature in assay buffer containing 50 mM Tris, pH 7.6, 0.5% Chaps, 15% glycerol, 2 mM GSH, and 0.01 mg/mL BSA in a final volume of 100 μL . A NS3/4A enzyme with the same volume of DMSO in place of each compound was also prepared in a same way as a control. A NS3/4A activity was measured in the same way as IC₅₀ measurements. Then enzyme–inhibitor solution was diluted 25-fold and 50-fold and incubated for 30 min before measuring % recovery of the enzyme activity. All reversibility assays were performed in triplicate.

Inhibition Assay with Other HCV Genotypes and Subgenotypes. Inhibitory activity of all the reversible hit compounds from virtual screening were tested against three other HCV genotypes (1a, 2a, and 4d) along with control (genotype 1b). The three HCV NS3/4A protease genotypes were purchased from Anaspec. Continuous kinetic assays were performed against all hit compounds in the same way as the genotype 1b NS3/4A. Full length NS3/4A (genotype 1b) was tested under the same conditions as a control for direct comparison for each plate.

RESULTS AND DISCUSSION

In Silico Screening. In Figure 3, we summarize the in silico screening steps implemented in this study to identify novel therapeutic compounds against the NS3/4A serine protease of HCV.

Through Protocol 1, we derived a pool of small molecules enriched with enzyme inhibition data (IC₅₀) that could act as a training set enabling us to optimize our computational models and workflow for the ensuing virtual screenings for this target. Hence, a small subset of 5,000 compounds from Step 3 (Figure 3) of the GOLD-ranked data set were redocked using three

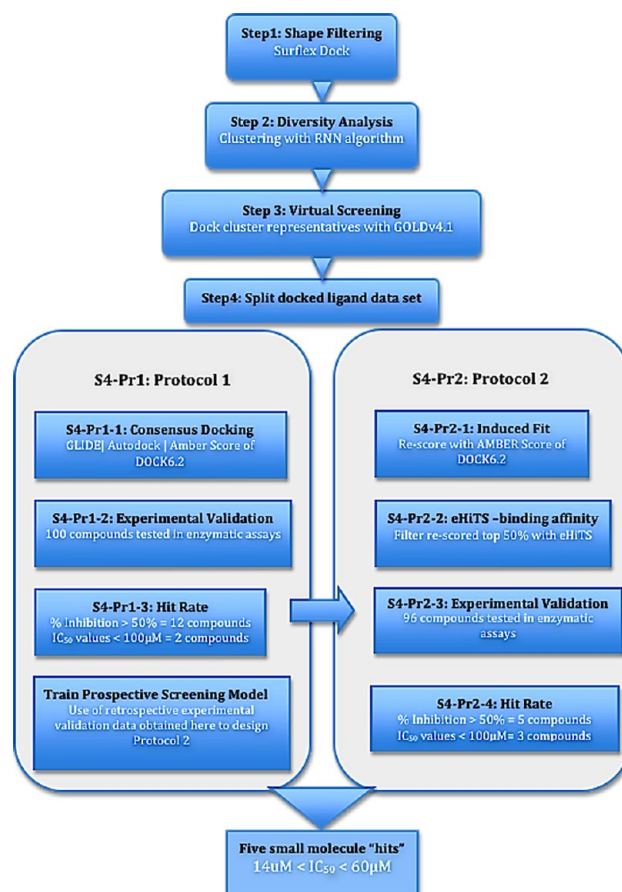
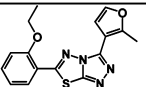
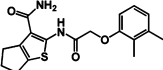
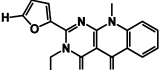
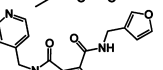
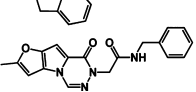


Figure 3. This workflow describes the in silico screening Protocols 1 and 2 implemented in this study. Protocols 1 and 2 identified five new non-macrocyclic small molecule inhibitors.

different algorithms: GLIDE,^{44,45} Autodock4.2,⁴⁶ and AMBER Score⁴⁷ from DOCK6.2. A consensus list of top scoring compounds among the three methods, Autodock4.2, GLIDE v5.0, and AMBER Score module of DOCK6.2 was generated and experimentally tested (2% of 5,000 compounds). An initial hit rate of 12% was achieved, with 12 compounds showing greater than 50% inhibition at 50 μM concentration. Two compounds, 1 (39 μM) and 2 (36 μM), from Protocol 1 subsequently showed moderately promising IC₅₀ values in enzymatic assays in the medium micromolar range (Table 1). Compound 1 was found to be active against all four subgenotypes (1b, 1a, 2a, and 4d) tested here (Table 1). Compounds 1 and 2 are the only molecules in this study that are active against genotype 4d with an inhibitory activity of less than 60 μM .

Protocol 2 was designed and implemented following Protocol 1 and was validated on the inhibitory data of the ligand set of 100 compounds that were purchased and tested from Protocol 1-based predictions. The AMBER Score rescoring improved the Pearson's correlation coefficient between the experimental and predicted ranking of these 100 compounds from 9% in original GoldScore based comparison to 13%. Hence, in the second protocol, the remaining 34,000 lead-like compounds from Step 4 in Figure 3 were first rescored with the AMBER Score module of DOCK6.2 to enable induced fit. (We chose not to include the initial 5,000 compound subset in this step, as it had already been "mined", but it could also have quite reasonably been included.) At this point, our goal

Table 1. Five Hit Compounds Identified from in Silico Screening Methods (Protocol 1 and Protocol 2) and Their IC₅₀ Value Comparison with Multiple HCV NS3 Genotypes

Compound and (Vendor ID)	Structure	Vendor	IC ₅₀ (μM)			
			sub-genotype 1b	sub-genotype 1a	sub-genotype 2a	sub-genotype 4d
1 (90227550)		Chembridge	39 ± 6	39 ± 24	24 ± 2	56 ± 25
2 (9014056)		Chembridge	36 ± 6	> 100	> 100	31 ± 17
3 (STOCK6S-01378)		IBScreen	15 ± 1	> 100	15 ± 12	> 100
4 (STOCK6S-03567)		IBScreen	50 ± 8	> 100	37 ± 41	> 100
5 (C784-0030)		ChemDiv	24 ± 4	> 100	> 100	> 100

was the careful selection of a scoring function that could correlate the rank-order of compounds to their binding affinity in the most accurate way, maximizing our “hit” rate. A literature survey pointed toward the scoring function implemented in eHiTS Score, 2009.⁵² The eHiTS rescoring further improved the correlation between the ranks of the 100 compounds to 22%. Hence, eHiTS was chosen for the final filtering steps. The top 50% of the rescored molecules obtained from induced fit docking were redocked using eHiTS⁵³ as a third filtering step; the bottom 50% showed poor scores and were excluded as unlikely to be hits. Approximately, 15,000 compounds were passed to eHiTS for docking and rescoring. The top 96 compounds were then purchased and tested in enzymatic assays. We further determined the IC₅₀ values for those showing over 50% inhibition, with three, namely, **3**, **4**, and **5**, showing promising inhibitory data with IC₅₀ values between 14 and 50 μM. Additionally, **3** and **4** are also active against genotype 2a.

Modeled Protein-Inhibitor Interactions and Their Implications. The docked orientations (with rigid receptor using GLIDEv5.0) of the five small molecule binders identified in this study are shown in the active sites of NS3/4A protease for subgenotypes 1b and 1a in Figure 4A and B. The goal was to identify the structural differences induced by the non-conserved residues in the binding site of the protease domain that dictate the binding affinity differences of the inhibitors. From Table 1, we note that except for **1** all the other four compounds that are active in HCV subtype 1b are inactive in subtype 1a. Although the pairwise sequence alignment is 90% conserved, there are 19 residue substitutions in the binding site from 1a to 1b, including I132 V and T42S among others. Figure 4 helps elucidate the structural diversity of the ligand-binding site induced by these substitutions that might partially explain the discriminatory binding affinity data between the two genotype 1 subtypes.

In Figure 4A, we observe a horizontally shaped binding pocket defined by the residues Phe154, Ala157, Leu135, Thr108, and Val132 on the left and His57, Ser42, and Phe43 to the right in the NS3/4A binding site of genotype 1b. This forms a favorable environment for ligands with a bulky hydrophobic core to be aligned optimally in the site. In

contrast, the larger residues of Ile132 and Thr42 on either side replacing Val132 and Ser42 in genotype 1a convert the binding site into a more longitudinal space. The volume and surface area of the binding site in 1a was calculated to be 192.4 and 147.5 Å², respectively, using the Castp server.⁵⁸ The calculation includes the volume surrounded by residues Gln41 to Phe43, Val55, His57, Gly58, and Gly137 to Ser139, Phe154, and Arg155. In contrast, the analogous binding site in genotype 1b is much smaller and more compact, measuring 112.9 Å³ in volume (V) and 101.3 Å² in surface area (SA). This site is defined by surrounding residues Gln41 to Phe43, Val55, His57, Gly58, and Ser138 to Gly140 and Phe154. Hence, these statistics suggest that the NS3/4A protease of subtype 1a has a more open and broad active site compared to that of subgenotype 1b. The surface area to volume ratio (SA:V) for 1a was 0.76 Å⁻¹ compared to 0.9 Å⁻¹ for 1b. For a given shape, the SA:V is inversely proportional to size. In this case, the higher the ratio, the more well-defined and compact with grooves and curves the pocket is, whereas, the lower the ratio, the more open and featureless it is. This probably suggests that the relatively open active site in 1a is unable to bind tightly to any of these “small” molecule inhibitors and hence they exhibit poor inhibition. In addition, the side chain rotamers of residues Arg155, His57, and Gln41 differ between the two enzyme X-ray crystal structures, which add to the differing topologies of the binding sites, introducing more steric hindrance in the case of 1a.

Pocket Druggability. The nature of the binding pockets of NS3/4A protease in the subgenotypes 1a and 1b were quantified and evaluated using the SiteMap algorithm.^{54–56} Table 2 illustrates the comparative differences between the two sites in terms of SiteScore, Druggability Score, exposure, enclosure, contact, balance (between hydrophobicity and hydrophilicity), and don/acc ratio. It is well-known from current literature that the balance between the hydrophobic and hydrophilic nature of the pocket is perhaps one of the most important measures for testing pocket druggability, i.e., druggable binding sites have a larger proportion of apolar surface than undruggable ones.^{56,59–63} Polar interactions are dependent on the environment, and a hydrogen bond is strengthened in hydrophobic surroundings. From Table 2, we

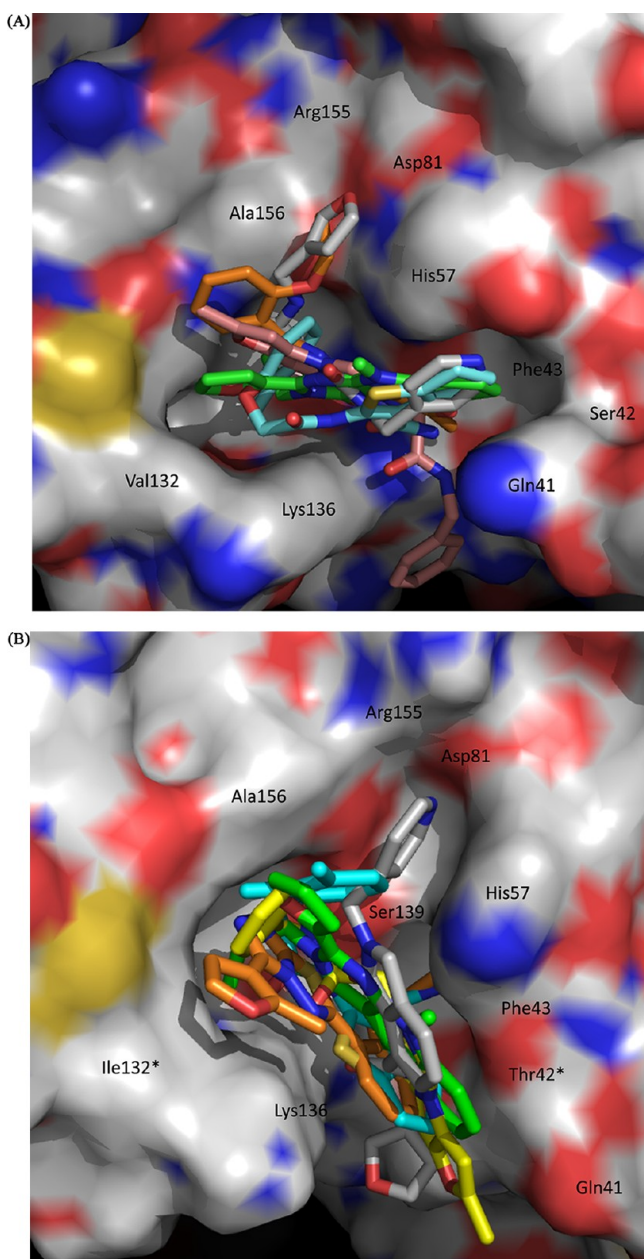


Figure 4. Differentiating features of the NS3/4A protease active site between subgenotypes 1b and 1a. (A) In subgenotype 1b, the active site shows a horizontal binding space, with residues Phe154, Ile153, Val132, Leu135, Val107, and Thr108 to the left, forming a hydrophobic pocket and residues Phe43 and His57 to the right, enabling another strong hydrophobic environment. However, surface residues Val132 and Ser42 are key players in determining the shape of the ligand-binding site. The docked conformations of the top five inhibitors from this study are shown above. (B) The NS3/4A active site of the subgenotype 1a is seen narrower and more longitudinal in shape. The relatively larger residues of Ile132 and Thr42 (marked with *), instead of Val132 and Ser42 in subgenotype 1a partially change the shape of the ligand-binding site and hence the alignment of the docked molecules in the site. (Discussion in the text).

note that the polar/apolar index termed “Balance” is much higher for 1b than for 1a. Also, the exposure and enclosure terms, which measure the amount of exposure to solvent and degree of enclosure by the protein, respectively, indicates that 1b is a deeper pocket (with increased binding site curvature for maximal interaction of the protein with the ligand) than is 1a. It has been shown that less druggable sites are longer, narrower,⁵⁹ and more open.^{56,60} In addition, a higher “Contact”, which is a measure of the average grid contact strength with the protein, quantitatively accounts for more protein–ligand contact in 1b than in 1a. Also, from a rigid receptor docking analysis, the ligands tend to score higher in the active site of the protease of subtype 1b compared to 1a. We thus conclude that the binding site of 1a is indeed more open, narrower, less complex, and unfavorable for binding small molecules than the binding site in 1b.

Induced Fit Docking Analysis. In Figure 5A–J, we illustrate the details of the interactions of the five compounds that show inhibitory data in at least one or more genotypes of HCV. The docked conformations of all five compounds are shown in the active sites of the subgenotype 1b NS3/4A serine protease. A standard docking exercise with a static receptor structure did not allow any of the compounds to form H-bonds with any of the catalytic residues, although they were all able to make a favorable interaction with the backbone amide of Gly137. However, when the same five compounds were docked allowing receptor flexibility and induced fit (using Schrödinger’s Induced Fit docking (IFD)^{64,65} module), all the compounds were able to make at least one H-bond with the backbone amide group of catalytic Ser139 or the side chains of His57 and Asp81. Information about each inhibitor and its protein interaction are detailed in the figure caption.

Protein–ligand interactions of the five hits were modeled in the active site using Induced Fit Docking of Glidev5.8 and are shown in Figure 5A–J. Initially, the compounds were docked into the rigid receptor structure with reduced van der Waals radii and an increased Coulomb–vdW cutoff. For each pose, the nearby side chains were reoriented using Prime3.1^{66–68} structure prediction to accommodate the ligand. Residues within 5 Å of the docked ligand were chosen for minimization and refinement (34 residues were selected), and only complex structures below 30 kcal/mol of energy were filtered out for further processing. These low-energy receptor structures were used to redock the ligands, and the resulting 105 complex structures were analyzed on the basis of their reranked GlideScores. The best scoring poses of the top five compounds from this study in complex with their receptor conformations are shown in Figure 5A–J. Compound 1 is the only compound that is active against genotype 1a with comparable affinity, and this could perhaps be attributed to the fact that 1 docks exactly in the same fashion in the active sites of NS3/4A in both subgenotypes 1a and 1b.

CONCLUSIONS

Attempts to identify inhibitors of the HCV NS3/4A serine protease through in silico methods have been a major challenge

Table 2. Quantitative Comparison of the Nature of the Binding Site between the NS3/4A Proteases of Subgenotypes 1a and 1b

Enzyme	SiteScore	DScore	Exposure	Enclosure	Contact	Balance	Don/Acc
NS3/4A subgenotype 1a	0.82	0.81	0.60	0.64	0.84	0.53	0.78
NS3/4A subgenotype 1b	0.84	0.82	0.55	0.68	0.92	0.64	1.01

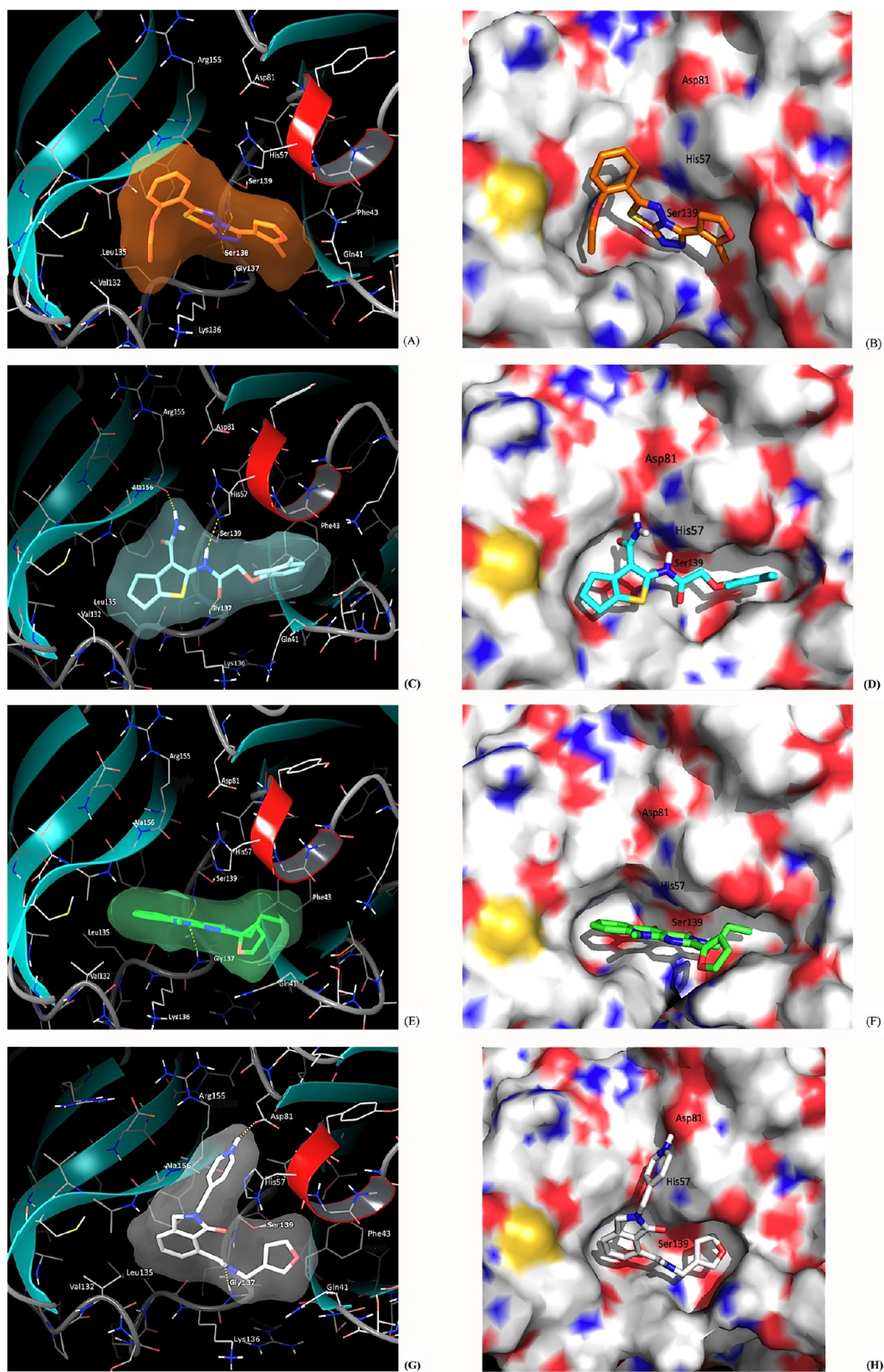


Figure 5. continued

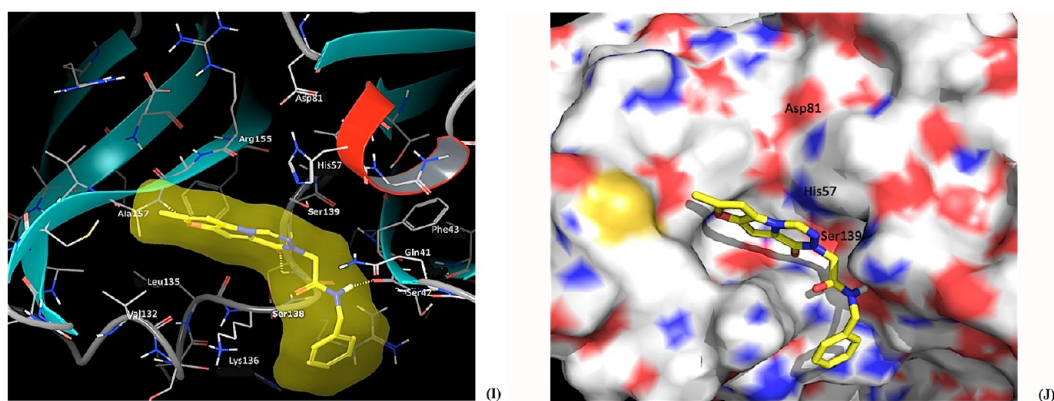


Figure 5. The binding mode of the five compounds that exhibit inhibition against one or more HCV genotypes are shown in the active site of NS3/4A serine protease in a ribbon and stick diagram (left) and in surface representation (right) for subgenotype 1b (PDBid: 4A92) as obtained from Induced Fit Docking. (A–B) The docked model of **1** (in pale orange) shows that the methylfuran moiety is able to nicely fit in a favorable environment made by Phe43, Ser42, and Gln41, while the ethyl phenyl ether group docks into the small hydrophobic pocket formed by Leu135, Val132, and Ala157, in addition to pi–pi stacking with Phe154. The compound is also able to make two H-bond interactions with the backbone amide groups of Ser138 and Ser139. (C–D) The model of **2** (in light blue) shows its potential to make three H-bonding interactions with the backbone amides of residue Gly137, the side chain of His57, and the carboxyl oxygen of Ala156. The compound optimally aligns itself in the horizontal active site by placing the o-xylene moiety in a hydrophobic environment toward Phe43. (E–F) The docked model of **3** (in green) shows the optimal van der Waals fit of the compound in the hydrophobic pocket with the ethyl group lodged inward toward Phe154. The compound is also able to make two hydrogen bonds with the backbone amides of Gly137 and Ser138 (none of which are highly mutable residues). (G–H) The **4** (in white) bound protease interactions are shown to be dominated by two possible H-bond interactions with the backbone of residue Gly157 and side chain of catalytic Asp81, anchoring the compound at two ends allowing the furan group to pi–pi stack with the side chain of Phe43 and the phenylamine group to pi–pi stack with His57. (I–J) The docked model of compound **5**, which is active only against subgenotype 1b, is shown (in yellow) to make several H-bonds with Ser42, Ser138, Ser139, and Ala157. This anchors the compound on both ends and in the middle to assume a nice fit in the binding pocket.

of HCV drug discovery and design. The active site of the enzyme is fairly broad and shallow, which makes docking methods particularly challenging for the discovery of “hits”. Although there are several recently published macrocyclic inhibitors that are now approved for treatment or are in advanced stage clinical trials, most of the existing treatments through monotherapy or combination therapy for the HCV infection have severe side effects. Thus, there is a continuing unmet need for the discovery and design of novel therapeutics for this target.

In this study, we identified five new small molecule inhibitors that show low micromolar range activity against the NS3/4A serine protease of HCV. Two different *in silico* screening protocols are presented here and are summarized in Figure 3. The workflow involves a two-tiered screening approach where the results (screened and tested molecules in enzymatic assays) from one stage, termed Protocol 1, are used as a training set to optimize the computational workflow in Protocol 2. Through this *in silico* screening strategy, starting from 96,990 compounds, we identified five small molecule inhibitors of the NS3/4A protease validated in enzymatic assays. Table 1 shows the structures and IC_{50} values of these inhibitors. The top binder, **3**, exhibits an IC_{50} value of 15 μ M and is active against genotypes 1b and 2a. Inhibitor **1** is active against all four subgenotypes tested here, including 1b, 1a, 2a and 4d. Compounds **1** and **2** are the only compounds that are active against genotype 4d.

In the absence of known non-macrocyclic small molecule inhibitors for the NS3/4A serine protease, it was difficult to train and test a computational model that would select active molecules with a moderate correlation between rank-order and binding affinity through a virtual screening exercise. Protocol 1, shown in Figure 3, was designed and implemented to derive a pool of compounds enriched with IC_{50} values that could be

used for optimizing and validating Protocol 2, which in turn could be used to predict “hits” with a potentially higher rate of enrichment. In Protocol 1, a consensus scoring method (based on the highest common rank) between three docking algorithms was used to select a set of 100 predicted binding compounds that were purchased and experimentally tested for their inhibition against the HCV NS3/4A serine protease. Induced fit docking implemented through the AMBER Score module of DOCK 6.2 provided the best correlation between the predicted and experimental ranks of these 100 compounds. This was not surprising given recent publications that indicate inhibitors of NS3/4A serine protease perhaps bind in a multi-step induced fit binding process.^{29,38}

Protocol 2 hence incorporated protein flexibility in the first step of the screening process with a short minimization and MD run for each of the protein–ligand complexes using the AMBER Score module of DOCK6.2. The top 50% of the ranked compounds were then rescored using eHiTS Score. According to a recent survey article that compares 18 scoring functions from various docking software,⁵² the scoring function potentially prioritizes molecules in a way that is well correlated with their target binding affinity. We validated the power of eHiTS on the 100 compounds from Protocol 1 and obtained a rank correlation of 22% (between predicted and experimental data), besting all the other docking algorithms used in this study. Hence, as a second filtering step in Protocol 2, we screened the ligand set using eHiTS Score. The resulting top-ranking 96 compounds were experimentally tested, resulting in the identification of three additional small molecule inhibitors **3**, **4**, and **5** with IC_{50} values in the range of 14 and 50 μ M. In summary, through the accommodation of protein flexibility in step 1 and the inclusion of an improved scoring function that rank orders active compounds well correlated to their experimental inhibitory data in step 2, we identified low

micromolar range non-macrocyclic small molecule inhibitors of the NS3/4A serine protease of HCV.

Although some studies have failed to correlate HCV genotype and clinical outcome,^{69,70} others have demonstrated that the subtype seems to influence the clinical course and the therapeutic outcome.^{71–73} The HCV-genotype 1 is prevalent over five continents and current drugs show poor efficacy against this genotype. Subgenotype 1a is mostly found in North America, South America, and Australia, while 1b is common in Europe and Asia. Thus, we compared the structural differences between these two subtypes to better understand the challenges associated with designing drugs for either of the subgenotypes 1a and 1b. The pairwise sequence identity between the NS3 protease domain (1–181aa) of the subgenotype 1a with that of 1b is ~90% with 19 nonconserved residues.

Interestingly, four out of five molecules that tested active for subgenotype 1b were found to be inactive for subgenotype 1a. To better understand the underlying reason behind this variable affinity data between the two subgenotypes, molecular modeling and docking studies helped us isolate the key structural differences between the binding sites of the NS3/4A proteases. The volume/surface area calculations, docked alignment of the small molecule binders, binding pocket druggability tests, and identification of nonconserved amino acids that dominate the topology of the active sites, all indicate that the NS3/4A protease of subgenotype 1a is broader, more open, and longitudinal in binding space compared to the smaller, compact, more “druggable”, and horizontal binding space of subgenotype 1b.

In conclusion, in this study, we introduced and implemented a tiered computational screening protocol through which we identified five new non-macrocyclic small molecule inhibitors of the HCV NS3/4A serine protease. In addition, we characterized their inhibitory activity across four different HCV genotypes including subtypes 1a, 1b, 2a, and 4d.

AUTHOR INFORMATION

Corresponding Author

* E-mail: mjohnson@uic.edu. Telephone: +1-312-996-9114. Fax: +1-312-431-9303.

Present Address

[†]Joint IRB-BSC Program on Computational Biology, Institute for Research in Biomedicine, Barcelona, C/Baldiri Reixac 10, Barcelona 08028, Spain.

Author Contributions

[§]These authors contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Professor Ajay Jain, UCSF, for providing the Surflex program, SymbioSys, and especially Dr. Orr Ravitz for providing access to eHits, Professor Arthur Olson, Scripps, for providing access to Autodock, and UCSF for access to DOCK. This work was supported in part by the National Science Foundation through TeraGrid resources provided by the Loni and Steele resources under Grant Numbers TGMCB090040 and TG-MCB090168 (M.E.J.).

REFERENCES

- (1) Volk, M. L.; Tocco, R.; Saini, S.; Lok, A. Public health impact of antiviral therapy for hepatitis C in the United States. *Hepatology* **2009**, *50*, 1750–1755.
- (2) Mitchell, A. E.; Colvin, H. M.; Palmer Beasley, R. Institute of Medicine recommendations for the prevention and control of hepatitis B and C. *Hepatology* **2010**, *51*, 729–733.
- (3) Hoofnagle, J. H. Course and outcome of hepatitis C. *Hepatology* **2002**, *36*, 21–29.
- (4) Bartschlag, R.; Lohmann, V. Replication of hepatitis C virus. *J. Gen. Virol.* **2000**, *81*, 1631–1648.
- (5) Lindenbach, B.; Rice, C. Unravelling hepatitis C virus replication from genome to function. *Nature* **2005**, *436*, 933–938.
- (6) Ryan, K. J.; Ray, C. G. *Sherris Medical Microbiology*, 4th ed.; McGraw Hill: New York, 2004.
- (7) Beld, M.; Penning, M.; Putten, M.; van; Lukashov, V.; Hoek, A.; van den; McMorrow, M.; Goudsmit, J. Quantitative antibody responses to structural (Core) and nonstructural (NS3, NS4, and NS5) hepatitis C virus proteins among seroconverting injecting drug users: impact of epitope variation and relationship to detection of HCV RNA in blood. *Hepatology* **1999**, *29*, 1288–1298.
- (8) Colvin, H. M.; Mitchell, A. E., Eds.; *Hepatitis and Liver Cancer: A National Strategy for Prevention and Control of Hepatitis B and C*; Committee on the Prevention and Control of Viral Hepatitis Infections, Institute of Medicine, National Academies Press: Washington, DC, 2010.
- (9) Hepatitis C; Fact Sheet 164; World Health Organization: Geneva, Switzerland, 2000.
- (10) WHO. Hepatitis C: global prevalence (update). *Wkly. Epidemiol. Rec.* **1999**, *74*, 425–427.
- (11) Wasley, A.; Alter, M. J. Epidemiology of hepatitis C: Geographic differences and temporal trends. *Semin. Liver Dis.* **2000**, *20*, 1–16.
- (12) Alter, M. J. Epidemiology of hepatitis C virus infection. *World J. Gastroenterol.* **2007**, *13*, 2436–2441.
- (13) Love, R. A.; Parge, H. E.; Wickersham, J. A.; Hostomsky, Z.; Habuka, N.; Moomaw, E. W.; Adachi, T.; Hostomska, Z. The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. *Cell* **1996**, *87*, 331–342.
- (14) Miller, R. H.; Purcell, R. H. Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 2057–2061.
- (15) Kim, J. L.; Morgenstern, K. A.; Lin, C.; Fox, T.; Dwyer, M. D.; Landro, J. A.; Chambers, S. P.; Markland, W.; Lepre, C. A.; O'Malley, E. T.; Harbeson, S. L.; Rice, C. M.; Murcko, M. A.; Caron, P. R.; Thomson, J. A. Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell* **1996**, *87*, 343–355.
- (16) Cheng, J. S.; Bonis, A. L. P.; Lau, J.; Pham, Q. N.; Wong, B. J. Interferon and ribavirin for patients with chronic hepatitis C who did not respond to previous interferon therapy: A meta-analysis of controlled and uncontrolled trials. *Hepatology* **2001**, *33*, 231–240.
- (17) Ghany, M. G.; Strader, D. B.; Thomas, D. L.; Seeff, L. B. Diagnosis, management, and treatment of hepatitis C: an update. *Hepatology* **2009**, *49*, 1335–1374.
- (18) Ferguson, M. Current therapies for chronic hepatitis C. *Pharmacotherapy* **2011**, *31*, 92–111.
- (19) Modi, A.; Liang, T. Hepatitis C: A clinical review. *Oral Dis.* **2008**, *14*, 10–14.
- (20) El-Zayadi, A.; Attia, M.; Barakat, E.; Badran, H.; Hamdy, H.; El-Tawil, A.; El-Nakeeb, A.; Selim, O.; Saied, A. Response of hepatitis C genotype-4 naïve patients to 24 weeks of Peg-interferon-alpha2b/ribavirin or induction-dose interferon-alpha2b/ribavirin/amantadine: A non-randomized controlled study. *Am. J. Gastroenterol.* **2005**, *100*, 2447–52.
- (21) Kwong, A. D.; Kauffman, R. S.; Hurter, P.; Mueller, P. Discovery and development of telaprevir: an NS3-4A protease inhibitor for treating genotype 1 chronic hepatitis C virus. *Nat. Biotechnol.* **2011**, *29*, 993–1003.

- (22) Lin, K.; R.B. P.; A.D. K.; C. L. VX-950, a novel hepatitis C virus (HCV) NS3-4A protease inhibitor, exhibits potent antiviral activities in HCV replicon cells. *Antimicrob. Agents Chemother.* **2006**, *50*, 1813–1822.
- (23) Perni, R. B.; Almquist, S. J.; Byrn, R. A.; Chandorkar, G.; Chaturvedi, P. R.; Courtney, L. F.; Decker, C. J.; Dinehart, K.; Gates, C. A.; Harbeson, S. L.; Heiser, A.; Kalkeri, G.; Kolaczowski, E.; Lin, K.; Luong, Y. P.; Rao, B. G.; Taylor, W. P.; Thomson, J. A.; Tung, R. D.; Wei, Y.; Kwong, A. D.; Lin, C. Preclinical profile of VX-950, a potent, selective, and orally bioavailable inhibitor of hepatitis C virus NS3-4A serine protease. *Antimicrob. Agents Chemother.* **2006**, *50*, 899–909.
- (24) Venkatraman, S.; Bogen, S. L.; Arasappan, A.; Bennett, F.; Chen, K.; Jao, E.; Liu, Y. T.; Lovey, R.; Hendrata, S.; Huang, Y.; Pan, W.; Parekh, T.; Pinto, P.; Popov, V.; Pike, R.; Ruan, S.; Santhanam, B.; Vibulbhan, B.; Wu, W.; Yang, W.; Kong, J.; Liang, X.; Wong, J.; Liu, R.; Butkiewicz, N.; Chase, R.; Hart, A.; Agrawal, S.; Ingravallo, P.; Pichardo, J.; Kong, R.; Baroudy, B.; Malcolm, B.; Guo, Z.; Prongay, A.; Madison, V.; Broske, L.; Cui, X.; Cheng, K. C.; Hsieh, Y.; Brisson, J. M.; Prelusky, D.; Korfmacher, W.; White, R.; Bogdanowich-Knipp, S.; Pavlovsky, A.; Bradley, P.; Saksena, A. K.; Ganguly, A.; Piwinski, J.; Girijavallabhan, V.; Njorog, F. G. Discovery of (1R,5S)-N-[3-amino-1-(cyclobutylmethyl)-2,3-dioxopropyl]-3-[2(S)-[[[(1,1-dimethylethyl)-amino]carbonyl]amino]-3,3-dimethyl-1-oxobutyl]-6, 6-dimethyl-3-azabicyclo[3.1.0]hexan-2(S)-carboxamide (SCH 503034), a selective, potent, orally bioavailable hepatitis C virus NS3 protease inhibitor: A potential therapeutic agent for the treatment of hepatitis C infection. *J. Med. Chem.* **2006**, *49*, 6074–6086.
- (25) Sarrazin, C.; Rouzier, R.; Wagner, F.; Forestier, N.; Larrey, D.; Gupta, S. K.; Hussain, M.; Shah, A.; Cutler, D.; Zhang, J.; Zeuzem, S. SCH 503034, a novel hepatitis C virus protease inhibitor, plus pegylated interferon alpha-2b for genotype 1 nonresponders. *Gastroenterology* **2007**, *132*, 1270–1278.
- (26) Seiwert, S. D.; Andrews, S. W.; Jiang, Y.; Serebryany, V.; Tan, H.; Kossen, K.; Rajagopalan, P. T.; Misialek, S.; Stevens, S. K.; Stoycheva, A.; Hong, J.; Lim, S. R.; Qin, X.; Rieger, R.; Condroski, K. R.; Zhang, H.; Do, M. G.; Lemieux, C.; Hingorani, G. P.; Hartley, D. P.; Josey, J. A.; Pan, L.; Beigelman, L.; Blatt, L. M. Preclinical characteristics of the hepatitis C virus NS3/4A protease inhibitor ITMN-191 (R7227). *Antimicrob. Agents Chemother.* **2008**, *52*, 4432–4441.
- (27) Lin, T.-I.; Lenz, O.; Fanning, G.; Verbinnen, T.; Delouvroy, F.; Scholliers, A.; Vermeiren, K.; Rosenquist, Å.; Edlund, M.; Samuelsson, B.; Vrang, L.; Kock, H.; de Wigerinck, P.; Raboisson, P.; Simmen, K. In vitro activity and preclinical profile of TMC435350, a potent hepatitis C virus protease inhibitor. *Antimicrob. Agents Chemother.* **2009**, *53*, 1377–1385.
- (28) Raboisson, P.; Kock, H. de; Rosenquist, A.; Nilsson, M.; Salvador-Oden, L.; T.I. Lin, N. R.; Ivanov, V.; Wähling, H.; Wickström, K.; Hamelink, E.; Edlund, M.; Vrang, L.; Vendeville, S.; Vreken, W. V.; de McGowan, D.; Tahri, A.; Hu, L.; Boutton, C.; Lenz, O.; Delouvroy, F.; Pille, G.; Surleraux, D.; Wigerinck, P.; Samuelsson, B.; Simmen, K. Structure-activity relationship study on a novel series of cyclopentane-containing macrocyclic inhibitors of the hepatitis C virus NS3/4A protease leading to the discovery of TMC435350. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4853–4858.
- (29) Marcellin, P.; Reesink, H.; Berg, T.; Cramp, M.; Flisiak, R.; Vlierberghe, H. Van; Verloes, R.; Lenz, O.; Peeters, M.; Sekar, V.; Smedt, G. De Presented at the EASL 44th Annual Meeting; Copenhagen, 2009.
- (30) Weiland, O. Interferon therapy in chronic hepatitis C virus infection. *FEMS Microbiol. Rev.* **1994**, *14*, 279–288.
- (31) Manns, M. P.; et al. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: A randomized trial. *Lancet* **2001**, *358*, 958–965.
- (32) Fried, M. W.; et al. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N. Engl. J. Med.* **2002**, *347*, 975–982.
- (33) Hadziyannis, S. J.; et al. Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: A randomized study of treatment duration and ribavirin dose. *Ann. Intern. Med.* **2004**, *140*, 346–355.
- (34) Bacon, B. R.; et al. Retreating chronic hepatitis C with daily interferon alfacon-1/ribavirin after nonresponse to pegylated interferon/ribavirin: DIRECT results. *Hepatology* **2009**, *49*, 1838–1846.
- (35) Jensen, D. M.; et al. Re-treatment of patients with chronic hepatitis C who do not respond to peginterferon-a2b: A randomized trial. *Ann. Intern. Med.* **2009**, *150*, 528–540.
- (36) Poynard, T.; et al. Peginterferon alfa-2b and ribavirin: Effective in patients with hepatitis C who failed interferon alfa/ribavirin therapy. *Gastroenterology* **2009**, *136*, 1618–1628.
- (37) Liverton, N. J.; Holloway, M. K.; McCauley, J. A.; Rudd, M. T.; Butcher, J. W.; Carroll, S. S.; DiMuzio, J.; Fandozzi, C.; Gilbert, K. F.; Mao, S.-S.; McIntyre, C. J.; Nguyen, K. T.; Romano, J. J.; Stahlhut, M.; Wan, B.-L.; Olsen, D. B.; Vacca, J. P. Molecular modeling based approach to potent P2-P4 macrocyclic inhibitors of hepatitis C NS3/4a protease. *J. Am. Chem. Soc.* **2008**, *130*, 4607–4609.
- (38) McCauley, J. A.; McIntyre, C. J.; Rudd, M. T.; Nguyen, K. T.; Romano, J. J.; Butcher, J. W.; Gilbert, K. F.; Bush, K. J.; Holloway, M. K.; Swestock, J.; Wan, B. L.; Carroll, S. S.; DiMuzio, J. M.; Graham, D. J.; Ludmerer, S. W.; Mao, S. S.; Stahlhut, M. W.; Fandozzi, C. M.; Trainor, N.; Olsen, D. B.; Vacca, J. P.; Liverton, N. J. Discovery of vaniprevir (MK-7009), a macrocyclic hepatitis C virus NS3/4a protease inhibitor. *J. Med. Chem.* **2010**, *53*, 2443–2463.
- (39) 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.
- (40) Jain, A. N. Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* **2003**, *46*, 499–511.
- (41) Jain, A. N. Surflex-Dock 2.1: Robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J. Comput. Aided Mol. Des.* **2007**, *21*, 281–306.
- (42) Jones, G.; Willett, P.; Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, *245*, 43–53.
- (43) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.
- (44) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *1739*–1749.
- (45) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750–1759.
- (46) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
- (47) Graves, A. P.; Shivakumar, D. M.; Boyce, S. E.; Jacobson, M. P.; Case, D. A.; Shoichet, B. K. Rescoring docking hit lists for model cavity sites: Predictions and experimental testing. *J. Mol. Biol.* **2008**, *377*, 914–934.
- (48) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.
- (49) Case, D. A.; T.E. Cheatham, I.; Darden, T.; Gohlke, H.; Luo, R.; Merz, J. K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (50) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.

- (51) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general AMBER force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (52) Englebienne, P.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 4. Are popular scoring functions accurate for this class of proteins? *J. Chem. Inf. Model.* **2009**, *49*, 1568–1580.
- (53) Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, S.; Johnson, A. eHiTS: A new fast, exhaustive flexible ligand docking system. *J. Mol. Graph. Model.* **2007**, *26*, 198–212.
- (54) SiteMap, version 2.5, Schrödinger, LLC. 2011.
- (55) Halgren, T. A. New method for fast and accurate binding-site identification and analysis. *Chem. Biol. Drug Des.* **2007**, *69*, 146–148.
- (56) Halgren, T. A. Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.* **2009**, *49*, 377–389.
- (57) Lee, H.; Torres, J.; Truong, L.; Chaudhuri, R.; Mittal, A.; Johnson, M. . Reducing agents affect inhibitory activities of compounds: Results from multiple drug targets. *Anal. Biochem.* **2012**, *423*, 46–53.
- (58) Dundas, J.; Ouyang, Z.; Tseng, J.; Binkowski, A.; Turpaz, Y.; Liang, J. CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* **2006**, *34*, 116–118.
- (59) Hajduk, P. J.; Huth, J. R.; Fesik, S. W. Druggability indices for protein targets derived from NMR-based screening data. *J. Med. Chem.* **2005**, *48*, 2518–2525.
- (60) Cheng, A. C.; Coleman, R. G.; Smyth, K. T.; Cao, Q.; Souland, P.; Caffrey, D. R.; Salzberg, Anna C.; Enoch, S. Huang structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.* **2007**, *25*, 71–75.
- (61) Schmidtke, P.; Barril, X. Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. *J. Med. Chem.* **2010**, *53*, 5858–5867.
- (62) Gupta, A.; Gupta, A. K.; Seshadri, K. Structural models in the assessment of protein druggability based on HTS data. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 583–592.
- (63) Krasowski, A.; Muthas, D.; Sarkar, A.; Schmitt, S.; Brenk, R. DrugPred: A structure-based approach to predict protein druggability developed using an extensive nonredundant data set. *J. Chem. Inf. Model.* **2011**, *51*, 2829–2842.
- (64) Sherman, W.; Beard, H. S.; Farid, R. Use of an induced fit receptor structure in virtual screening. *Chem. Biol. Drug Des.* **2006**, *67*, 83–84.
- (65) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **2006**, *49*, 534–554.
- (66) Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. On the Role of crystal packing forces in determining protein sidechain conformations. *J. Mol. Biol.* **2002**, *320*, 597–608.
- (67) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. A hierarchical approach to all-atom protein loop prediction. *Proteins* **2004**, *55*, 351–367.
- (68) Prime, version 3.1, Schrödinger, LLC, New York, 2009.
- (69) Romeo, R.; Colombo, M.; Rumi, M.; Soffredini, R.; Ninno, E.; Del, Donato, M. F.; Russo, A.; Simmonds, P. Lack of association between type of hepatitis C virus, serum load and severity of liver disease. *J. Viral Hepatitis* **1996**, *3*, 183–190.
- (70) Zhou, S.; Terrault, N. A.; Ferrell, L.; Hahn, J. A.; Lau, J. Y.; Simmonds, P.; Roberts, J. P.; Lake, J. R.; Ascher, N. L.; Wright, T. L. Severity of liver disease in liver transplantation recipients with hepatitis C virus infection: Relationship to genotype and level of viremia. *Hepatology* **1996**, *24*, 1041–1046.
- (71) McOmish, F.; Yap, P. L.; Bow, B. C.; Follet, E. A. C.; Seed, C.; Keller, A. J.; Cobain, T. J.; Krusius, T.; Kolho, E.; Naukkarinen, R.; Lin, C.; Lai, C.; Leong, S.; Medgyesi, G. A.; Hejjas, M.; Kiyokawa, H.; Fukuda, K.; Cuyppers, T.; Saeed, A. A.; Al-Rasheed, A. M.; Lin, M.; Simmonds, P. Geographical distribution of hepatitis C virus genotypes in blood donors: An international collaborative survey. *J. Clin. Microbiol.* **1994**, *32*, 884–892.
- (72) Prati, D.; Capelli, C.; Zanella, A.; Mozzi, F.; Bosoni, P.; Pappalettera, M.; Zanuso, F.; Vianello, L.; Locatelli, E.; Fazio, C.; De; Ronchi, G.; Ninno, E.; Del; Colombo, M.; Sirchia, G. Influence of different hepatitis C virus genotypes on the course of asymptomatic hepatitis C virus infection. *Gastroenterology* **1996**, *110*, 178–183.
- (73) Takada, A.; Tsutsumi, M.; Zhang, S. C.; Okanou, T.; Matsushima, T.; Fujiyama, S.; Komatsu, M. Relationship between hepatocellular carcinoma and subtypes of hepatitis C virus: A nationwide analysis. *J. Gastroenterol. Hepatol.* **1996**, *11*, 166–169.