Multiple e-Pharmacophore Modeling, 3D-QSAR, and High-Throughput Virtual Screening of Hepatitis C Virus NS5B Polymerase Inhibitors

Patrisha Joseph Therese,[†] Dinesh Manvar,[‡] Sridevi Kondepudi,[†] Madhu Babu Battu,[†] Dharmarajan Sriram,[†] Amartya Basu,[‡] Perumal Yogeeswari,*,[†] and Neerja Kaushik-Basu*,[‡,§]

[†]Computer-Aided Drug Design Lab, Department of Pharmacy, Birla Institute of Technology & Science−Pilani, Hyderabad campus, Jawahar Nagar, Hyderabad−500078, Andhra Pradesh, India

[‡]Department of Biochemistry and Molecular Biology, New Jersey Medical School, Rutgers Biomedical and Health Sciences, 185 South Orange Avenue, Newark, New Jersey 07103, United States

[§]Rutgers Cancer Institute of New Jersey, 120 Albany Street, New Brunswick, New Jersey 08901, United States

Ⓢ *Supporting Information*

ABSTRACT: The hepatitis C virus (HCV) NS5B RNA-dependent RNA polymerase (RdRP) is a crucial and unique component of the HCV RNA replication machinery and a validated target for drug discovery. Multiple crystal structures of NS5B inhibitor complexes have facilitated the identification of novel compound scaffolds through in silico analysis. With the goal of discovering new NS5B inhibitor leads, HCV NS5B crystal structures bound with inhibitors in the palm and thumb allosteric pockets in combination with ligands with known inhibitory potential were explored for a comparative pharmacophore analyses. The energy-based and 3D-QSAR-based pharmacophore models were validated using enrichment analysis, and the six models thus developed were employed for high-throughput virtual screening and docking to identify nonpeptidic leads. The hits derived at each stage were analyzed for diversity based on the six pharmacophore models, followed by molecular docking and filtering based on their interaction with amino acids in the NS5B allosteric pocket and 3D-QSAR predictions. The resulting 10 hits displaying diverse scaffold were then screened employing biochemical and cell-based NS5B and anti-HCV inhibition assays. Of these, two molecules H-5 and H-6 were the most promising, exhibiting IC_{50} values of 28.8 and 47.3 μM against NS5B polymerase and anti-HCV inhibition of 96% and 86% at 50 μM, respectively. The identified leads comprised of benzimidazole (H-5) and pyridine (H-6) scaffolds thus constitute prototypical molecules for further optimization and development as NS5B inhibitors.
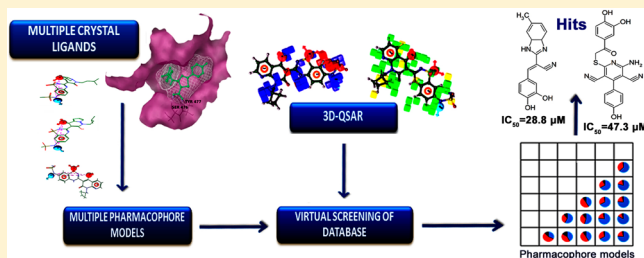
1. INTRODUCTION

Hepatitis C virus (HCV), a positive-strand RNA virus, is a significant global pathogen causing acute hepatitis and chronic liver disease, including cirrhosis and liver cancer. About 200 million people are chronically infected with HCV, of which ~350 000 die every year from hepatitis C-related liver diseases.[1] The current HCV therapy suffers from inadequate sustained viral response rate, rapid emergence of drug resistance, in particular for patients infected with genotype 1 HCV, along with significant side effects, resulting in a discontinuation of therapy.[2,3] Therefore, massive efforts are ongoing to identify new treatment modalities for HCV.

Replication of the positive strand RNA genome of HCV is mediated by the RNA-dependent RNA polymerase (RdRp) activity of its nonstructural protein NS5B, in concert with other HCV nonstructural proteins and host factors.[4,5] Given the absolute requirement of NS5B to synthesize nascent HCV RNA, NS5B represents an attractive target for the development of anti-HCV inhibitors.[6,7] Furthermore, host cells lack RdRp.

Therefore, an inhibitor that blocks NS5B RdRp activity should, in theory, have minimal or no effect on host biological processes. In addition, given the success paradigm of HIV-reverse transcriptase for drug development, efforts to develop HCV-directed antivirals have focused on this essential viral-encoded enzyme.

HCV NS5B is a 66 kDa protein with a hydrophobic C-terminal membrane anchorage domain, which poses a challenge for full-length NS5B purification. Consequently, structure−function investigations on NS5B have been achieved employing recombinant NS5B with 21−55 amino acid C-terminal truncations, which does not compromise its enzymatic activity.[6,8] Crystal structures of NS5B have revealed a classical "right-hand" topology of the polymerase family, with discernible fingers, palm, and thumb subdomains. An unusual feature of NS5B is the extensive interactions between its finger

Structure based drug design   Ligand based drug design
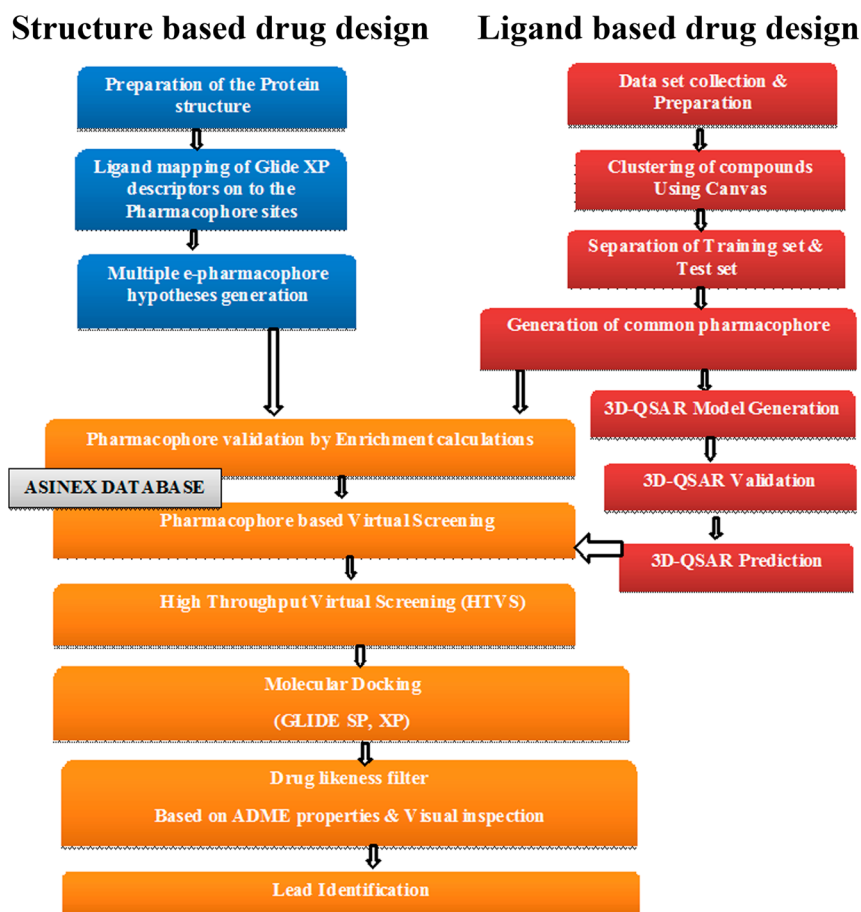
Figure 1. Virtual screening workflow. Systematic representation of the workflow for lead identification based on structure-based drug design (e-pharmacophore based) and ligand-based drug design (3D-QSAR based).

and thumb subdomains, resulting in an encircled active site.[9] Several crystal structures of NS5B either alone or in complex with diverse ligands ranging from ions and small molecules to nucleotides and non-nucleoside analogue inhibitors (NNI) have been elucidated.[9−15] The NNI are reported to bind to five distinct allosteric pockets on NS5B located on its thumb or palm domain.[16,17] These structures offer an excellent platform for the identification and development of new NS5B inhibitors. The structure-based strategy has proved to be successful over random screening of existing chemical libraries.[18−20] In addition, ligand-based and three-dimensional quantitative structure−activity relationship (3D-QSAR) approaches have also been explored for the development of novel and potent NS5B inhibitors.[21−24]

This Article summarizes efforts for the discovery of hit compounds via integrating multiple cocrystal structures of NS5B−inhibitor complexes and 3D-QSAR of ligands and evaluating the effectiveness of the pharmacophores toward identification of small molecule NS5B inhibitor leads. Toward this goal, we have constructed multiple energy-based pharmacophore models based on five good resolution crystal structures with high affinity bound inhibitors displaying in vitro HCV NS5B IC_{50} values between 10 and 290 nM. The method combines pharmacophore perception and database screening with protein−ligand energetic terms computed by the Glide XP (extra precision) scoring function to rank the importance of pharmacophore features in continuation, as we believe that integration of multiple crystal structure information is fruitful as
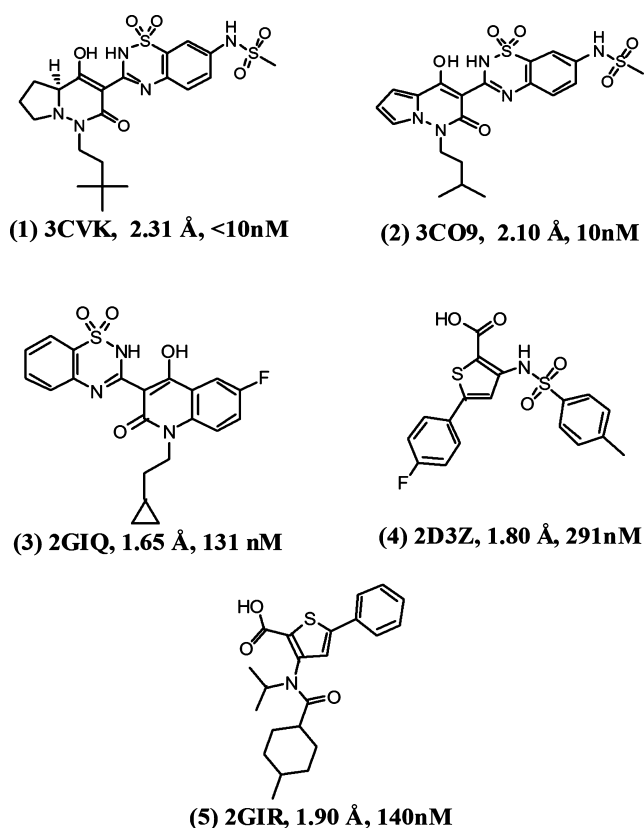
envisaged successfully in our earlier work with β-secretase.[25] The energy-based and ligand-based pharmacophore models were validated using enrichment analysis, and all six models were further employed for the high-throughput virtual screening and docking. The flow of activities performed is depicted in Figure 1, and the hits derived at each stage were analyzed for diversity based on each pharmacophore model. Combining these hits on the basis of the six pharmacophore models, followed by molecular docking and application of filtering criteria pertaining to their interaction with amino acids in the NS5B allosteric pocket and 3D-QSAR predictions, 10 compounds were finally short-listed for testing against HCV.

2. MATERIALS AND METHODS

2.1. Computational Details. All computations were carried out on an Intel Core 2 Duo E7400 2.80 GHz capacity processor with a memory of 2GB RAM running with the RHEL 5.2 operating system. GLIDE 5.8 module of Schrödinger suite was used to screen the public database like Asinex. PHASE 3.4 implemented in the Maestro 9.3 software package (Schrödinger, LLC) was used to generate pharmacophore and 3D-QSAR models.

2.2. Protein Structure Preparation. We selected five crystal structures of HCV NS5B polymerase bound with inhibitors from PDB on the basis of their good resolution and their activity ranging from ≤10 to 291 nm. For palm region, selected crystal structures were 3CVK, 3CO9, and 2GIQ, whereas 2D3Z and 2GIR were considered for thumb region.

Protein preparation wizard in Maestro software was used to prepare the proteins (Maestro, v9.2, Schrodinger, LLC, New York, NY). The crystal ligand structure with resolution and affinities are presented in Figure 2.

(1) 3CVK, 2.31 Å, <10nM

(2) 3CO9, 2.10 Å, 10nM

(3) 2GIQ, 1.65 Å, 131 nM

(4) 2D3Z, 1.80 Å, 291nM

(5) 2GIR, 1.90 Å, 140nM

Figure 2. Structures of bound inhibitors from indicated cocrystals selected for docking studies listed with their respective PDB IDs, IC_{50} values, and resolutions.

2.3. Energy-Optimized Structure-Based Pharmacophore Hypothesis Generation. Receptor Grid Generation tool in Maestro software package was used to generate energy grids for all prepared protein structures. We docked the crystal ligands using the None (refine only) option in the XP (extra precision) docking in Glide 4.2 X. Default settings were employed to minimize and optimize the structure. On the basis of the XP descriptor information, pharmacophore features were generated using PHASE 3.4. Pharmacophore sites were automatically generated with phase (Phase, v3.0, Schrodinger, LLC, New York, NY) using the default set of six chemical features: hydrogen-bond acceptor (A), hydrogen-bond donor (D), hydrophobe (H), negative ionizable (N), positive ionizable (P), and aromatic ring (R). Hydrogen-bond acceptor sites were represented as vectors along the hydrogen-bond axis in accordance with the hybridization of the acceptor atom. Hydrogen-bond donors were represented as projected points, located at the corresponding hydrogen-bond acceptor positions in the binding site. Projected points allow the possibility for structurally dissimilar active compounds to form hydrogen bonds to the same location, regardless of their point of origin and directionality. Each pharmacophore feature site is first assigned an energetic value equal to the sum of the Glide XP contributions of the atoms comprising the site, allowing sites to be quantified and ranked on the basis of the energetic terms.

Glide XP descriptors include terms for hydrophobic enclosure, hydrophobically packed correlated hydrogen bonds, electrostatic rewards, π–π stacking, π–cation, and other interactions. ChemScore, hydrogen bonding, and lipophilic atom pair interaction terms are included when the Glide XP terms for hydrogen binding and hydrophobic enclosure are zero.

2.4. Generation of Data Sets for Ligand-Based Pharmacophore Hypothesis Generation. We choose 1568 NS5B inhibitors from literature with known IC_{50} values. These inhibitors showed a wide range of activity (0.0004–500 μM) and structural diversity. The LigPrep 2.5 application from Schrödinger software package was utilized to build and energetically minimize structures and to add hydrogens and generate stereoisomers at neutral pH 7 using ionizer subprogram. The Canvas 1.4 chemoinformatics package (version 1.4, 2011, Schrödinger, LLC, New York, NY) was employed to cluster molecules based on tanimoto similarities between a set of linear fingerprint descriptors to determine the structural diversity among the compounds. Finally, 132 compounds were selected from 151 clusters as representative molecules. The structures and references are provided in the Supporting Information. Clustered molecules with structural diversity were utilized for 3D-QSAR development. Among these, a set of ninety inhibitors used as actives to validate the pharmacophores were also prepared using LigPrep 2.5. The IC_{50} values were converted to pIC_{50} to get the linear relationship in the QSAR equation, using the following formula:

$$pIC_{50} = -\log IC_{50} \tag{1}$$

where IC_{50} is the concentration of the antagonist producing 50% inhibition of NS5B enzyme. The data set consisted of some highly active and inactive molecules with few molecules as moderately active.

2.5. Molecular Alignment and Development of Pharmacophore Model for 3D-QSAR Approach. To develop the 3D-QSAR model, pharmacophore models and statistical analyses were performed using PHASE. Conformers were generated using a Macromodel torsion angle search approach followed by minimization of each generated structure using OPLS_2005 as force field with implicit distance-dependent dielectric solvent model. A maximum of 1000 conformers were generated per structure using the Macro-Model search method (ConfGen) in the preprocess minimization of 100 steps, including post process minimization of 50 steps. Each generated conformer was further filtered using a relative energy scale of 10.0 kcal/mol and a RMSD of 1 Å. After conformers generation step, there was an essential step of creating pharmacophore sites on each ligand structure responsible for facilitating noncovalent binding interaction between the ligand and the receptor. The threshold range of the active and inactive pIC_{50} was 9.398 and 3.301, respectively. Pharmacophore features were created using the clean minimized structure. The last step in pharmacophore generation was "scoring hypothesis" in which hypotheses were ranked to make rational choices among the hypotheses and the most appropriate one for further exploration. Common pharmacophores were examined by a scoring protocol to identify the pharmacophore from each surviving n-dimensional box that yielded the best alignment of the active set ligands. The inactive molecules were scored to observe the alignment of these molecules with respect to the pharmacophore hypothesis to enable selection of the hypothesis. The larger is the difference between the scores of actives and inactives, the better

is the hypothesis in distinguishing the actives from inactives.[26] The 132 ligands were aligned with the pharmacophore template of compound with high active score shown in the Supporting Information.

2.6. 3D-QSAR Modeling. Pharmacophore data set comprising of 132 molecules was utilized to develop our 3D-QSAR model. The data were divided randomly as 70 for training set and 62 for test set (Supporting Information), by using the method "Automated Random Selection" option present in the PHASE module. PHASE provided the means to build QSAR models using the activities of the ligands that match a given hypothesis. PHASE QSAR models were based on PLS regression, applied to a large set of binary valued variables, which were individually derived from a regular grid of cubic volume elements, with each cubic element represented by a set of bit values (0 or 1) to account for the different type of pharmacophore features in the training set.[27] The independent variables in the QSAR model were derived from a regular grid of cubic volume elements that span the space occupied by the training set ligands. The ligand sets used were diverse structures, and hence pharmacophore-based QSAR models were generated using the 72-member training set and a grid spacing of 1.00 Å. QSAR models containing one to five PLS factors were generated. A model with PLS factor five was considered as the best statistical model. This model was validated by predicting activities of test set molecules.[28]

2.7. Enrichment Calculations. Enrichment factor (EF) was employed for the fraction of known actives recovered when a fraction of database is screened. EF(X%) is the fraction of actives recovered after X% of decoy database has been screened.[29] A decoy set consists of 1090 molecules, which includes 90 known active molecules of NS5B inhibitors. For this, we focused primarily on EF(1%), the enrichment in the top 1% of the decoys.[30] A second enrichment metrics, the Boltzmann-enhanced discrimination of receiver operating characteristic (BEDROC),[31] was also used as a way to ensure that the results and conclusions were significant. We used $\alpha = 20.0$ and $\alpha = 160.9$ for the comparison. The value of $\alpha = 20.0$ was suggested as a reasonable choice for virtual screening evaluations.[29] We have also found a value of $\alpha = 20.0$ to be useful in virtual screening, which corresponds to 80% of the score being accounted in the top 1% of the database.

2.8. Database Preparation. For validating the reliability of the 5 e-pharmacophore and 3 ligand-based pharmacophore models, an initial data set of 1000 drug-like decoys, with an average molecular weight of 400 D (the "dl-400" data set), was employed. Ligand decoy sets were available for download (http://www.schrodinger.com/glide_decoy_set). The commercial chemical database Asinex with over 500 000 compounds was processed through redundancy checking and Lipinski filters to select compounds that have better drugability. Database molecules were prepared using LigPrep (LigPrep, version 2.5, Schrödinger, LLC, New York, NY, 201.) with Epik (Epik, version 2.2, Schrödinger, LLC, New York, NY, 201.). Conformational sampling was performed on all database molecules using the ConfGen search algorithm.[32] ConfGen samples the conformations based on a heuristic search algorithm and energetic evaluations to efficiently explore diversity around rotatable bonds, flexible ring systems, and nitrogen inversions. We employed ConfGen with the OPLS_2005 force field and a duplicate pose elimination criterion of 1.0 Å rmsd (root-mean-square deviation) to remove redundant conformers. A distance-dependent dielectric sol-

vation treatment was used to screen electrostatic interactions. The maximum relative energy difference of 10 kcal/mol was chosen to exclude high-energy structures. Using phase, the database was indexed with the automatic creation of pharmacophore sites for each conformer to allow rapid database alignments and screening.

2.9. Pharmacophore Filtering of Database. For the e-pharmacophores and ligand-based pharmacophore, explicit matching was required for the most energetically favorable site, conditional on it scoring better than −1.0 kcal/mol. Screening molecules required one to match a minimum of 3 sites and 4 sites for hypotheses with 5 or more sites. Distance matching tolerance was set to 2.0 Å as a balance between stringent and loose-fitting matching alignment. Database hits were ranked in order of their fitness score[26] a measure of how well the aligned ligand conformer matches the hypothesis based on rmsd, site matching, vector alignments, and volume terms. The fitness scoring function is an equally weighted composite of these three terms and ranges from 0 to 3, as implemented in the default database screening in phase (Phase, version 3.3, Schrödinger, LLC, New York, NY, 2011).

2.10. PLS Analysis and External Statistical Validation of QSAR Models. The 3D-QSAR models were generated by using significant statistical method of partial least-squares analysis. The cross validation analysis was performed using the "leave one out" (LOO) method, which evaluates the predictive ability of QSAR model in accordance with procedure and equations described previously.[30,32,33] 3D-QSAR models with good predictive ability and the corresponding pharmacophores were selected as the top three pharmacophore models and employed to find new leads.

2.11. High-Throughput Virtual Screening and Molecular Docking. Database ligands were docked into the binding sites of the crystal structures in the palm and thumb regions with Glide 5.7 (Glide, version 5.7, Schrödinger, LLC, New York, NY, 2011), utilizing the high-throughput virtual screening (HTVS) scoring function to estimate protein−ligand binding affinities. Glide HTVS is faster and more tolerant to suboptimal fits than is Glide XP, making it better for comparison in this work.[32] The center of the Glide grid was defined by the position of the cocrystallized ligand. Default settings were used for both the grid generation and the docking. Post docking minimization was implemented to optimize the ligand geometries. Compounds with best docking and Glide scores were then subjected to Glide SP (standard precision) and XP (extra precision) screening. Finally, the top hits were docked using the GOLD program for comparison (http://www.ccdc. cam.ac.uk). The generated grid information for each crystal ligand for docking is given in Table 1.

2.12. Biological Assays. All compounds were reconstituted in DMSO as 10 mM stocks and diluted just prior to evaluation. Huh7/Rep-Feo1b cells were grown at 37 °C with

Table 1. Grid Information along with Their PDB IDs Employed for Docking Studies

sr. no.	PDB code	X-center	Y-center	Z-center
1	3CVK	96.23	9.30	54.85
2	3CO9	33.54	44.51	53.95
3	2D3Z	9.33	33.76	74.92
4	2GIQ	9.18	−8.33	−12.80
5	2GIR	−33.86	−18.84	33.39

5% CO_2 in Dulbecco's Modified Eagle's Medium (DMEM) containing 10% FBS, 1% penicillin/streptomycin, 1% glutamine, and 500 μg/mL of G418. IC_{50} values were calculated using CalcuSyn software (Biosoft).

2.13. Anti-HCV Activity and Cytotoxicity in Cell-Based Assays. Anti-HCV activity of compounds in cell-based assays was evaluated in the Huh7/Rep-Feo1b cells replicating the subgenomic HCV genotype 1b replicon RNA carrying the firefly luciferase reporter.[34] The evaluation of compounds was performed similarly to that described earlier.[34,35] In brief, the cells were grown in 96-well plates and treated with the compound (50 μM) or equal amount of DMSO (control). The final concentration of DMSO was maintained at 0.5%. After 48 h of treatment, cells were washed with phosphate buffered saline, and the inhibitory effect of compounds on HCV replication was measured as the relative levels of the luciferase signals in compound treated cells versus DMSO controls employing luciferase assay system (Promega).

The effect of the compounds on cell viability was investigated in Huh7/Rep-Feo1b cells under conditions similar to the antiviral assays. Briefly, cells plated in 96-well plates were treated with compounds or DMSO (control) for 48 h. Cell viability was assessed employing CellTiter 96 AQ_{ueous} One solution in accordance with the manufacturer's protocol (Promega, U.S.).

2.14. HCV NS5B RdRp Inhibition Assay. Recombinant HCV NS5B carrying a deletion of 21 hydrophobic amino acids from C-terminus and bearing N-terminal hexa-histidine tag was purified employing Ni-NTA column chromatography as described earlier.[36,37] Inhibition of HCV NS5B was conducted on poly rA-U_{12} template primer with $[\alpha\text{-}^{32}P]UTP$ and $MnCl_2$ as the divalent cation in the presence of DMSO or the compound as described previously.[36] The reactions were incubated at 30 °C for 1 h, and reactions were terminated by adding ice cold 5% TCA solution. The nascent radiolabeled RNA was precipitated with 5% TCA, spotted on GF-B filters, and counted in a liquid scintillation counter. NS5B activity in the presence of DMSO was set at 100%, and that in the presence of the compounds was determined relative to this control. Compounds exhibiting ≥50% inhibition at 50 μM concentration were subjected to 50% inhibition (IC_{50}) evaluation. Graphpad prism 3.03 software was employed to extrapolate the IC_{50} values of compounds from dose−response curves at 8−10 concentrations of the serially diluted compounds in duplicate in two independent experiments.

3. RESULTS AND DISCUSSION

3.1. Protein Preparation. The HCV NS5B polymerase crystal structures with low resolution and with high active bound inhibitor were retrieved from PDB. Here, we used a total of five crystal structures with a resolution between 1.65 Å and 2.31 Å and activity of inhibitors (IC_{50}) ranging from ≤10 to 200 nM to enhance the chances of retrieving similarly active and diverse hits from the database screening. Three of the crystal structures (3CVK, 3CO9, and 2GIQ) represent bound inhibitors in the palm region, and the other two (2GIR and 2D3Z) represent inhibitors bound to the thumb region. We prepared these proteins using protein preparation wizard by adding hydrogens, and energy minimization by converging the heavy atoms to RMSD 0.30 Å using OPLS_2005 as force field, removed water molecules, and finally refined the protein structure shown in Figure 3. All of the crystal ligands were redocked onto the prepared protein to generate e-pharmaco-
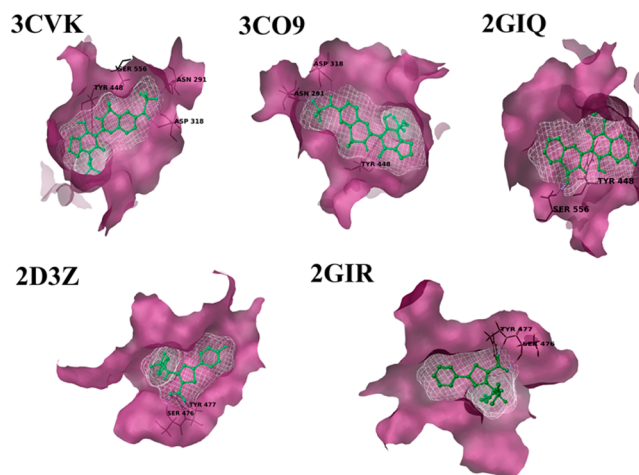
Figure 3. Binding site pocket of the five indicated crystal structures, demonstrated in white and magenta color created using the "create binding site surfaces" module of Schrödinger Suite.

phore. The root-mean-square deviation (rmsd) was less than 1 Å for all five of the crystal ligands. Ligand interaction analysis showed that the important residues in palm region were Asp 318, Asn 291, and Tyr 448 and for the thumb region Ser 376 and Tyr 377.[10,13]

3.2. Energy-Based Pharmacophore Generation. The e-pharmacophore method combines the aspects of structure-based and ligand-based techniques and was explored for five crystal structures of HCV NS5B polymerase used in this study. The pharmacophore hypotheses were developed by mapping Glide XP energetic terms onto pharmacophore sites, which were calculated on the basis of the structural and energy information between the protein and the ligand. The initial number of pharmacophore sites was set up to 10 for all of the crystal structures. Table 2 lists the total number of

Table 2. e-Pharmacophore Features[a]

sr. no.	PDB code	no. of possible sites	no. of selected sites	hypothesis
1	3CVK	3	3	ADR
2	3CO9	3	3	ADR
3	2D3Z	4	3	ARR
4	2GIQ	3	3	NRR
5	2GIR	3	3	NHR

[a]A, acceptor; D, donor; R, ring aromatic; N, negative ionic; and H, hydrophobic.

pharmacophore sites for each ligand prior to energy-based site selection and the optimized sites for hypothesis generation for the five crystal structures of NS5B (Figure 4). The pharmacophoric features for the palm region were ADR (3CVK and 3CO9) and ARR (2GIQ) and for the thumb region NRR (2D3Z) and NHR (2GIR). One ring (R) feature was common in all of the hypotheses. All of the crystal structures yielded three features of pharmacophore. The maximum and minimum distances between the features were 8.489 and 2.341 Å, respectively. The distance mapping among the features is presented in Table 3. Acceptor (A) and ring (R) features were common among the pharmacophoric sites derived from palm domain crystal structures (3CVK, 3CO9, and 2D3Z), where in thumb domain crystal structures (2GIQ and 2GIR) ring (R) and negatively ionizable (N) features were common. Regarding
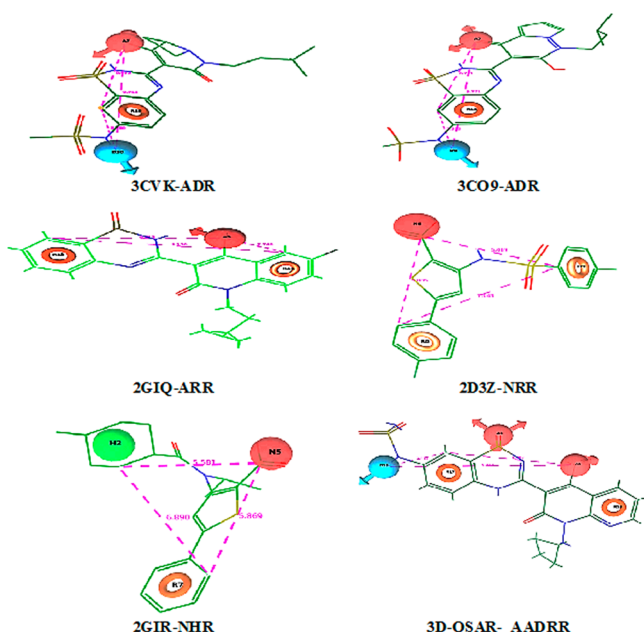
Figure 4. Bound inhibitors from crystal structure selected for docking studies, depicted with their respective PDB IDs (3A) 3CVK, (3B) 3CO9, (3C) 2GIQ, (3D) 2D3Z, (3E) 2GIR and their potential e-pharmacophore hypotheses, for screening the compound database with their respective PDB IDs (3F) 3CVK, (3G) 3CO9, (3H) 2GIQ, (3I) 2D3Z, (3J) 2GIR and the final selected best pharmacophore hypothesis on the basis of the 3D-QSAR approach. Pink sphere with arrow, hydrogen-bond acceptor (A); yellow open circle, aromatic ring (R); blue sphere with arrow, hydrogen-bond donor (D); green sphere, hydrphobic (H); pink sphere, negatively ionizable (N).

the distance between A and R when compared among the palm region structures, the ligands of 3CVK and 3CO9 were similar (3.73 Å), while 2GIQ generated two R features and the distances with the A feature were 5.31 and 2.34 Å, respectively. This depicts that the e-pharmacophore derived from 2GIQ is very much different from other e-pharmacophores of the palm region and hence has the possibility of yielding diverse structure libraries. The enrichment results for all five crystal ligands using the e-pharmacophore method were compared for the enrichment factor (EF1%), BEDROC ($\alpha = 160.9$ and $\alpha = 20.4$), based on the recovery rate of actives against the ranked decoy database that consisted of 1090 compounds in which 90 were known inhibitors. The results are as shown in Table 5 where 3CO9 and 2GIR pharmacophores showed the highest enrichment at 1% (EF1%) and BEDROC values. Also, the overall enrichment factor (EF) and goodness of fit (GH) were better for 2GIR and 2D3Z.

3.3. Ligand-Based Pharmacophore Modeling. On the other hand, we attempted to perform ligand-based pharmaco-

phore modeling and 3D-QSAR to compare and integrate with structure-based approach. PHASE QSAR models may be either atom based or pharmacophore based, the difference being whether all atoms were taken into account, or merely the pharmacophore sites that can be matched to the hypothesis. The training set molecules were sufficiently rigid and congeneric, and hence our 3D-QSAR approach involved the generation of a common pharmacophore hypothesis built on the principle of identification and alignment of pharmacophoric features of the chemical structures. We divided our data set into actives, inactives, and moderately actives. 721 hypotheses were produced. On the basis of the scoring functions (survival active, survival inactive, posthoc), we selected the top three pharmacophore hypotheses based on good survival activity, vector, volume, energy scores, best active alignment, and number of matches (Table 4). Hypothesis 1 (AADRR) showed

Table 4. Ligand-Based Pharmacophore Hypotheses with Their Scores

hypothesis	survival score	survival-inactive score	vector score	volume score	score site
AADRR	4.03	2.94	1.000	0.84	0.98
AAADP	3.96	3.11	1.000	0.81	0.94
AADPR	3.99	3.07	1.000	0.80	0.98

2 hydrogen-bond acceptors, 1 hydrogen-bond donor, and two aromatic rings, hypothesis 2 (AAADP) showed 3 hydrogen-bond acceptors, 1 hydrogen-bond donor, and 1 positively ionizable, and hypothesis 3 (AADPR) showed 2 hydrogen-bond acceptors, 1 hydrogen-bond donor, 1 positively ionizable, and 1 aromatic ring as features (Supporting Information). Among these hypotheses, model 1 (AADRR) showed the highest survival score (Figure 4). To further validate these pharmacophores for virtual screening, we evaluated the enrichment factor similar to that performed for e-pharmacophores. The decoy set consisted of 1082 compounds in which 82 compounds were known inhibitors, which were not utilized in the pharmacophore hypotheses building. The results as shown in Table 5 rated hypothesis 1 (AADRR) as the best one as indicated by the highest EF1%, BEDROC, and GH values. 3D-QSAR models were then developed for the pharmacophore hypothesis using the training set structures that match the pharmacophore on three sites. However, we utilized the three models for the 3D-QSAR studies by generation of pharmacophore-based 3D-QSAR models and PLS analysis.

3.4. 3D-QSAR Models Generation and PLS Analysis. To develop superlative 3D-QSAR models, which were meant to exhibit reliable predictions, it necessitated internal and external statistical validation. Models that are capable of fulfilling statistical validation parameter boundaries can display more reliable predictions. Randomly, we chose 70 compounds in the

Table 3. Distance between the Features of e-Pharmacophores and 3D-QSAR-Based Pharmacophores[a]

sr. no.	PDB code	distance from A to R (Å)	distance from R to D (Å)	distance from D to A (Å)	distance from N to R (Å)	distance from H to R (Å)	distance from R to R (Å)	distance from R to A (Å)	distance from N to H (Å)	distance from R to N (Å)
1	3CVK	3.75	4.75	8.49[max]						
2	3CO9	3.73	4.73	7.92						
3	2GIQ	5.31					6.76	2.34		
4	2D3Z				5.03		6.68			5.31
5	2GIR					5.45			4.18	5.32

[a]Distance shown in angstroms; A, acceptor; D, donor; R, ring aromatic; N, negatively ionizable; H, hydrophobic; and P, positively ionizable.

Table 5. Validation of e-Pharmacophore and Ligand-Based Hypotheses

PDB code	e-pharmacophore features	EF 1%[a]	RIE[b]	ROC[c]	BEDROC[d] (α-160.9)	BEDROC (α-20)	EF[e]	GH[f]
3CVK	ADR	2.2	1.01	2.33	0.16	0.27	0.50	0.05
3CO9	ADR	7.7	1.02	1.55	0.24	0.67	0.51	0.05
2GIQ	ARR	6.6	0.96	2.46	0.25	0.51	0.76	0.06
2D3Z	NRR	3.3	1.04	1.97	0.20	0.19	1.57	0.13
2GIR	NHR	7.7	1.02	4.4	0.45	0.59	2.66	0.24
	AADRR	11	2.63	2.33	0.74	0.26	1.12	0.05
ligand-based	AAADP	8.4	1.81	1.55	0.65	0.18	0.77	0.04
	AADPR	9.6	2.70	1.97	0.68	0.26	0.88	0.02

[a]EF: Enrichment factor at 1% of the decoy data set. [b]RIE: Robust initial enhancement. [c]ROC: Receiver operating characteristic curve value. [d]BEDROC: Boltzmann-enhanced discrimination of receiver operating characteristic. [e]EF: Overall enrichment factor. [f]GH: Goodness of fit.

training set and 62 compounds in the test to develop the 3D-QSAR model. Important parameters obtained based on the LOO method (Table 6) favored the internal statistical

Table 6. PHASE 3D-QSAR and PLS Statistic for Internal Validation of the Data Set

statistical parameters	AADRR	AAADP	AADPR
number of molecules in training set	70	70	70
number of molecules in test set	62	62	62
R^{2b}	0.97	0.98	0.98
Q^{2f}	0.58	0.34	0.32
SD[a]	0.29	0.25	0.20
F-value[c]	381.8	451.3	781.3
Pearson-R[g]	0.77	0.58	0.62
RMSE[e]	0.81	1.051	1.02
r^2_{pred}[d]	0.59	0.51	0.12

[a]SD: Standard deviation of the regression. [b]R^2: For the regression. [c]F: Variance ratio. [d]r^2_{pred}: Predictive correlation coefficient value. [e]RMSE: Root-mean-square error. [f]Q: Squared (Q^2) value for the predicted activities. [g]Pearson R: Correlation between the predicted and observed activities for the test set.

validation by PLS analysis. Among the three models, hypothesis 1 (AADRR) showed good external predictive ability for each combination as compared to others. Hypothesis 1 showed a good R^2 value for the training set of 0.9666, good predictive power with Q^2 of 0.5810 for the test sets, with SD of 0.2936, and F value of 381.8. Further, the integrity of the model was predicted by r^2_{pred} for the test set with the value of 0.5955 (Table 7). The accepted LOO-cross validated value of training set (R^2) should be greater than 0.591, LOO cross validated value for test set (Q^2) should show a value greater than 0.55 to attain good predictive capacity, and standard deviation (SD) below 0.3, with minimum root-mean-square error (RMSE), and high value of variance ratio (F) to provide conventional QSAR validation limits. Also, the predictive correlation coefficient (r^2_{pred}) value generated on the basis of molecules of the test set demonstrated real predictive capacity and robustness of the QSAR model.

For a QSAR model, internal validation of LOO cross validated Q^2 is commonly used to assess predictive ability, where a high value of Q^2 is necessary and important but Q^2 alone is not a sufficient condition for a model to have a high predictive power. A reliable model should also be characterized by a high correlation coefficient R (or r^2) between the predicted and observed activities of compounds from an external test set.[38] In the present study, the best predictive ability of the model was characterized by correlation coefficient $R = 0.7679$

Table 7. External Statistical Validation of Quantitative Structure−Activity Relationship (QSAR) for Hypothesis 1 (AADRR) and Common Pharmacophore Hypotheses

external validation	parameter calculated	limitations
r^2_{cv}[a]	0.59	$r^2_{cv} > 0.5$
R[b]	0.77	must be close to 1
r^2	0.59	$r^2 > 0.5$
k value[c]	1.02	$0.85 \leq k \leq 1.15$
k' value	0.97	$0.85 \leq k' \leq 1.15$
R^2_0[d]	0.99	R^2_0 or R'^2_0 close to r^2
R'^2_0	0.97	R^2_0 or R'^2_0 close to r^2
$[(r^2 - R^2_0)/r^2]$[e]	−0.68	$[(r^2 - R^2_0/r^2] < 0.1$
$[(r^2 - R'^2_0/r^2]$	−0.65	$[(r^2 - R'^2_0/r^2] < 0.1$
r^2_m(LOO)[f]	0.825	r^2_m(LOO) > 0.5

[a]r^2_{cv}: Cross validated coefficient. [b]R (or r^2): Correlation coefficient between the actual and predicted activities. [c]k and k': Slope values of regression lines. [d]R^2_0 and R'^2_0: Correlation coefficients for the regression lines through the origin. [e]$[(r^2 - R^2_0)/r^2]$ and $[(r^2 - R'^2_0)/r^2]$: To calculate the relation between r^2, R^2_0, and R'^2_0. [f]R^2_m(LOO): Modified squared correlation coefficient for the "Leave One Out" method.

$(r^2 = 0.5895)$. High slope of regression lines through the origin k value of 1.018 and k' value of 0.9651 (either k or k' should be close to 1)[25] gave the substantial values of R^2_0 value 0.9883 and the R'^2_0 value 0.9722, which were obtained by calculating the correlation coefficient of regression lines of the scatter plot obtained by means of actual activity versus predicted activity and predicted activity versus actual activity plots, respectively (Figure 5). The relation between r^2, R^2_0, and R'^2_0 gave $(r^2 - R^2_0/r^2)$ values of −0.6765, and second relation $(r^2 - R'^2_0/r^2)$ value −0.6492 showed optimum values within the statistical limits (Table 7). Yet, our established QSAR model from hypothesis 1 (finalized after PLS analysis) gave an r^2_{cv} value of 0.592. A parameter of modified r^2 $[r^2_m(LOO)]$[30] was considered as a better external predictive potential for the whole set of compounds, which was of 0.8242 (>0.5) defined through scatter plot best fit line values (Figure 5). This appeared to be truly predictive by fulfilling the requirements of every parameter in the external validation (Table 7). We therefore consider this model as statistically significant. Thus, we resumed further steps to predict the activities of new leads from the compound libraries by using hypothesis 1.

3.5. Contour Map Analyses. The final validated hypothesis 1 obtained from 3D-QSAR was used to generate contour maps. These contour maps were important to identify the positions of the substitutions or replacements of atoms to enhance bioactivity. Inhibitory activity can be gained by visualizing and understanding the maps against most active (1) and least active (131) compounds. This could help in
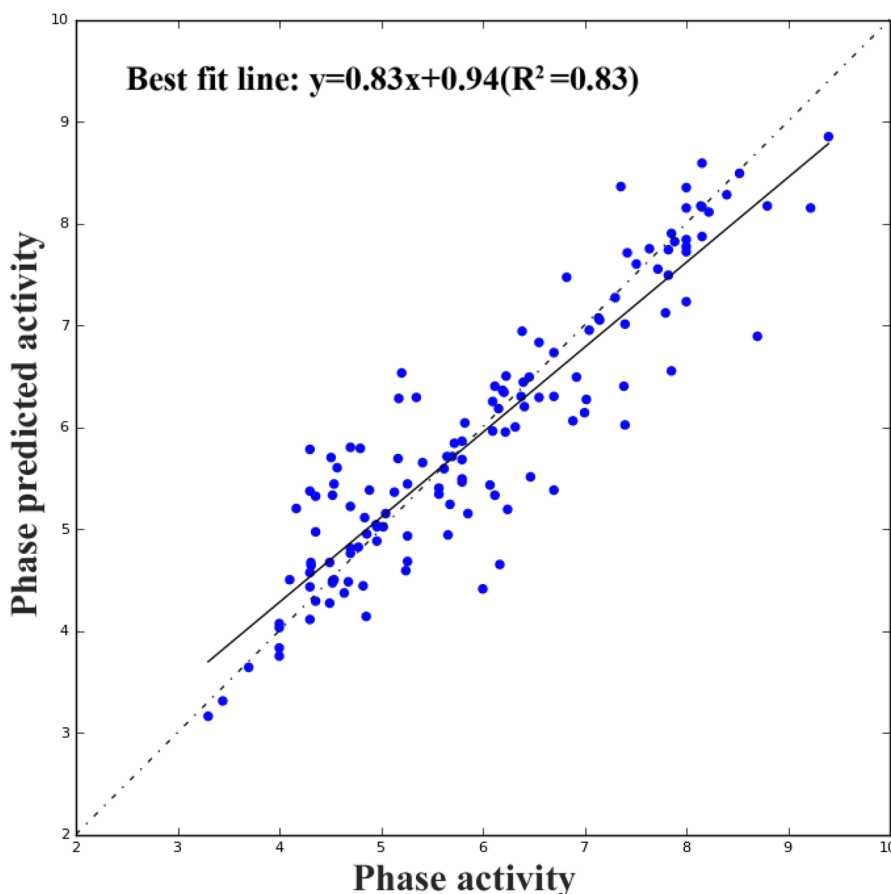
Figure 5. Scatter plot of the observed versus predicted activity of NS5B inhibitors generated by the best model obtained employing 70 compounds as the training set and validated using 62 compounds as the test set. Inclination of the best fit line from the origin was considered in the context of calculating values of external r^2, R_0^2, $R_0'^2$ and particularly to select the best predictive models from among comparable models by using the overall $r_m^2(\text{LOO})$ value (0.8208).

discovering novel scaffolds with good biological activity. The most and least active ligand contour maps were generated and are shown in Figure 6. Contour maps indicated H-bond donor effect on the most active ligand (1) and least active ligand (131) (Figure 6A and B), the hydrophobic effect of the ligands (Figure 6C and D), and the electron-withdrawing nature (Figure 6E and F) of both ligands represented in the figure are discussed further.

The hydrogen-bond donor nature for the most active compound 1 and the least active compound 131 when compared showed their most favorable region blue color and unfavorable regions red color. Hydrogen-bond donor mapping revealed that favorable regions lied near the nitrogens of naphthyridin and in benzothiadiazine, indicating their importance for activity as compared to the least active compound 131. Therefore, the presence of cyclic rings with hydrogen donor group in the scaffold backbone is very much needed for the activity.

Figure 6C and D when compared for their hydrophobic nature for the most active compound 1 and least active compound 131 revealed that the favored green region around the naphthyridin, benzothiadiazine, and cyclobutyl rings indicated that the terminal hydrophobic rings were very much needed for the activity of the compound, and the unfavorable yellow region on ring hydrogen moiety revealed that an increase in the carbon chain could increase the activity. In Figure 6E, the favored red regions were observed near

hydrogen-bond acceptors along with respective acceptor hypothesis features of most active compound, which indicated that these features were crucial for the activity and these groups should be unsubstituted with further lead modifications. In the least active compound as in Figure 6F, the unfavorable blue region blue surrounded the napthyridin ring moiety, which indicated that a decrease in the ring size could increase the biological activity of the compound.

3.6. Multiple Pharmacophore Models-Based Virtual Screening and Docking. Virtual screening studies of the commercial database are fruitful resource for initial lead identification. A fit value is a measure of how well the ligand fits the pharmacophore. Therefore, the hits with a high fit value are probably very active. In one method, we employed the best e-pharmacophores, and in a second method, the ideas based on the 3D-QSAR and its corresponding pharmacophore were utilized to find important features for the inhibition of NS5B, to help in designing the lead molecules. The best validated five e-pharmacophore model and the pharmacophore from 3D-QSAR (hypothesis 1) were used to screen against the databases of 500 000 compounds (Asinex) as presented in the flowchart (Figure 1). The number of hits from pharmacophore screening is presented in Table 8. From the phase find matches module output for the e-pharmacophore ADR for 3CVK, fitness of more than 1.4 was taken as limit for the HTVS (high throughput virtual screening), which yielded 13 226 compounds belonging to more than 1800 clusters. These 13 226
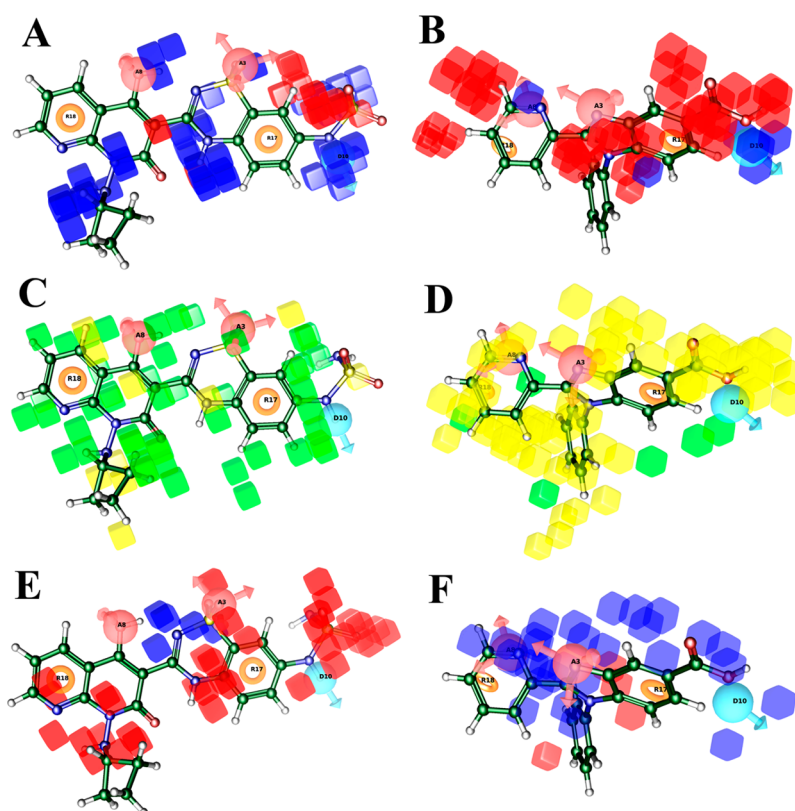
Figure 6. (A) H-bond donor effect, most active; (B) least active (blue, favorable; red, unfavorable); (C) hydrophobic effect, most active; (D) least active (green, favorable; yellow, unfavorable); (E) electron-withdrawing effect, most active; and (F) least active (red, favorable; blue, unfavorable).

Table 8. Number of Compounds (Hits) Retrieved at Each Stage of Screening

PDB code	phase find matches hits	clusters	HTVS hits	clusters	SP hits	clusters	XP hits	clusters	nonpeptidic	peptidic
3CVK	13 226	1812	7257	1026	1549	168	58	18	11	47
3CO9	7576	541	2092	356	1513	306	391	98	59	332
2D3Z	28 687	1923	2035	364	961	183	482	126	32	450
2GIQ	13 367	1646	9987	1345	2110	430	78	8	24	54
2GIR	11 256	978	6334	478	3843	334	205	28	61	144
3D-QSAR (2GIQ)	50 000	3458	4231	352	1552	269	108	19	22	86

Table 9. Lead Compounds with Their Fitness, Docking Score, Number of H-Bonds, Interaction with Important Amino Acids, and Two Respective Gold Scores

no.	fitness	docking score	H-bond	predicted activity	ligand interactions	gold score1	gold score2
H-1	1.11	−9.96	5	5.08	Tyr 477, Ser 476 (2), Arg 501, Lys 533	68	63
H-2	1.71	−8.67	4	6.18	Tyr 415, Ser 556, Asp 318 (2)	75	70
H-3	1.57	−8.22	2	5.20	Tyr 448, Ser 407	49	45
H-4	1.58	−8.13	5	5.92	Asn 291, Asp 318, Gln 446, Ser 556, Tyr 448	63	62
H-5	2.06	−6.56	2	5.08	Asp 318 (2)	61	62
H-6	2.02	−7.47	4	6.22	Asp 318, Tyr 448, Ser 556, Met 414	72	71
H-7	1.48	−7.82	4	6.05	Ser 476, Trp 528, Lys 533, Arg 422	58	59
H-8	1.80	−7.75	3	5.01	Asn 291, Tyr 448, Ser 556	67	61
H-9	1.41	−7.12	2	5.69	Tyr 448, Gln 446	42	43
H-10	1.86	−6.52	2	5.92	Asp 318, Ser 288	67	74

ligands were docked to palm region of 3CVK using default grid, and finally we selected 7257 ligand molecules as hits from HTVS with the criteria that the docking score was above −6 kcal/mol, fitness was above 1.6, and the number of H-bonds was above 2. This yielded 1026 clusters. These hits were further docked using the Glide SP (standard precision) docking module, and 1549 ligand molecules were selected on the basis

of the docking score −7 kcal/mol, fitness above 1.7, 2 H-bonds, and visual occupancy of ligand into the pocket. Finally, we subjected the Glide SP filtered ligands to Glide XP (extra precision) docking simulation. The top 58 ligand molecules with more than a docking score of −7.00 kcal/mol were visually inspected for the pose and for important binding residues. Top 58 ligands belonged to 18 diverse structural scaffolds, matching
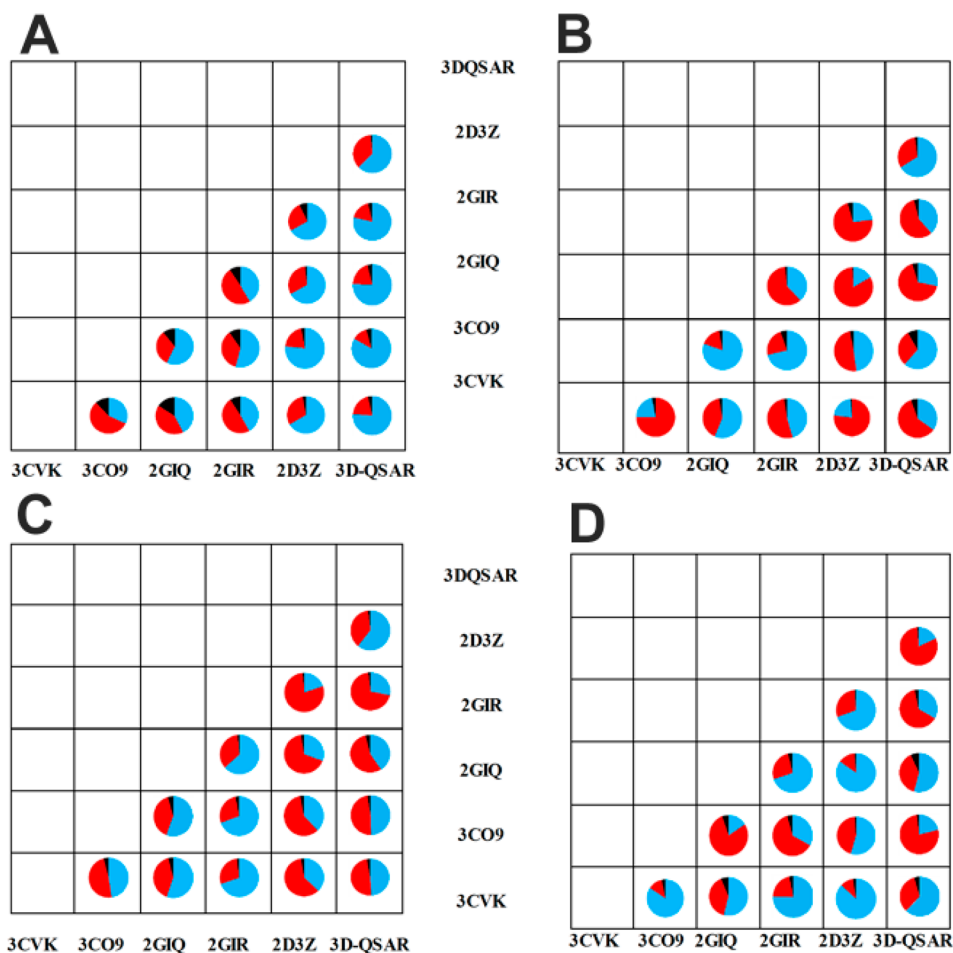
Figure 7. Pie graph shows the extent of overlapping of compounds at each stage of screening for all six pharmacophore models. (A) Phase find match hits, (B) HTVS docking hits, (C) SP docking hits, and (D) XP docking hits. Red and blue regions represent the number of screened hit compounds from any two pharmacophore models, and the black region represents the overall overlap of the screened compounds.

all three e-pharmacophoric features. The selected best pharmacophore based on the 3D-QSAR approach was ADDRR. Using this pharmacophore, phase find matches yielded the highest number of ligands. We chose 2GIQ PDB because of its very low resolution value. We docked our top 50 000 compounds in the palm active site pocket. Fitness above 1.4 (50 000 ligands of 3458 clusters) was selected for HTVS docking. For SP docking, we short-listed 4231 (352 clusters); they had a docking score above −5 kcal/mol, fitness above 1.5, and 2 H-bonds. For XP docking, we short-listed 1552 (269 clusters); they had a docking score above −6 kcal/mol, fitness above 1.7, and 2 H-bonds. The top 108 ligands clustered, which gave 19 diverged structural scaffolds with docking score above −6.5 kcal/mol, fitness above 2.0, and 2 H were selected. In contrast, the pharmacophore model from 3CO9 was found very restrictive and retrieved very limited hits from Asinex with only 7576 hits demonstrating that different pharmacophore models may have quite different performance in screening a chemical database. Therefore, multiple pharmacophore models can be used to improve the overall screening efficacy. Similar evaluations were performed for other pharmacophoric hypotheses of 3CO9, 2GIQ, 2D3Z, and 2GIR. All of the final hits were analyzed for their ranking, and the final hit lists were prepared on the basis of their ranking in more than one pharmacophore models. From these, finally we selected 10 hits whose predicted

activities, Glide scores, fitness, H-bond data, and GOLD scores are presented in Table 9.

Structural diversity is an important index for the quality of the hits by an in silico approach. Accordingly, we compared the structural diversity of the hits retrieved from the Asinex database. The compounds retrieved from each of the five crystal structures and 3D-QSAR model after docking results were clustered using Canvas (Canvas, version 1.4, Schrödinger, LLC, New York, NY, 2011), a clustering algorithm from Schrodinger. The total number of clusters from each structure as represented in Table 8 was analyzed for structural diversity among the compounds generated. For further analyzing the performance of different pharmacophore models, the Glide XP screening, which is based on the protein active site pocket grids, was employed, and the extent of overlap of the hits retrieved from each pair of different pharmacophore models was studied and presented as a pie graph in Figure 7. When compared to the similar pharmacophore models of 3CVK, 3CO9, and 2GIQ (Figure 7A), based on the similar features of A and R, there were more common hits as compared to other pharmacophore models, which were diverse. Figure 7A−D depicts similar analyses at every stage of the design cycle from pharmacophore filtering, HTVS, Glide SP, and XP steps, and there was little overlap in later stages, indicating the possibility of more diverse ligand generation. Analyses of the hits based on peptidic versus nonpetidic leads revealed that most of the leads generated were

Table 10. Lead Compounds with Their Qikprop Properties

no.	mol MW[a]	QPlogP o/w[b]	QPPCaco[c]	QPlogHERG[d]	percent human oral absorption[e]	rule of five[f]
H-1	421.45	3.52	16.56	−4.26	69.35	0
H-2	420.48	3.43	139.17	−5.64	85.38	0
H-3	357.41	4.43	777.37	−6.35	100.00	0
H-4	400.43	−1.54	52.24	−4.87	35.75	1
H-5	291.31	2.01	130.09	−5.73	76.55	0
H-6	418.43	0.59	2.97	−6.00	38.83	0
H-7	478.53	3.03	27.36	−6.95	70.38	0
H-8	382.33	1.88	205.60	−4.95	79.38	0
H-9	381.34	2.50	112.27	−5.97	78.27	0
H-10	357.38	1.78	91.85	−4.88	72.51	0

[a]Molecular weight (acceptable range <500). [b]Predicted octanol/water partition coefficient log p (acceptable range from −2.0 to 6.5). [c]Predicted Caco-2 cell permeability in nm/s (acceptable range: <25 is poor and >500 is great). [d]Predicted value for blockage of HERG K+ channels (concern below −6.5). [e]Percentage of human oral absorption (<25% is poor and >80% is high). [f]Rule of five (no. of violations of Lipinski's rule of five: 0 is good and 4 is bad).
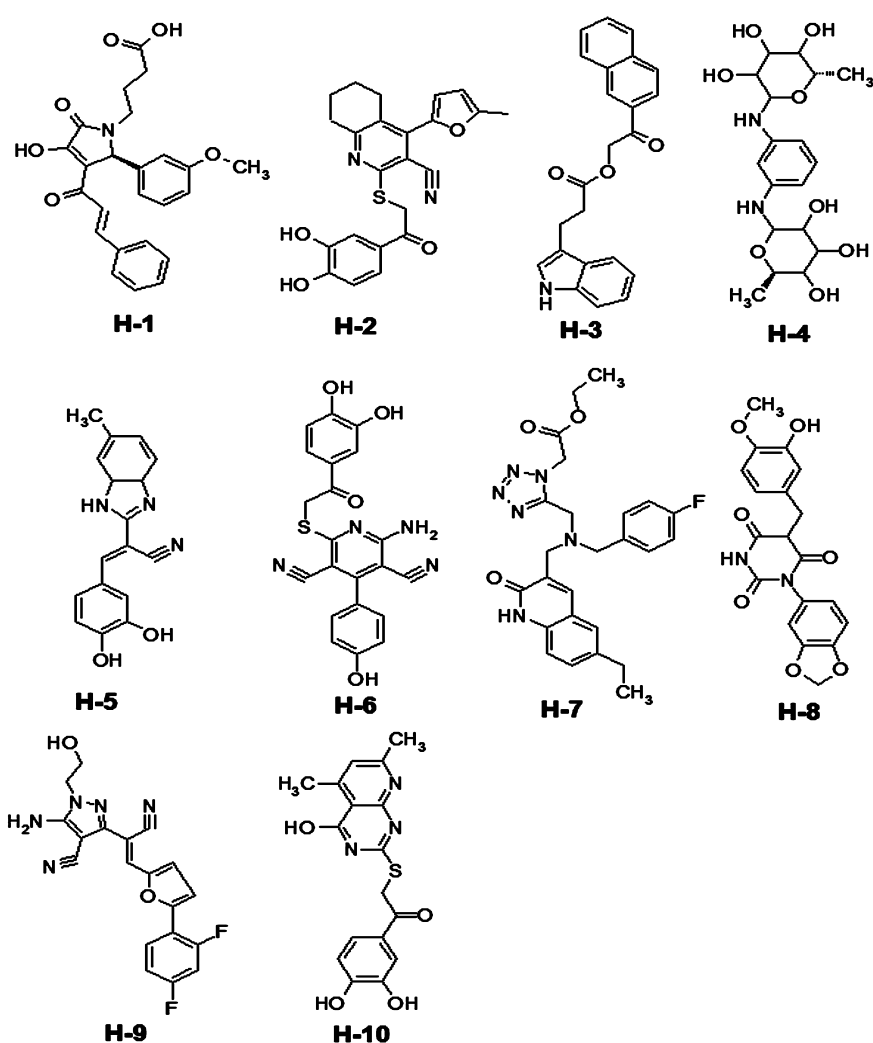

Figure 8. Structures of the top 10 hit compounds.

peptidic ligands (Table 8). Short-listing was performed by comparing among nonpeptidic leads ranked in more than one pharmacophore model.

3.7. ADME Predictions. We evaluated the 10 lead compounds for the pharmaceutically relevant properties to check drug likeness and predictions for drug's pharmacokinetics including ADME. The QikProp module of the Schrodinger suite was used for evaluation of drug-like behavior through analysis of pharmacokinetic parameters required for absorption, distribution, metabolism, and excretion (ADME) (Table 10). All 10 lead compounds showed good partition coefficient (QPlogPo/w) values, which were critical for understanding of absorption and distribution of drugs, to range from −1.535 to 4.448. Factor QPPCaco indicating permeability of the 10 lead

compounds ranged from 2.966 to 777.368, where QPPCaco was a predicted apparent Caco-2 cell permeability in nm/s value, a key factor for the estimation of cell permeability in biological membranes. All of the lead compounds passed the entire pharmacokinetic requirement for a drug-like compound and were within the acceptable range defined for human use. Additional parameters such as molecular weight, H-bond donors, H-bond acceptors, and human oral absorption according to Lipinski's rule of five, etc., were also evaluated for their drug-like behavior. Thus, compounds with binding interaction and good predicted pharmacokinetic properties were finalized and are shown in Figure 8. The short-listed hits belonged to diverse scaffolds like pyridine, pyrrole, benzimidazole, quinolone, pyrimidinotrione, pyrimidopyridine, etc. For hits H1−10 when analyzed for their binding interactions in the allosteric pockets of the palm and thumb regions, it was found that most of the hits (H-2−6, H-8−10) were retrieved from palm region docking, while hits H-1 and H-7 were retrieved from thumb region binding. Among these hits, H-2, H-6, and H-7 showed higher predicted bioactivity based on the 3D-QSAR model. The binding pose of all of these hit compounds is presented in the Supporting Information. It is clear that important amino acid residues reported for palm or thumb region have been involved in the interaction. The hit compounds (H-1−10) were procured from the Asinex database for further biological screening.

3.8. Anti-HCV Activity and Cellular Cytotoxicity. Having identified the hit molecules from the in silico screening approach, we hypothesized that the compounds may possess the ability to bind with HCV NS5B and consequently inactivate its RdRp activity. To test this hypothesis, the compounds were initially screened at 50 μM concentrations in standard in vitro RdRp assays, and the relative inhibition of NS5B RdRp activity was evaluated (Table 11). Among the tested compounds, H-5 and H-6 displayed ≥50% inhibition of HCV NS5B RdRp activity at 50 μM compound concentration. Therefore, compounds H-5 and H-6 were evaluated in a dose-dependent manner to estimate their potency against NS5B polymerase, and their IC$_{50}$ values were found to be 28.8 and 47.3 μM, respectively. Among the other compounds, H-10 showed weak inhibition of NS5B RdRp activity displaying 27% inhibition at 50 μM. All other compounds exhibited negligible inhibition of NS5B RdRp activity.

We further evaluated the anti-HCV activity and cellular cytotoxicity of the hit compounds, in cell-based HCV reporter assays employing the Huh7/Rep-Feo1b cells bearing the HCV subgenomic replicons.[34,35] In this system, compounds H-5 and H-6 displayed no cytotoxicity, but strongly inhibited HCV RNA replication by 97% and 86%, respectively, at 50 μM concentration. Compounds H-2, H-7, H-9, and H-10 also displayed <80% inhibition of HCV RNA replication at 50 μM concentration but with decreased cell viability of 64%, 68%, 77%, and 113%, respectively. Thus, the inhibition observed by these compounds was partly due to cellular toxicity. Furthermore, because these compounds did not inhibit NS5B, their ability to inhibit HCV replication in cell-based assays could be attributed to targeting of other nonstructural proteins of HCV or cellular factors facilitating HCV replication. Compound H-3 showed approximately 50% inhibition of HCV replication at 50 μM concentration, whereas compounds H-4 and H-8 exhibited weak inhibition but displayed no cellular cytotoxicity at the concentration tested.

Table 11. Anti-HCV and Anti-NS5B Activity and Cytotoxicity of Hit Compounds

no.	Huh7/Rep-Feo1b inhibition[a] (%)	viability[b] (%)	NS5B RdRp inhibition[c] (%)	IC$_{50}$[d] (μM)
H-1	NI	87.9 ± 0.9	5.9 ± 2.2	ND
H-2	96.7 ± 1.3	64.8 ± 2.1	9.1 ± 1.0	ND
H-3	51.1 ± 5.1	88.4 ± 0.9	1.9 ± 1.0	ND
H-4	16.9 ± 4.6	109.8 ± 3.4	NI	ND
H-5	97.0 ± 1.1	114.7 ± 11.1	67.8 ± 4.3	28.8 ± 2.0
H-6	85.8 ± 1.8	89.6 ± 4.5	50.9 ± 3.4	47.3 ± 5.0
H-7	91.2 ± 2.9	68.4 ± 9.3	8.1 ± 3.2	ND
H-8	25.3 ± 8.1	106.2 ± 5.1	NI	ND
H-9	79.7 ± 4.5	77.0 ± 6.2	13.4 ± 5.1	ND
H-10	80.9 ± 1.0	113.5 ± 12.4	27.3 ± 1.3	ND

[a]The anti-HCV activity of the compounds in Huh7/Rep-Feo1b reporter cells at 50 μM concentration. Cells incubated in the presence of equal amounts of DMSO served as control. The data represent an average of three independent experiments. [b]Cell viability was determined at 50 μM concentrations using the CellTiter 96 AQueous One Solution Cell Proliferation kit (Promega) in Huh7/Rep-Feo1b cells. Cells treated with equal amounts of DMSO were considered as 100% viable and served as control. [c]Inhibition of NS5B RdRp activity at 50 μM concentrations of indicated compounds on homopolymeric template-primer (polyrA-U$_{12}$) and MnCl$_2$ as divalent metal ion. The RdRp activity in the presence of DMSO was set at 100%, and that in the presence of the compounds was determined relative to this control. The results represent an average of three independent experiments in triplicate. [d]Ten quarter log dilutions of compounds were employed in RdRp assay. NI, no inhibition. ND, not determined.

4. CONCLUSION

The available information on HCV NS5B crystal structures bound with inhibitors in the palm and thumb allosteric pockets in combination with ligands with known inhibitory potential was explored for a comparative pharmacophore analysis followed by a high-throughput virtual screening and docking to identify diverse nonpeptidic leads. The hits were retrieved by combining multiple pharmacophore hypothesis and 3D-QSAR predictions, and the resulting 10 hits displayed diverse scaffold, which were then validated employing biochemical and cell-based NS5B and anti-HCV inhibition assays. Of these, H-5 and H-6 were the most promising, exhibiting IC$_{50}$ values of 28.8 and 47.3 μM against NS5B polymerase and anti-HCV inhibition of 97% and 86% at 50 μM, respectively. The lead compounds identified bore benzimidazole (H-5) and pyridine (H-6) scaffolds, which are both amenable for further optimization and development as NS5B inhibitors.

■ ASSOCIATED CONTENT

⑤ Supporting Information

Compounds utilized for ligand-based pharmacophore and 3D-QSAR, alignment of all ligands based on the pharmacophore model AADRR for 3D-QSAR, ligand-based pharmacophore models, alignment of hits retrieved from each pharmacophore, binding pose of H-1−H-10 in the active site pocket, 3D-QSAR data set of 132 compounds with their fitness to the best pharmacophore selected, activity, predicted activity, and the difference between their predicted and actual activity, distance between the features for ligand-based pharmacophores, and comparison of distance between the central core A−D, D−R, and R−A. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*Phone: +91-40-66303515. Fax: +91-40-66303998. E-mail: pyogee@hyderabad.bits-pilani.ac.in.

*Phone: (973) 972-8653. Fax: (973) 972-5594. E-mail: kaushik@njms.rutgers.edu.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) WHO. Hepatitis C. Fact sheet no. 164, 2011; accessed at www.who.int/mediacenter/factsheet/fs164/en/.

(2) Manns, M. P.; Foster, G. R.; Rockstroh, J. K.; Zeuzem, S.; Zoulim, F.; Houghton, M. The way forward in HCV treatment–finding the right path. *Nat. Rev. Drug Discovery* **2007**, *6*, 991−1000.

(3) Sarrazin, C.; Zeuzem, S. Resistance to direct antiviral agents in patients with hepatitis C virus infection. *Gastroenterology* **2010**, *138*, 447−62.

(4) Behrens, S. E.; Tomei, L.; De Francesco, R. Identification and properties of the RNA-dependent RNA polymerase of hepatitis C virus. *EMBO J.* **1996**, *15*, 12−22.

(5) Moradpour, D.; Brass, V.; Bieck, E.; Friebe, P.; Gosert, R.; Blum, H. E.; Bartenschlager, R.; Penin, F.; Lohmann, V. Membrane association of the RNA-dependent RNA polymerase is essential for hepatitis C virus RNA replication. *J. Virol.* **2004**, *78*, 13278−84.

(6) Rice, C. Perspective: miles to go before we sleep. *Nature* **2011**, *474*, S8.

(7) Hagedorn, C. H.; van Beers, E. H.; De Staercke, C. Hepatitis C virus RNA-dependent RNA polymerase (NS5B polymerase). *Curr. Top. Microbiol. Immunol.* **2000**, *242*, 225−60.

(8) Tomei, L.; Vitale, R. L.; Incitti, I.; Serafini, S.; Altamura, S.; Vitelli, A.; De Francesco, R. Biochemical characterization of a hepatitis C virus RNA-dependent RNA polymerase mutant lacking the C-terminal hydrophobic sequence. *J. Gen. Virol.* **2000**, *81*, 759−67.

(9) 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*, 937−43.

(10) Ruebsam, F.; Sun, Z.; Ayida, B. K.; Webber, S. E.; Zhou, Y.; Zhao, Q.; Kissinger, C. R.; Showalter, R. E.; 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. Hexahydro-pyrrolo- and hexahydro-1H-pyrido[1,2-b]pyridazin-2-ones as potent inhibitors of HCV NS5B polymerase. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5002−5.

(11) 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*, 3616−21.

(12) Maynard, A.; Crosby, R. M.; Ellis, B.; Hamatake, R.; Hong, Z.; Johns, B. A.; Kahler, K. M.; Koble, C.; Leivers, A. L.; Leivers, M. R.; Mathis, A.; Peat, A. J.; Pouliot, J. J.; Roberts, C. D.; Samano, V.; Schmidt, R. M.; Smith, G. K.; Spaltenstein, A.; Stewart, E. L.; Thommes, P.; Turner, E. M.; Voitenleitner, C.; Walker, J. T.; Waitt, G.; Weatherhead, J.; Weaver, K. L.; Williams, S.; Wright, L.; Xiong, Z. Z.; Haigh, D.; Shotwell, J. B. Discovery of a potent boronic acid derived inhibitor of the HCV RNA-dependent RNA polymerase. *J. Med. Chem.* **2013**, DOI: 10.1021/jm400317w.

(13) Biswal, B. K.; Wang, M.; Cherney, M. M.; Chan, L.; Yannopoulos, C. G.; Bilimoria, D.; Bedard, J.; James, M. N. Non-nucleoside inhibitors binding to hepatitis C virus NS5B polymerase reveal a novel mechanism of inhibition. *J. Mol. Biol.* **2006**, *361*, 33−45.

(14) Bressanelli, S.; Tomei, L.; Rey, F. A.; De Francesco, R. Structural analysis of the hepatitis C virus RNA polymerase in complex with ribonucleotides. *J. Virol.* **2002**, *76*, 3482−92.

(15) Bressanelli, S.; Tomei, L.; Roussel, A.; Incitti, I.; Vitale, R. L.; Mathieu, M.; De Francesco, R.; Rey, F. A. Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13034−9.

(16) Patil, V. M.; Gupta, S. P.; Samanta, S.; Masand, N. Current perspective of HCV NS5B inhibitors: a review. *Curr. Med. Chem.* **2011**, *18*, 5564−97.

(17) Yan, S.; Appleby, T.; Larson, G.; Wu, J. Z.; Hamatake, R. K.; Hong, Z.; Yao, N. Thiazolone-acylsulfonamides as novel HCV NS5B polymerase allosteric inhibitors: convergence of structure-based drug design and X-ray crystallographic study. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1991−5.

(18) Schoenfeld, R. C.; Bourdet, D. L.; Brameld, K. A.; Chin, E.; de Vicente, J.; Fung, A.; Harris, S. F.; Lee, E. K.; Le Pogam, S.; Leveque, V.; Li, J.; Lui, A. S.; Najera, I.; Rajyaguru, S.; Sangi, M.; Steiner, S.; Talamas, F. X.; Taygerly, J. P.; Zhao, J. Discovery of a novel series of potent non-nucleoside inhibitors of Hepatitis C virus NS5B. *J. Med. Chem.* **2013**, *56*, 8163−82.

(19) Hucke, O.; Coulombe, R.; Bonneau, P.; Bertrand-Laperle, M.; Brochu, C.; Gillard, J.; Joly, M. A.; Landry, S.; Lepage, O.; Llinas-Brunet, M.; Pesant, M.; Poirier, M.; Poirier, M.; McKercher, G.; Marquis, M.; Kukolj, G.; Beaulieu, P. L.; Stammers, T. A. Molecular dynamics simulations and structure-based rational design lead to allosteric HCV NS5B polymerase thumb pocket 2 inhibitor with picomolar cellular replicon potency. *J. Med. Chem.* **2013**, DOI: 10.1021/jm4004522.

(20) Beaulieu, P. L.; Coulombe, R.; Duan, J.; Fazal, G.; Godbout, C.; Hucke, O.; Jakalian, A.; Joly, M. A.; Lepage, O.; Llinas-Brunet, M.; Naud, J.; Poirier, M.; Rioux, N.; Thavonekham, B.; Kukolj, G.; Stammers, T. A. Structure-based design of novel HCV NS5B thumb pocket 2 allosteric inhibitors with submicromolar gt1 replicon potency: discovery of a quinazolinone chemotype. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 4132−40.

(21) Yu, H.; Fang, Y.; Lu, X.; Liu, Y.; Zhang, H. Combined 3D-QSAR, molecular docking, molecular dynamics simulation, and binding free energy calculation studies on the 5-hydroxy-2H-pyridazin-3-one derivatives as HCV NS5B polymerase inhibitors. *Chem. Biol. Drug. Des.* **2013**, DOI: 10.1111/cbdd.12203.

(22) Zhang, H. X.; Li, Y.; Wang, X.; Xiao, Z. T.; Wang, Y. H. Insight into the structural requirements of benzothiadiazine scaffold-based derivatives as hepatitis C virus NS5B polymerase inhibitors using 3D-QSAR, molecular docking and molecular dynamics. *Curr. Med. Chem.* **2011**, *18*, 4019−28.

(23) Wang, X.; Yang, W.; Xu, X.; Zhang, H.; Li, Y.; Wang, Y. Studies of benzothiadiazine derivatives as hepatitis C virus NS5B polymerase inhibitors using 3D-QSAR, molecular docking and molecular dynamics. *Curr. Med. Chem.* **2010**, *17*, 2788−803.

(24) Musmuca, I.; Caroli, A.; Mai, A.; Kaushik-Basu, N.; Arora, P.; Ragno, R. Combining 3-D quantitative structure-activity relationship with ligand based and structure based alignment procedures for in silico screening of new hepatitis C virus NS5B polymerase inhibitors. *J. Chem. Inf. Model.* **2010**, *50*, 662−76.

(25) Palakurti, R.; Sriram, D.; Yogeeswari, P.; Vadrevu, R. Multiple e-pharmacophore modeling combined with high-throughput virtual screening and docking to identify potential inhibitors of β-secretase (BACE1). *Mol. Inf.* **2013**, *32*, 385−398.

(26) Dixon, S. L.; Smondyrev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. A. PHASE: a new engine for pharmacophore perception, 3D-QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 647−71.

(27) Lewis, R. A. A general method for exploiting QSAR models in lead optimization. *J. Med. Chem.* **2005**, *48*, 1638−48.

(28) Consonni, V.; Ballabio, D.; Todeschini, R. Comments on the definition of the Q2 parameter for QSAR validation. *J .Chem. Inf. Model.* **2009**, *49*, 1669−78.

(29) 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*, 1750−9.

(30) Loving, K.; Salam, N. K.; Sherman, W. Energetic analysis of fragment docking and application to structure-based pharmacophore hypothesis generation. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 541−54.

(31) Truchon, J. F.; Bayly, C. I. Evaluating virtual screening methods: good and bad metrics for the "early recognition" problem. *J. Chem. Inf. Model.* **2007**, *47*, 488−508.

(32) Pratim Roy, P.; Paul, S.; Mitra, I.; Roy, K. On two novel parameters for validation of predictive QSAR models. *Molecules* **2009**, *14*, 1660−701.

(33) Battu, M. B.; Chandra, A. M.; Sriram, D.; Yogeeswari, P. Pharmacophore-based 3D-QSAR and molecular docking studies to identify new non-peptidic inhibitors of cathepsin S. *Curr. Med. Chem.* **2013**, *Sep*, 24059223.

(34) Kim, K.; Kim, K. H.; Kim, H. Y.; Cho, H. K.; Sakamoto, N.; Cheong, J. Curcumin inhibits hepatitis C virus replication via suppressing the Akt-SREBP-1 pathway. *FEBS Lett.* **2010**, *584*, 707−712.

(35) Nichols, D. B.; Fournet, G.; Gurukumar, K. R.; Basu, A.; Lee, J. C.; Sakamoto, N.; Kozielski, F.; Musmuca, I.; Joseph, B.; Ragno, R.; Kaushik-Basu, N. Inhibition of hepatitis C virus NS5B polymerase by S-trityl-L-cysteine derivatives. *Eur. J. Med. Chem.* **2012**, *49*, 191−9.

(36) Kaushik-Basu, N.; Bopda-Waffo, A.; Talele, T. T.; Basu, A.; Costa, P. R.; da Silva, A. J.; Sarafianos, S. G.; Noel, F. Identification and characterization of coumestans as novel HCV NS5B polymerase inhibitors. *Nucleic Acids Res.* **2008**, *36*, 1482−96.

(37) Chen, Y.; Bopda-Waffo, A.; Basu, A.; Krishnan, R.; Silberstein, E.; Taylor, D. R.; Talele, T. T.; Arora, P.; Kaushik-Basu, N. Characterization of aurintricarboxylic acid as a potent hepatitis C virus replicase inhibitor. *Antiviral Chem. Chemother.* **2009**, *20*, 19−36.

(38) Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P. Random forest: a classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1947−58.