

Published in final edited form as:

Bioorg Med Chem. 2013 June 1; 21(11): 3127–3137. doi:10.1016/j.bmc.2013.03.032.

Inhibitors for the hepatitis C virus RNA polymerase explored by SAR with advanced machine learning methods

Iwona E. Weidlich^{a,b,c,1}, Igor V. Filippov^d, Jodian Brown^a, Neerja Kaushik-Basu^e, Ramalingam Krishnan^e, Marc C. Nicklaus^b, and Ian F. Thorpe^{a,1}

^aDepartment of Chemistry and Biochemistry, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250

^bChemical Biology Laboratory, Center for Cancer Research, National Cancer Institute, NIH, DHHS, Frederick National Laboratory for Cancer Research, 376 Boyles Street, Frederick, MD 21702

^cComputational Drug Design Systems (CODDES) LLC, Rockville, MD

^dChemical Biology Laboratory, Center for Cancer Research, SAIC-Frederick, Inc., 376 Boyles Street, Frederick, MD 21702

^eDepartment of Biochemistry and Molecular Biology, UMDNJ New Jersey Medical School, 185 South Orange Ave, Newark, NJ 07103

Abstract

Hepatitis C virus (HCV) is a global health challenge, affecting approximately 200 million people worldwide. In this study we developed SAR models with advanced machine learning classifiers Random Forest and k Nearest Neighbor Simulated Annealing for 679 small molecules with measured inhibition activity for NS5B genotype 1b. The activity was expressed as a binary value (active/inactive), where actives were considered molecules with IC₅₀ < 0.95 μM. We applied our SAR models to various drug-like databases and identified novel chemical scaffolds for NS5B inhibitors. Subsequent in vitro antiviral assays suggested a new activity for an existing prodrug, Candesartan cilexetil, which is currently used to treat hypertension and heart failure but has not been previously tested for anti-HCV activity. We also identified NS5B inhibitors with two novel non-nucleoside chemical motifs.

Keywords

NS5B; RdRp; HCV; SAR; HCV-796; RDKit; Random Forest; k Nearest Neighbor Simulated Annealing; Candesartan cilexetil; Computational Drug Repositioning

© 2013 Elsevier Ltd. All rights reserved.

¹Corresponding author. Tel.: +1 240 409 2467, Fax: +1 240 629 8023, iweidlic@coddes.com; Tel.: +1 410 455 5728, Fax +1 410 455 2608, ithorpe@ummc.edu.

Note:

The authors declare no competing financial interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

The World Health Organization estimates that nearly three percent of the world's population is infected with hepatitis C virus (HCV).¹ The Centers for Disease Control and Prevention reports that approximately 5 million people are infected within the United States and that there are up to 230,000 new HCV infections in the U.S. every year which result in 8,000 to 10,000 deaths annually. Chronic HCV infections can cause liver cirrhosis and hepatocellular carcinoma.² There is no vaccine and the current standard of care (ribavirin and interferon alpha) can result in severe side-effects.³ Due to the large number of people infected, HCV is an attractive target for the development of antiviral drugs.

The HCV genotype 1 subtypes "a" and "b" in particular have been widely studied, as the current standard of care is less effective against this strain compared to the other strains.⁴ Patients infected with HCV genotype 1 fail to achieve a lasting virological response under current treatment in about 50% of cases.⁵ Thus, new HCV therapeutics are sorely needed. The HCV NS5B gene product is an RNA-dependent RNA polymerase that is essential for viral replication and has no human homolog, making it a prime target for drug development. NS5B has a "right hand" structure common to several viral polymerases, with three major sub-domains known as the finger, palm, and thumb (Fig. 1).

In this work we employ machine learning techniques and structure-activity relationship (SAR) analysis to identify novel inhibitors of NS5B of HCV genotype 1b (hereafter referred to simply as NS5B). We apply these ligand-based drug design approaches in an effort to reposition known drugs as potential HCV therapeutics and to identify new chemical scaffolds for inhibitors.

2. Design of SAR Models

SAR techniques are widely used in the study of anti-viral therapeutics. Unfortunately many such efforts have failed in the past due to a variety of factors: small training data sets^{6,7} lack of input data curation, overfitting of the data and lack of external validation. Our approach reported here aims at avoiding all such pitfalls.

In this work we selected large datasets and curated them to construct high quality training sets. We also made sure to employ machine learning methods that are more stable vis-à-vis training sets that include imperfect data, such as cases where bioactivity of some compounds may be unreliable. No statistical method can assure the reliability of SAR prediction for structures significantly different from those seen in the training set.⁸ Therefore, in addition to statistical modeling methods which included cross-validation and an external test set, we applied our predictions and tested prospective novel compounds using *in vitro* NS5B activity assays.

2.1. Datasets

PubChem, a component of the NIH Roadmap (now called Common Core) Molecular Libraries Initiative^{9,10,11} provides information on a large number of biological activities of small molecules. The assay results obtained directly through the NIH Roadmap project (not through other projects and entities such as ChEMBL) are currently available for more than 500,000 chemically diverse screening compounds, obtained from both commercial and academic sources that are distributed by the NIH Molecular Libraries Small Molecule Repository to the Molecular Libraries Probe Production Centers Network screening centers.

Starting in about 2010, the Computer-Aided Drug Design (CADD) Group of the National Cancer Institute (NCI) has periodically made available an aggregated set of structures and

assay results from these “NIH Roadmap assays”, consolidated into one SD file. The current version of this freely available file comprises more than 470,000 unique structures.¹² It contains NIH Roadmap biological data indexed by PubChem Assay IDs (AID) and is available from the NCI CADD Group’s public web server.¹³

From this set we extracted 45 assays specifically for HCV NS5B genotype 1b. Only IC₅₀ results identified as “actual” and converted to micromolar range were used. We removed all the records marked with “>” sign (IC₅₀ not actual). Additionally for all cases where there was more than one value we checked for consistency. Some assays were expressed in different units such as EC₅₀, DC₅₀ and they were excluded from our data set. In the end 29 assays with IC₅₀ values in micromoles with the type “=” (IC₅₀ actual) were used to construct the models.

Selected assays fell into the deposited category called “literature extracted”. We reviewed the literature to identify the experimental conditions of these assays.^{14–28}

We selected 23 biochemical binding assays and 2 functional assays. All of them measured NS5B polymerase activity and some of the experimental conditions were different. The scientific literature indicated various laboratories.

Most compounds were assayed for their inhibitory activity against recombinant HCV NS5B Δ 21 using a biochemical assay in which the template label mixture (MgCl₂, DTT, ATP, UTP, CTP, 3′ UTR RNA template, μ Ci ³²P CTP, water) was added to the assay plates followed by each compound dilution. In some cases buffer conditions were different where Tris HCl was used. The NS5B enzyme used in the assays was derived from genotype 1b.

All IC₅₀ values were extracted and connected to 418 structures with CID’s (Compound ID’s) and put into one sd file.

An additional set of 261 compounds²⁹ and corresponding assay data was manually extracted from the literature. All assays were secondary biochemical binding assays with or without detergent. Assays were against HCV NS5B Δ 21 from genotype 1b. Some used Mg²⁺ as a cofactor while others used Mn²⁺. A number of buffer conditions were used.^{30–39}

We feel that even with some variability in testing approaches mentioned above, the data, after the described filtering and standardization, can be useful for the development of computational models.

We used the Optical Structure Recognition Application to facilitate information extraction.⁴⁰ Our training set was therefore constructed from the combined data for 679 small molecules to predict anti-viral activity against NS5B. All assay data were carefully analyzed and curated manually. We also selected 323 molecules from ChEMBL⁴¹ previously tested for NS5B inhibition as an external validation set. There was no overlap by structure between the training and validation sets.

2.2. Training Procedures

To create a predictive SAR model, we tested open source and commercial products for machine learning: PASS⁴², Phase⁴³ and Eureqa⁴⁴. Finally we decided to use an in-house developed python script based on the RDKit⁴⁵ Random Forest⁴⁶ machine learning and chemistry modules (ML and Chem).

RDKit is an open source toolkit for cheminformatics and machine learning written in C++ and Python.⁴⁷ While some of the other software tools have shown higher precision for activity prediction, RDKit was our top choice for not overfitting the data (a common

problem in machine learning approaches) based on our tests of overfitting tendencies by the Y-randomization procedure.

Random forest is one of the most accurate modern machine learning methods. It is an ensemble classifier that is a collection of simpler individual classifiers. The individual classifiers in the case of random forest are decision trees. A decision tree can be thought of as an algorithmic implementation of the 20 question game, each node of the tree is a question such as “is feature N smaller than the threshold T?”, and the branches are “Yes/No” answers. The leaves in this case will be the prediction results such as “Active/Inactive”. The prediction of the whole forest can be an average of the individual tree predictions.⁴⁶

We set a threshold of 0.95 μ M below which a compound was considered active. The activity was expressed as a binary value (Y/N) for the purposes of machine modeling. 285 out of 679 molecules were therefore assigned an “active” label, while the remaining structures were designated as “inactive”. 210 in the external validation set were found to be active according to the above threshold.

A mean Pairwise Tanimoto Similarity coefficient was calculated for the whole set, as well as for the active and inactive subsets separately, using Accelrys Pipeline Pilot extended connectivity fingerprints.⁴⁸ In all three cases, the similarity was in the low 20% range.

We used 2048-bit Morgan fingerprints of radius 2 as features and employed a Random Forest model consisting of 5000 trees. Morgan Fingerprints is a bit vector where bits are set according to molecular neighborhoods of atoms.⁴⁹ It is similar to Extended Connectivity Fingerprints used in Accelrys Pipeline Pilot and MNA descriptors used in PASS.

Concurrently we investigated a novel classification algorithm k Nearest Neighbor Simulated Annealing (kNN-sa).⁵⁰ kNN-sa is an ensemble classifier, with individual classifiers being the k-nearest neighbors method.⁵¹ Each individual classifier finds k (usually 3 or 5) neighbors which are the closest (according to some metric) to the point to be predicted. The prediction then is a weighted average of the values found on the neighbors.

Area Under the Curve (AUC) for the Receiver Operating Characteristic (ROC) curve was used to measure performance of the model for four cases – self-prediction, cross-validation, Y-randomization and prediction for the external validation set. (Table 1 and Figures 2–3). Cross-validation was performed by randomly splitting the training set into training (70%) and validation (30%) sets. The procedure was repeated 5 times and the average AUC is reported in Table 1.

2.3. External Validation - ChEMBL

3. Results and Discussion

SAR models are usually expected to be reliable only within a certain area of chemistry, i.e. a range of compounds somewhat similar to the structures from the training set: the applicability domain. However when the goal is to find novel chemical motifs, it is often necessary to venture outside of the applicability domain while being aware that the accuracy of the model might degrade to some extent. We employed this approach to find new chemotypes.^{52,53}

3.1. Filtering Drug-like Databases

Re-investigating existing drugs for new indications reduces the cost of drug discovery and has become increasingly popular in academia and the pharmaceutical industry. In this study

we perform re-investigation, but present results only for compounds that are commercially available.

The computed SAR models were applied to several databases of known drugs and biologically active compounds: DrugBank, Traditional Chinese Medicine, MDL Drug Data Report, MDL Patent Chemistry database, CHMIS-A Comprehensive Herbal Medicine Information System for Cancer, ComGenex - now AMRI Hungary, Beilstein Natural Products and structures extracted from US patents by the IBM Strategic IP Insight platform (SIIP) and donated to NIH.⁵⁴ The results are shown in Table 2.

All compounds predicted as active by both SAR models were cross-docked into 14 models of the NS5B polymerase based on high resolution X-ray crystal structures of the enzyme.

We were not able to purchase selected compounds from vendors identified by ChemSpider.⁵⁵ We used CACTVS,⁵⁶ Perl and OpenBabel⁵⁷ to extract information from ZINC,⁵⁸ ChemNavigator-Sigma-Aldrich, Inc.⁵⁹ and eMolecules⁶⁰ using InChIkey. The resulting number of commercially available compounds among the predicted molecules is shown in Table 2.

3.2. Cross-Docking

Computational screening of NS5B inhibitors with a machine learning paradigm requires additional computational and experimental analysis to succeed. Docking represents one such additional computational approach. However, binding affinity of NS5B inhibitors may be affected by enzyme conformational changes, thus making the predicted docking poses less reliable. The crystal structures of the NS5B apoprotein show that the finger and thumb domains interact to encircle the active site. Allosteric inhibitors that bind to the thumb or finger domain can displace the finger loop and disrupt this structure.⁶¹ Therefore rigid crystal structures may not account for all the relevant enzyme conformations. Exceptions to this general observation include inhibitors that bind to the palm domain, which are not observed to induce enzyme conformational changes.

We identified twenty-nine crystal structures of HCV NS5B in the Protein Data Bank.⁶² Visual inspection of the downloaded crystal structures revealed that the structures could be classified into seven groups based on their different binding sites in the enzyme as well as on enzyme conformation. 14 PDB IDs were selected for cross-docking in the seven groups as follows: **1]** 1NHU⁶³ (Thumb) **2]** 2BRK, 2WHO⁶⁴ (Thumb) **3]** 2D3Z⁶⁵, 2D41⁶⁵, 2GIR [Thumb] **4]** 3HKY⁶⁶ (Palm) **5]** 2HAI⁶⁷ (Thumb, Finger), 2WRM (Thumb 2), 3FRZ⁶⁸ (Thumb) **6]** 2GIQ (Thumb 2), 3H2L⁶⁹ (Palm), 3H5S⁷⁰ (Palm) **7]** 3E51⁷¹, 3H5U⁶⁹ (Palm).

All compounds predicted as active by both SAR models were filtered based on the physico-chemical properties of known actives ($267 < MW < 663$, $0.4 < \text{clogP} < 8$, $42 < PSA < 237$ Å², $10 < \text{NRB} < 10$, $10 < \text{HBA} < 10$, $5 < \text{HBD} < 5$, $1 < \text{NAR} < 5$).⁷²

The SAR generated set was docked into thumb, palm and finger binding sites of NS5B using the extra-precision mode in the program Glide (Schrödinger).⁷³ In the cross-docking procedure, each flexible ligand was docked into a NS5B crystal structure that is bound to a different ligand. The generated poses were filtered using a cutoff filter for H-bond energy of -1.3 kcal/mol using the Glide “sort” utility from Schrödinger.

We superimposed the generated docking poses with the conformation of the small molecule as found in 14 selected crystal structures using the option “absolute position” in the program MOE.⁷⁴ This procedure allowed us to select only poses with similar binding modes from a few models and use these for further analysis.

The final selection of compounds was based on the assumption that SAR analysis and docking used separately do not always successfully predict potential drug candidates by themselves. Thus, we used both methods together to increase the chance of identifying successful hits. The number of compounds ultimately selected for purchase was nevertheless significantly lower due to limited commercial availability of the identified hits. (Table 2)

3.3. Prediction and experimental testing of novel chemical structures

3.3.1. Identification of Candesartan Cilexetil through SAR and Computational Drug Repositioning—Eleven compounds were selected for initial in vitro anti-viral testing. The compounds were initially selected based on the procedure described in section 3.2 and earlier, less accurate version of the Random Forest SAR model combined with commercial availability.

To make sure that selected compounds were not registered drugs against HCV we checked the interactions between small molecules and proteins at STITCH⁷⁵, FDA⁷⁶ and Pipeline Citeline⁷⁷. We also searched for side effects at SIDER⁷⁸ and Drugcite.⁷⁹

We identified three compounds predicted by the RF model with activity scores above the threshold for active compounds and simultaneously predicted four compounds by the kNN-sa model (which included the same three predicted by RF).

The compound with the highest predicted score was experimentally confirmed to be active, with $IC_{50}=14.2\text{ }\mu\text{M}$ and $IC_{90}=28\text{ }\mu\text{M}$ where IC_{50} indicates a 50% inhibitory concentration and IC_{90} indicates 90% inhibitory concentration (see Biological Evaluation section below).

The identified active compound is known as Atacand (Candesartan cilexetil), (*InChIKey:GHOSNRCGJFBJIB-UHFFFAOYSA-N*, source: MDL database), a drug which is known as an angiotensin II receptor inhibitor used to treat high blood pressure, congestive heart failure and other conditions. It also has been reported that Candesartan decreases liver fibrosis^{80,81}

Liver fibrosis is formation of scar tissue in response to liver damage. This scar tissue secretes stellate cells in an activated state. The stellate cells carry a receptor for the angiotensin hormone when activated. If this receptor is blocked, the cell becomes inactive and does not participate in scar tissue formation.⁸² Thus, some hypothesize that using a drug such as Candesartan cilexetil, which blocks angiotensin receptors, should result in less scar tissue formation in the livers of patients with hepatitis C. However Candesartan cilexetil has not yet been specifically tested for direct action against HCV NS5B genotype 1 b.

Cyclohexyl 1-hydroxyethyl carbonate (cilexetil) ester, known as Candesartan cilexetil is mostly used in combination therapies. Pipeline Citeline reports several therapeutic classes for the drug: Antihypertensive, renin system, Cardiovascular, Symptomatic antidiabetic, and Ophthalmological. The ClinicalTrials.gov database⁸³ reported clinical studies in the following areas: in hypertensive HCV positive patients to prevent hypertension, for antihypertensive treatment in HIV and HCV co-infected patients, and as anti-fibrotic intervention in patients with chronic Hepatitis C.

3.3.1.1. Predicted Key Binding Interactions: Using the absolute position approach we identified one pose of candesartan cilexetil which had similar binding mode to the small molecule (Figure 7, Compound 2) found in the crystal structure with PDB code 3E51. The conformational analysis of docking poses of Candesartan cilexetil indicated that the palm site is the best fit for Candesartan. This pose is shown in Figure 6 Left.

We noticed that Candesartan cilexetil (Figure 6 Right) forms similar hydrogen bonds with NS5B as the compound, called thereafter “compound **2**” bound to the Palm domain of NS5B in the Co-crystal X-ray structure with PDB ID 3E51 (Figure 7). These similarities in binding interactions play an important role in the identification of the best pose of Candesartan cilexetil.

The 5-hydroxy group on the pyridazinone ring of compound **2** forms hydrogen bonds with the backbone amide NH of TYR 448 and with a structural water molecule that interacts with the backbone NH of GLY 449 (Figure 7). One of the oxygen atoms of the carbonate ester group of Candesartan cilexetil (**1**) forms a hydrogen bond directly with TYR 448 without mediation by a water molecule (Figure 6).

Simultaneously the X-ray crystal structure of compound **2** bound to the Palm domain of NS5B shows a hydrogen bond to a second structural water molecule (hydrogen-bonded to the side chain of SER 288) (Figure 7). However we did not predict similar binding for candesartan cilexetil (**1**) but instead a hydrogen bond to ASP 318 which occurs in the X-ray structure of NS5B complex with compound **2** as well (Figure 6).

Such hydrogen bonding interactions are not the only means of stabilizing the NS5B enzyme structure. Additionally the tetrazole ring present in Candesartan cilexetil (**1**) forms hydrogen bond with TYR 415 which also belongs to the deep hydrophobic pocket (*shown schematically as a dark blue line*).

In contrast, the pyrrolidine ring present in compound **2** binds to a much more shallow cleft near the surface of the enzyme that is made up of residues GLY 410, MET 414, and GLN 446 (*shown schematically as a light blue line*).⁷¹ (Figure 7)

Making the assumption that active compounds have specific contacts with NS5B to display activity, and that predicted docking poses are reliable we carefully analyzed two compounds containing the same biphenyl-tetrazole substituent. Our results revealed the binding motif for active compound **1** which was different than for compound **3** (Figure 8). This noticeable difference may be associated with large activity changes.

This analysis helps us to determine whether a new molecule will perform better or worse compared to the specific molecules in the dataset. The binding motif for compound **3** is different than for compound **1** and **3** was inactive as both our models predicted. It forms hydrogen bonds with ASP 220 and THR 221 not observed in the X-ray crystal structure of compound **2** bound to the Palm domain of NS5B. It has been observed that many potent polymerase pyridazinone inhibitors^{84,85} contain an isoamyl fragment bound in a deep hydrophobic pocket comprised of NS5B residues PRO 197, ARG 200, LEU 384, CYS 366, MET 414, TYR 415, and TYR 448 (Figure 7).

Candesartan (**1**) contains two phenyl rings in its biphenyl-tetrazole substituent where one of the rings inserts into this deep hydrophobic pocket. Two polar carbonyl groups present in inactive compound **3** make this hydrophobic interaction impossible (*circled by a dashed line in Fig. 8 Left*).

3.3.1.2. Candesartan vs the active moiety: We speculate that Candesartan cilexetil will act as a prodrug and be metabolized to a more active form in vivo. Prodrugs have the potential to improve the drug delivery by enhanced aqueous solubility or permeability: 14% of all new approved drugs worldwide in 2001–2002 were prodrugs, and currently 5–7% of all marketed medicines can be classified as prodrugs.⁸⁶ Therefore, prodrugs should be considered as a potential design targets in the early stages of drug design and development.

Candesartan cilexetil selected by our SAR models is metabolized completely to the active Candesartan moiety,⁸⁷ which falls into the same binding motif as the Candesartan prodrug (Figure 9). Thus, we believe that the metabolized active form of this compound will have the same or better IC₅₀.

The Candesartan active moiety loses two hydrogen bonds with GLY 449 and TYR 448. However the tetrazole ring is present in both the prodrug and the active moiety of candesartan. Therefore we hypothesize that this deep hydrophobic interaction will stabilize the bound conformation and maintain the predicted binding affinity.

Prodrugs have already achieved certain commercial success in the pharmaceutical market but the concept is still relatively unknown or undervalued. This approach is not always the easiest and fastest way to improve the drug delivery. However, prodrugs may overcome unwanted pharmaceutical, pharmacokinetic, pharmacodynamic, or economical barriers of the parent drug molecule. To be successful in the future, more studies are needed to identify novel and creative prodrug structures. This may help to reduce the significant risks of attrition for lead compounds in drug development due to poor pharmacokinetics, efficacy or toxicological adverse effects.⁸⁸

3.3.2. Virtual Screening of the ChemNavigator Data base to Identify Novel Scaffolds—In drug discovery it is desirable to identify new chemical motifs and a diverse subset of molecules to avoid retesting of what are effectively duplicates. We employed the following procedure to identify novel chemical scaffolds relevant to NS5B inhibitors.

The ChemNavigator database⁵⁹ was refined to remove salts, inorganic compounds and filtered based on the physico-chemical properties for known actives using Pipeline Pilot 8.5.⁴⁸ From the resulting 24,870 compounds 10,399 scored highly in both SAR models. We selected a set of 200 diverse compounds by K-means clustering based on extended fingerprints from Accelrys Pipeline Pilot. From this set 87 compounds were purchased.

4. Biological Evaluation

4.1. Candesartan Cilexetil

The reference compounds were tested in vitro against HCV NS5B CON1 strain at concentrations ranging from 0.03 to 100 μ M. HCV-796, an allosteric inhibitor of HCV NS5B polymerase was used as a reference compound and yielded an IC₅₀ value of 0.04 μ M under our assay conditions. Candesartan Cilexetil (prodrug) exhibited an IC₅₀ value of 14.2 μ M in this assay. The activity for all assay points is available in the supplementary materials.

4.2. Two New Chemotypes Identified from ChemNavigator

Out of 87 diverse compounds selected from ChemNavigator two had IC₅₀ values below 50 μ M. These new chemotypes, compound **4** and compound **5**, are shown in Figure 10. The calculated physicochemical properties of both compounds are shown in Table 3.

The IC₅₀ values of the compounds were determined from dose–response curves using 8–12 concentrations of each compound in duplicate in two independent experiments. Curves were fitted to data points using nonlinear regression and IC₅₀ values were interpolated from the resulting curves using the GraphPad Prism 3.03 software.

The best pose of compound **4** has binding motif similar to compound **2** from the crystal structure 3E51 while the best pose of compound **5** is partially outside of the Palm binding pocket (Figure 11)

5. Experimental Procedures

5.1. NS5B inhibition assay for compounds selected from computational drug repositioning

The genotype used for the assays was 1b and the CON1 strain was used. The assay was performed in a 96-well plate format. NS5B activity was determined by measuring incorporation of radiolabeled ribonucleotides into synthesized RNA. The standard NS5B reaction mixture contained an in vitro transcribed viral RNA that is complementary to the HCV (–) strand 3'UTR region, radiolabeled ribonucleotides and rNTPs (for de novo polymerization). For compound screening, serially diluted test articles were added into the reactions. The reaction was initiated with the addition of the enzyme. After 2 hours incubation and extensively washing, radioactivity was measured using a microplate scintillation counter. The reduction of radioactivity was calculated to derive IC₅₀ (50% reduction of the signals compared to no compound control).

5.2. NS5B Inhibition Assay for Compounds Selected from ChemNavigator

Genotype 1b, strain J4L6 was used for the assays. This strain is 99.9% identical to CON1 and the NS5B enzymes in the two strains are identical, with the same susceptibility to inhibitors. Both strains are considered prototypical NS5B enzymes.

The anti-NS5B activity of the compounds was evaluated in vitro. The inhibition assay utilized poly rA/U₁₂ as the template-primer (TP) and wild-type recombinant NS5B Δ 21 protein as the enzyme source. Percent inhibition was determined at 100 μ M concentration of the indicated compound with 25 μ M UTP substrate and with 1 mM Mn as the divalent cation. All compounds were dissolved in dimethylsulfoxide (DMSO) as a 10 mM stock solution and stored at –20°C. Serial dilutions were made in DMSO immediately prior to the assay. Preliminary screening was conducted at 100 μ M compound concentration. Reactions were initiated by the addition of 1.0 mM MnCl₂, incubated at 30°C for 60 mins and terminated by the addition of ice-cold 10% (v/v) trichloroacetic acid (TCA) containing 0.5 mM pyrophosphate. The amount of radioactive UTP incorporated into nascent RNA was subjected to trichloroacetic acid (TCA) precipitation, spotted on a GF-B filter and quantified on a liquid scintillation counter (Packard). Activity of NS5B in the presence of equivalent amounts of DMSO was set at 100% and that in the presence of the inhibitor was calculated relative to this control. Percent inhibition was determined at 100 μ M concentration of the indicated compound on a polyA-U₁₂ template-primer, 25 μ M UTP substrate and with 1 mM Mn as the divalent cation. The data represents an average of at least two independent measurements in triplicate.

Conclusions

SAR models are being used more frequently for decision making nowadays but poor quality of predictions is still a major concern. The reliability of predictions for chemicals belonging to the model's applicability domain is much higher than for chemicals outside of this domain. However, staying within the applicability domain runs the risk of not identifying novel drug candidates. Going outside of the domain is a calculated risk in that it makes the model more susceptible to produce less accurate results. Occasionally taking this risk, albeit in a controlled way is necessary in the context of drug discovery.

As a result of our studies we identified Candesartan cilexetil, an existing prodrug for hypertension as a potential anti-HCV drug candidate. It has not previously been tested for direct anti-HCV activity. Given the predicted poses and SAR rationale, we suggest ways to improve the activities of the highlighted compound **1**. Our preliminary docking results indicates that replacement of benzyl ring in benzimidazole in candesartan cilexetil (**1**) with

thiazine-1,1-dioxide heterocycle can improve activity. The cilexetil ester is metabolized in human body. N-phenyl methanesulfonamide derivatives could prevent metabolism and improve activity.^{93,94}

We also found two novel non-nucleotide chemical motifs for NS5B inhibitors. We do not consider this to be the end of the story but rather an open door for further investigations to support decision making in the drug discovery process.

The lead hits selected during this study that are not commercially available have not been shown here but will be the subject of future studies on this target.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Some of the hardware used in these studies is part of the UMBC High Performance Computing Facility (HPCF). The facility is supported by the U.S. National Science Foundation through the MRI program (grant no. CNS-0821258, CNS-1228778) and the SCREMS program (grant no. DMS-0821311), with additional substantial support from the University of Maryland, Baltimore County (UMBC).

We gratefully acknowledge Dr. Michael G. Murray, Dr. Zhuhui Huang and Southern Research Institute for their support with NS5B inhibition assays and helpful discussions.

Funding Disclaimer:

This project has been funded in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract N01-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. This Research was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

This project has been funded in part with Federal Funds from the Division of Microbiology and Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institute of Health, Department of Health and Human Services, under contract HHSN272201100012I. HCV NS5B inhibition studies were supported by the National Institute of Health research grant CA153147 to N.K-B.

References and Notes

1. Tossing G. Management of chronic hepatitis C in HIV-co-infected patients-results from the First International Workshop on HIV and Hepatitis Co-infection. December 2–4, 2004, Amsterdam, Netherlands. *Eur J Med Res.* 2005; 10 (1):43–5. [PubMed: 15737953]
2. Chander G, Sulkowski MS, Jenckes MW, Torbenson MS, Herlong HF, Bass EB, Gebo KA. Treatment of chronic hepatitis C: a systematic review. *Hepatology.* 2002; 36 (5):S135–S144. [PubMed: 12407587]
3. Delang L, Coelmont L, Neyts J. Antiviral Therapy for Hepatitis C Virus: Beyond the Standard of Care. *Viruses.* 2010; 2 (4):826–866. [PubMed: 21994657]
4. McGuigan C, Gilles A, Madela K, Aljarah M, Holl S, Jones S, Vernachio J, Hutchins J, Ames B, Dawn Bryant K, Gorovits E, Ganguly B, Hunley D, Hall A, Kolykhalov A, Liu Y, Muhammad J, Raja N, Walters R, Wang J, Chamberlain S, Henson G. Phosphoramidate ProTides of 2'-C-Methylguanosine as Highly Potent Inhibitors of Hepatitis C Virus. Study of Their in Vitro and in Vivo Properties. *J Med Chem.* 2010; 53 (13):4949–4957. [PubMed: 20527890]
5. Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, Reindollar R, Goodman ZD, Koury K, Ling M, Albrecht JK. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. *Lancet.* 2001; 358 (9286):958–965. [PubMed: 11583749]

6. Singh S. Comparative QSAR Studies on the Novel Series of Thiazolones and Tetrazole Derivatives as HCV NS5B Polymerase Allosteric Inhibitors. *Lett Drug Des Discov*. 2009; 6 (4):286–297.
7. Musmuca I, Caroli A, Mai A, Kaushik-Basu N, Arora P, Ragno R. Combining 3-D QSAR 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 (4):662–676. [PubMed: 20225870]
8. Liu Q, Zhou H, Liu L, Chen X, Zhu R, Cao Z. Multi-target QSAR modelling in the analysis and design of HIV-HCV co-inhibitors: an in-silico study. *BMC Bioinformatics*. 2011; 12:294. [PubMed: 21774796]
9. Li Q, Wang Y, Bryant SH. A novel method for mining highly imbalanced high-throughput screening data in PubChem. *Bioinformatics*. 2009; 25 (24):3310–3316. [PubMed: 19825798]
10. Xie X, Chen J. Data mining a small molecule drug screening representative subset from NIH PubChem. *J Chem Inf Model*. 2008; 48 (3):465–475. [PubMed: 18302356]
11. <http://pubchem.ncbi.nlm.nih.gov/assay/>
12. The process of making this Roadmap file is available from cactus website.
13. <http://cactus.nci.nih.gov/download/roadmap/>
14. Stansfield I, Avolio S, Colarusso S, Gennari N, Narjes F, Pacini B, Ponzi S, Harper S. Active site inhibitors of HCV NS5B polymerase. The development and pharmacophore of 2-thienyl-5,6-dihydroxypyrimidine-4- carboxylic acid. *J Bioorg Med Chem Lett*. 2004; 14 (20):5085–5088.
15. Burton G, Ku TW, Carr TJ, Kiesow T, Sarisky RT, Lin-Goerke J, Baker A, Earnshaw DL, Hofmann GA, Keenan RM, Dhanak D. Identification of small molecule inhibitors of the hepatitis C virus RNA-dependent RNA polymerase from a pyrrolidine combinatorial mixture. *Bioorg Med Chem Lett*. 2005; 15 (6):1553–1556. [PubMed: 15745795]
16. Summa V, Petrocchi A, Matassa VG, Taliani M, Laufer R, De Francesco R, Altamura S, Pace P. HCV NS5b RNA- Dependent RNA Polymerase Inhibitors: From α,γ -Diketoacids to 4,5-Dihydroxypyrimidine- or 3-Methyl-5- hydroxypyrimidinonecarboxylic Acids. Design and Synthesis. *J Med Chem*. 2004; 47 (22):5336–5339. [PubMed: 15481971]
17. Di Santo R, Fermeiglia R, Ferrone M, Paneni MS, Costi R, Artico M, Roux A, Gabriele M, Tardif KD, Siddiqui A, Priol S. Simple but Highly Effective Three-Dimensional Chemical-Feature-Based \ Pharmacophore Model for Diketo Acid Derivatives as Hepatitis C Virus RNA-Dependent RNA Polymerase Inhibitors. *J Med Chem*. 2005; 48(20):6304–6314. [PubMed: 16190757]
18. Gopalsamy A, Aplasca A, Ciszewski G, Park K, Ellingboe JW, Orlowski M, Feld B, Howe AYM. Design and synthesis of 3,4- dihydro-1*H*-[1]-benzothieno[2,3-*c*]pyran and 3,4-dihydro-1*H*-pyrano[3,4- *b*]benzofuran derivatives as non-nucleoside inhibitors of HCV NS5B RNA dependent RNA polymerase. *Bioorg Med Chem Lett*. 2006; 16 (2):457–460. [PubMed: 16274990]
19. Powers JP, Piper DE, Li Y, Mayorga V, Anzola J, Chen JM, Jaen JC, Lee G, Liu J, Peterson MG, Tonn, Ye Q, Nigel PC, Walker, Wang Z. SAR and Mode of Action of Novel Non-Nucleoside Inhibitors of Hepatitis C NS5b RNA Polymerase. *J Med Chem*. 2006; 49 (3):1034–1046. [PubMed: 16451069]
20. Evans KA, Chai D, Graybill TL, Burton G, Sarisky RT, Lin-Goerke J, Johnston VK, Rivero RA. An efficient, asymmetric solid- phase synthesis of benzothiadiazine-substituted tetramic acids: Potent inhibitors of the hepatitis C virus RNA-dependent RNA polymerase. *Bioorg Med Chem Lett*. 2006; 16 (8):2205–2208. [PubMed: 16455253]
- 21G. opalsamy A, Chopra R, Lim K, Ciszewski G, Shi M, Curran KJ, Sukits SF, Svenson K, Bard J, Ellingboe JW, Agarwal A, Krishnamurthy G, Anita Y, Howe M, Orlowski M, Feld B, O'Connell J, Mansour TA. Discovery of Proline Sulfonamides as Potent and Selective Hepatitis C Virus NS5b Polymerase Inhibitors. Evidence for a New NS5b Polymerase Binding Site. *J Med Chem*. 2006; 49 (11):3052–3055. [PubMed: 16722622]
22. Summa V, Petrocchi A, Matassa VG, Gardelli C, Muraglia E, Rowley M, Paz OG, Laufer R, Monteagudo E, Pace P. 4,5- Dihydroxypyrimidine Carboxamides and *N*-Alkyl-5- hydroxypyrimidinone Carboxamides Are Potent, Selective HIV Integrase Inhibitors with Good Pharmacokinetic Profiles in Preclinical Species. *J Med Chem*. 2006; 49 (23):6646–6649. [PubMed: 17154493]

23. Beaulieu PL, Bös M, Bousquet Y, DeRoy P, Fazal G, Gauthier J, Gillard J, Goulet S, McKercher G, Poupart MA, Valois S, Kukolj G. Non-nucleoside inhibitors of the hepatitis C virus NS5B polymerase: discovery of benzimidazole 5-carboxylic amide derivatives with low-nanomolar potency. *Bioorg Med Chem Lett*. 2004; 14 (4):967–971. [PubMed: 15013003]
24. Pace P, Nizi E, Pacini B, Pesci S, Matassa V, De Francesco R, Altamura S, Summa V. The monoethyl ester of meconic acid is an active site inhibitor of HCV NS5B RNA-dependent RNA polymerase. *Chem Lett*. 2004; 14 (1):119–124.
25. Chan L, Reddy TJ, Proulx M, Das SK, Pereira O, Wang W, Siddiqui A, Yannopoulos CG, Poisson C, Turcotte N, Drouin A, Alaoui-Ismaili MH, Bethell R, Hamel M, L'Heureux L, Bilimoria D, Nguyen-Ba N. Identification of N,N-Disubstituted Phenylalanines as a Novel Class of Inhibitors of Hepatitis C NS5B Polymerase. *J Med Chem*. 2003; 46 (8):1283–1285. [PubMed: 12672227]
26. Pfefferkorn JA, Greene ML, Nugent RA, Gross RJ, Mitchell MA, Finzel BC, Harris MS, Wells PA, Shelly JA, Anstadt RA, Kilkuskie RE, Kopta LA, Schwende FJ. Inhibitors of HCV NS5B polymerase. Part 1: Evaluation of the southern region of (2*Z*)-2- (benzoylamino)-3-(5-phenyl-2-furyl)acrylic acid. *Bioorg Med Chem Lett*. 2005; 15(10):2481–2486. [PubMed: 15863301]
27. Mashino T, Shimotohno K, Ikegami N, Nishikawa D, Okuda K, Takahashi K, Nakamura S, Mochizuki M. Human immunodeficiency virus-reverse transcriptase inhibition and hepatitis C virus RNA-dependent RNA polymerase inhibition activities of fullerene derivatives. *Bioorg Med Chem Lett*. 2005; 15 (4):1107–1109. [PubMed: 15686922]
28. Gopalsamy A, Lim K, Ellingboe JW, Krishnamurthy G, Orlowski M, Feld B, Zeijl M, Howe AYM. Identification of [(naphthalene-1- carbonyl)-amino]-acetic acid derivatives as nonnucleoside inhibitors of HCV NS5B RNA dependent RNA polymerase. *Bioorg Med Chem Lett*. 2004; 14 (16):4221–4224. [PubMed: 15261274]
29. Dihydropyrones, Benzothiazine-substituted Quinolinediones, Carbon-linked Dihydropyrones, Cyclic N-Acetamideindolecarboxylic Acids, Acetamide-C2 Combinations, Benzothiadiazine and 1,4-benzothiazine Inhibitors, Dihydropyrones, Tetracyclic Inhibitors, N,N-Disubstituted Phenylalanine Inhibitors, Alpha-gamma Diketoacid Inhibitors, Thiophene-2-carboxylic acids.
30. Li H, Linton A, Tatlock J, Gonzalez J, Borchardt A, Abreo M, Jewell T, Patel L, Drowns M, Ludlum S, Goble M, Yang M, Blazel Y, Rahavendran R, Skor H, Shi S, Lewis C, Fuhrman S. Allosteric Inhibitors of Hepatitis C Polymerase: Discovery of Potent and Orally Bioavailable Carbon-Linked Dihydropyrones. *J Med Chem*. 2007; 50 (17):3969–3972.
31. Vicente J, Hendricks RT, Smith DB, Fell JB, Fischer J, Spencer SR, Stengel PJ, Mohr P, Robinson JE, Blake JF, Hilgenkamp RK, Yee C, Adjabeng G, Elworthy TR, Tracy J, Chin E, Li J, Wang B, Bamberg JT, Stephenson R, Oshiro C, Harris SF, Ghate M, Leveque V, Najera I, Pogam S, Rajyaguru S, Ao-Ieong G, Alexandrova L, Larrabee S, Brandl M, Briggs A, Sukhtankar S, Farrell R, Xu B. Non-nucleoside inhibitors of HCV polymerase NS5B. Part 2: Synthesis and structure–activity relationships of benzothiazine-substituted quinolinediones. *Bioorg Med Chem Lett*. 2009; 19:3642–3646. [PubMed: 19457662]
32. Li H, Tatlock J, Linton A, Gonzalez J, Jewell T, Patel L, Ludlum S, Drowns M, Rahavendran SV, Skor H, Hunter R, Shi ST, Herlihy KJ, Parge H, Hickey M, Yu X, Chau F, Nonomiya J, Lewis C. Discovery of (R)-6-Cyclopentyl-6-(2-(2,6-diethylpyridin-4-yl)ethyl)-3-((5,7-dimethyl-[1,2,4]triazolo[1,5-a]pyrimidin-2-yl)methyl)-4-hydroxy-5,6-dihydropyran-2-one(PF- 00868554) as a Potent and Orally Available Hepatitis C Virus Polymerase Inhibitor. *J Med Chem*. 2009; 52:1255–1258. [PubMed: 19209845]
33. Beaulieu PL, Jolicoeur E, Gillard J, Brochu C, Coulombe R, Dansereau N, Duan J, Garneau M, Jakalian A, Kühn P, Lagacé L, LaPlante S, McKercher G, Perrault S, Poirier M, Poupart M, Stammers T, Thauvette L, Thavonekham B, Kukolj G. *N*-Acetamideindolecarboxylic acid allosteric finger-loop' inhibitors of the hepatitis C virus NS5B polymerase: discovery and initial optimization studies. *Bioorg Med Chem Lett*. 2010; 20:857–861. [PubMed: 20074949]
34. McKercher G, Beaulieu PL, Lamarre D, LaPlante S, Lefebvre S, Pellerin C, Thauvette L, Kukolj G. Specific inhibitors of HCV polymerase identified using an NS5B with lower affinity for template/primer substrate. *Nucleic Acids*. 2004; 32 (2):422–431.
35. Li H, Tatlock J, Linton A, Gonzalez J, Borchardt A, Dragovich P, Jewell T, Prins T, Zhou R, Blazel J, Parge H, Love R, Hickey M, Doan C, Shi S, Duggal R, Lewis C, Fuhrman S.

- Identification structure-based optimization of novel dihydropyrones as potent HCV RNA polymerase inhibitors. *Bioorg Med Chem Lett*. 2006; 16:4834–4838. [PubMed: 16824756]
36. Hirashima S, Suzuki T, Ishida T, Noji S, Yata S, Ando I, Komatsu M, Ikeda S, Hashimoto H. Benzimidazole Derivatives Bearing Substituted Biphenyls as Hepatitis C Virus NS5B RNA-Dependent RNA Polymerase Inhibitors: Structure Activity Relationship Studies and Identification of a Potent and Highly Selective Inhibitor JTK- 109. *J Med Chem*. 2006; 49:6950–6953. [PubMed: 17125247]
37. Chan L, Reddy TJ, Proulx M, Das SK, Pereira O, Wang W, Siddiqui A, Yannopoulos CG, Poisson C, Turcotte N, Drouin A, Alaoui-Ismaili MH, Bethell R, Hamel M, L'Heureux L, Bilimoria D, Nguyen Ba. Identification of *N,N*-Disubstituted Phenylalanines as a Novel Class of Inhibitors of Hepatitis C NS5B Polymerase. *J Med Chem*. 2003; 46:1283–1285. [PubMed: 12672227]
38. Summa V, Petrocchi A, Pace P, Matassa VG, Francesco R, Altamura S, Tomei L, Koch U, Neuner P. Discovery of α,γ -Diketo Acids as Potent Selective and Reversible Inhibitors of Hepatitis C Virus NS5b RNA-Dependent RNA Polymerase. *J Med Chem*. 2004; 47:14–17. [PubMed: 14695815]
39. Chan L, Pereira O, Reddy TJ, Das SK, Poisson C, Courchesne M, Proulx M, Siddiqui A, Yannopoulos CG, Nguyen-Ba N, Roy C, Nasturica D, Moinet C, Bethell R, Hamel M, L'Heureux L, David M, Nicolas O, Courtemanche-Asselin P, Brunette S, et al. Discovery of thiophene-2-carboxylic acids as potent inhibitors of HCV NS5B polymerase and HCV subgenomic RNA replication. Part 2: Tertiary amides. *Bioorg Med Chem Lett*. 2004; 14:797–800. [PubMed: 14741292]
40. Filippov IV, Nicklaus MC. Optical Structure Recognition Software To Recover Chemical Information: OSRA -An Open Source Solution. *J Chem Inf Model*. 2009; 49 (3):740–743. [PubMed: 19434905]
41. <https://www.ebi.ac.uk/chembl>
42. <http://pharmaexpert.ru/passonline>
43. Phase, Schrödinger 2011.
44. <http://creativemachines.cornell.edu/eureqa>
45. <http://rdkit.org/>
46. Breiman, Leo. Random Forests. *Mach Learn*. 2001; 45(1):5–32.10.1023/A:101093340432
47. <http://code.google.com/p/rdkit/>
48. <http://accelrys.com/products/pipeline-pilot/>
49. <http://code.google.com/p/rdkit/wiki/ExplainingMorganFingerprintBits>
50. chembench.mml.unc.edu/theme/kNNQSAR.pdf, page 6
51. <http://en.wikipedia.org/wiki/KNN>
52. Fourches D, Muratov E, Tropsha A. Trust, but verify: On the importance of chemical structure curation in cheminformatics and QSAR modeling research. *J Chem Inf Model*. 2010; 50 (7):1189–1204. [PubMed: 20572635]
53. Sushko I, Novotarskyi S, Körner R, Pandey AK, Cherkasov A, Li J, Gramatica P, Hansen K, Schroeter T, Müller KM, Xi L, Liu H, Yao X, Oberg T, Hormozdiari F, Dao P, Sahinalp C, Todeschini R, Polishchuk P, Artemenko A, Kuz'min V, Martin TM, Young DM, Fourches D, Muratov E, Tropsha A, Baskin I, Horvath D, Marcou G, Muller C, Varnek A, Prokopenko VV, Tetko IV. Applicability Domains for Classification Problems: Benchmarking of Distance to Models for Ames Mutagenicity Set. *J Chem Inf Model*. 2010; 50(12):2094–2111. Published Online: October 29, 2010. 10.1021/ci100253r [PubMed: 21033656]
54. <http://www-935.ibm.com/services/us/gbs/bao/siip/nih/>
55. <http://www.chemspider.com/>
56. <http://xemistry.com/>
57. http://openbabel.org/wiki/Main_Page
58. <http://zinc.docking.org/>
59. <http://www.chemnavigator.com/>
60. <https://www.emolecules.com/>

61. Rigat K, Wang Y, Hudyma TW, Ding M, Zheng X, Gentles RG, Beno BN, Gao M, Roberts SB. Ligand-induced changes in hepatitis C virus NS5B polymerase structure. *Antivir Res.* 2010; 88(2): 197–206.10.1016/j.antiviral.2010.08.014 [PubMed: 20813137]
62. Berman HM, Westbrook J. *Nucleic Acids Res.* 2000; 28:235–242. [PubMed: 10592235]
63. Wang M, Ng KKS, Cherney MM, Chan L, Yannopoulos CG, Bedard J, Morin N, Nguyen-Ba N, Alaoui-Ismaili MH, Bethell RC, James MNG. Non-nucleoside Analogue Inhibitors Bind to an Allosteric Site on HCV NS5B Polymerase. Crystal Structures and mechanism of Inhibition. *JBiol Chem.* 2003; 278 (11):9489–9495. [PubMed: 12509436]
64. Ontoria JM, Rydberg EH, Marco S, Tomei L, Attenni B, Malancona S, Hernando JIM, Gennari N, Koch U, Narjes F, Rowley M, Summa V, Carroll SS, Olsen DB, De Francesco Raffaele, Altamura S, Migliaccio G, Carfi A. Identification and Biological Evaluation of a Series of 1*H*-Benzo[*de*]isoquinoline-1,3(2*H*)-diones as Hepatitis C Virus NS5B Polymerase Inhibitors. *J Med Chem.* 2009; 52(16):5217–5227. Published Online: July 24, 2009. 10.1021/jm900517t [PubMed: 19877603]
65. Biswal BK, Wang M, Cherney MM, Chan L, Yannopoulos CG, Bilimoria D, Bedard J, James MNG. Non- nucleoside Inhibitors Binding to Hepatitis C Virus NS5B Polymerase Reveal a Novel Mechanism of Inhibition. *J Mol Biol.* 2006; 361 (1):33–45. [PubMed: 16828488]
66. Nyanguile O, Devogelaere B, Vijgen L, Van den Broeck W, Pauwels F, Cummings MD, De Bondt HL, Vos AM, Berke JM, Lenz OL, Vandercruyssen G, Vermeiren K, Mostmans W, Dehertogh P, Delouvroy F, Vendeville S, VanDyck K, Dockx K, Cleiren E, Raboisson P, Simmen KA, Fanning GC. 1a/1b Subtype Profiling of Nonnucleoside Polymerase Inhibitors of Hepatitis C Virus. *Journal Virol.* 2010; 84 (6):2923–2934.
67. Li H, Tatlock J, Linton A, Gonzalez J, Borchardt A, Dragovich P, Jewell T, Prins T, Zhou Ru, Blazel J, Parge H, Love R, Hickey M, Doan C, Shi S, Duggal R, Lewis C, Fuhrman S. Identification and structure-based optimization of novel dihydropyrones as potent HCV RNA polymerase inhibitors. *Bioorg Med Chem Lett.* 2006; 16 (18):4834–4838. [PubMed: 16824756]
68. Li H, Tatlock J, Linton A, Gonzalez J, Jewell T, Patel L, Ludlum S, Drowns M, Rahavendran SV, Skor H, Hunter R, Ski ST, Herlihy KJ, Parge H, Hickey M, Yu X, Chau F, Nonomiya J, Lewis C. Discovery of (*R*)-6-Cyclopentyl-6-(2-(2,6-diethylpyridin-4-yl)ethyl)-3-((5,7-dimethyl-[1,2,4]triazolo[1,5-*a*]pyrimidin-2-yl)methyl)-4-hydroxy-5,6-dihydropyran-2-one (PF-00868554) as a Potent and Orally Available Hepatitis C Virus Polymerase Inhibitor. *J Med Chem.* 2009; 52 (5):1255–1258. [PubMed: 19209845]
69. Ruebsam F, Murphy DE, Tran CV, Li S, Zhao J, Dragovich PS, McGuire HM, Xiang AX, Sun Z, Ayida BK, Blazel JK, Kim SH, Zhou Y, Han Q, Kissinger CR, Webber SE, Showalter RE, Shah AM, Tsan M, Patel RA, Thompson PA, LeBrun LA, Hou HJ, Kamran R, Sergeeva MV, Bartkowski DM, Nolan TG, Norris DA, Khandurina J, Brooks J, Okamoto E, Kirkovsky L. Discovery of tricyclic 5, 6-dihydro-1*H*-pyridin-2-ones as novel, potent, and orally bioavailable inhibitors of HCV NS5B polymerase. *Bioorg Med Chem Lett.* 2009; 19 (22):6404–6412. [PubMed: 19818610]
70. Vicente J, Hendricks RT, Smith DB, Fell JB, Fischer J, Spenser SR, Stengel PJ, Mohr P, Robinson JE, Blake JF, Hilgenkamp RK, Yee C, Adjabeng G, Elworthy TR, Li J, Wang B, Bamberg JT, Harris SF, Wong A, Leveque VJP, Najera I, Pogam S, Rajyaguru S, Ao-leong G, Alexandrova L, Larrabee S, Brandi M, Briggs A, Sukhtankar S, Farrell R. Non- nucleoside inhibitors of HCV polymerase NS5B. Part 4: Structure-based design, synthesis, and biological evaluation of benzo[*d*]isothiazole-1,1-dioxides. *Bioorg Med Chem Lett.* 2009; 19 (19):5652–5656. [PubMed: 19709881]
71. Dragovich PS, Blazel JK, Ellis DA, Kamran QR, Kissinger SR, LeBrun LA, Lian-Sheng Li, Murphy DE, Noble M, Patel RA, Ruebsam F, Sergeeva MV, Shah AM, Showalter RE, Tran CV, Tsan M, Webber SE, Kirkovsky L, Ahou Y. Novel HCV NS5B polymerase inhibitors derived from 4-(1',1'-dioxo-1',4'-dihydro-1' λ -benzo[1',2',4']thiadiazin-3'-yl)-5-hydroxy-2*H*-pyridazin-3-ones. Part 5: Exploration of pyridazinones containing 6-amino-substituents. *Bioorg Med Chem Lett.* 2008; 18 (20):5635–5639. [PubMed: 18796353]
72. NRB—number of rotatable bonds; NAR—number of aromatic rings; NHA-number of hydrogen bond acceptors; NHD- number of hydrogen bond donors; MW-Molecule Weight; PSA-Polar Surface Area.

73. GLIDE. Schrödinger, LLC; New York, NY: 2011.
74. Chemical Computing Group's Molecular Operating Environment (MOE), 2010.
75. http://stitch.embl.de/cgi/show_input_page.pl
76. <http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm>
77. <http://nih-pipeline.citeline.com/CpAccount.aspx>
78. <http://sideeffects.embl.de/>
79. <http://www.drugcite.com/>
80. Tuncer I, Ozbek H, Ugras S, Bayram I. Anti-fibrogenic effects of captopril and candesartan cilexetil on the hepatic fibrosis development in rat: The effect of AT1-R blocker on the hepatic fibrosis. *Exp Toxic Pathol*. 2003; 55 (2–3):159–166.
81. Ueki M, Koda M, Shimizu T, Mitsuta A, Yamamoto T, Murawaki Y. Effect of an angiotensin-II type-1 receptor blocker candesartan on hepatic fibrosis in chronic hepatitis C: a prospective study. *Hepato-gastroenterol*. 2009; 56 (93):1100–4.
82. http://en.wikipedia.org/wiki/Hepatic_stellate_cell
83. <http://clinicaltrials.gov/ct2/results?term=candesartan+cilexetil&Search=Search>
84. Zhou Y, Webber SE, Murphy DE, Li LS, Dragovich PS, Tran CV, Sun Z, Ruebsam F, Shah AM, Tsan M, Showalter RE, Patel R, Li B, Zhao Q, Han Q, Hermann T, Kissinger CR, Lebrun L, Sergeeva MV, Kirkovsky L. Novel HCV NS5B polymerase inhibitors derived from 4-(1', 1'-dioxo-1', 4'-dihydro-1'lambda6-benzo[1', 2', 4']thiadiazin-3'-yl)-5-hydroxy-2H-pyridazin-3-ones. Part 1: exploration of 7'-substitution of benzothiadiazine. *Bioorg Med Chem Lett*. 2008; 18 (14):1413–8. [PubMed: 18242088]
85. Li LS, Zhou Y, Murphy DE, Stankovic N, Zhao J, Dragovich PS, Bertolini T, Sun Z, Ayida B, Tran CV, Ruebsam F, Webber SE, Shah AM, Tsan M, Showalter RE, Patel R, Lebrun LA, Bartkowski DM, Nolan TG, Norris DA, Kamran R, Brooks J, Sergeeva MV, Kirkovsky L, Zhao Q, Kissinger CR. Novel HCV NS5B polymerase inhibitors derived from 4-(1', 1'-dioxo-1', 4'-dihydro-1'lambda(6)-benzo[1', 2', 4']thiadiazin-3'-yl)-5-hydroxy-2H-pyridazin-3-ones. Part 3: Further optimization of the 2-, 6- and 7'-substituents and initial pharmacokinetic assessments. *Bioorg Med Chem Lett*. 2008; 18 (11):3446–55. [PubMed: 18457949]
86. Borchardt, RT.; Middaugh, CR. *Biotechnology: Pharmaceutical Aspects*. In: Stella, VJ.; Borchardt, RT.; Hageman, MJ.; Oliyai, R.; Maag, H.; Tilley, JW., editors. *Prodrugs: Challenges and Rewards Part 1 & 2*. Vol. V. Springer Science + Business Media; New York: 2007.
87. http://en.wikipedia.org/wiki/Candesartan_cilexetil
88. Huttunen KM, Rautio J. Prodrugs - An Efficient Way to Breach Delivery and Targeting Barriers. *Curr Top Med Chem*. 2011; 11 (18):2265–2287. [PubMed: 21671868]
89. We calculated aqueous solubility using program MOE. It outputs the aqueous Solubility expressed as logS, where S is the solubility in mol/L. The method used to estimate the solubility is the multiple linear regression model based on electrotopological State indices published by Tetko, et al. *J Chem Inf Comput Sci*. 2001; 41:1488–1493. "Estimation of Aqueous Solubility of Chemical Compounds Using E-State Indices. [PubMed: 11749573]
90. S+Sw – native water solubility [mg/ml]. Calculated using ADMET Predictor from Simulation Plus.
91. We calculated logP (o/w) using program MOE.
92. Ghose AK, Viswanadhan VN, Wendoloski JJ. Alog P was calculated using Pipeline Pilot from Accelrys. The Ghose/Crippen estimate of molar refractivity which contains information about molecular volume and polarizability of a compound. Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragment Methods: An Analysis of AlogP and CLogP Methods. *J Phys Chem A*. 1998; 102:3762–3772.
93. <http://www.drugbank.ca/drugs/DB00489>
94. Smith, DA. *Metabolism, Pharmacokinetics and Toxicity of Functional Groups*. Royal Society of Chemistry; 2010. p. 242

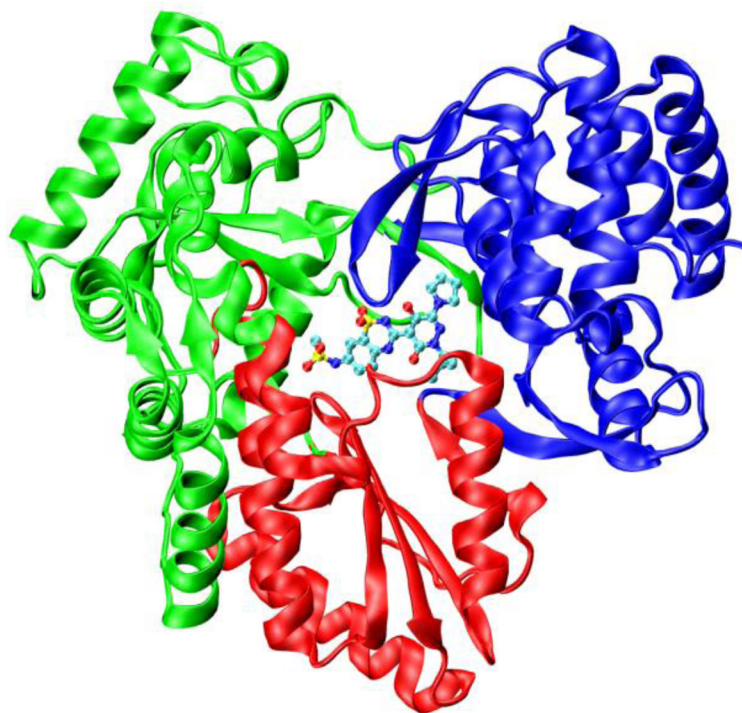


Fig. 1. NS5B/inhibitor complex from PDB structure 3E51 with the characteristic fingers, palm, and thumb domains, shown in green, red and blue, respectively.

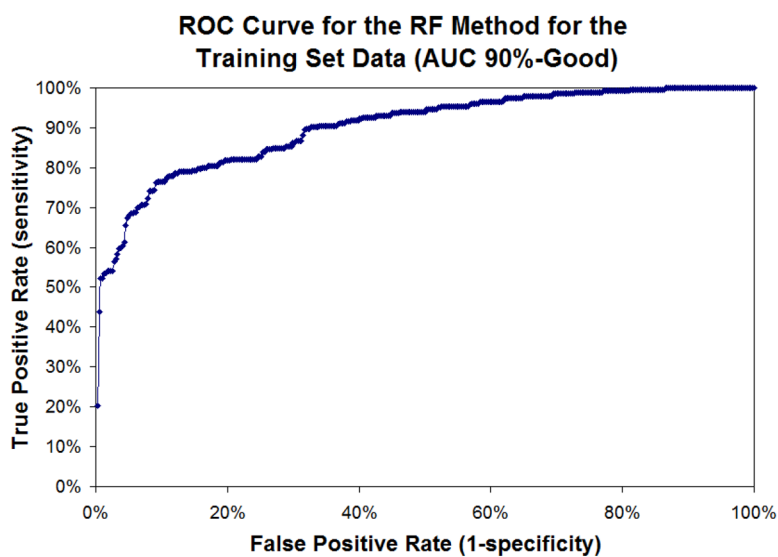


Fig. 2. ROC curve for the RF method for the training set data. The AUC is 90%. The ROC curve shows that the model is able to differentiate active vs. inactive compounds when applied to the training set itself.

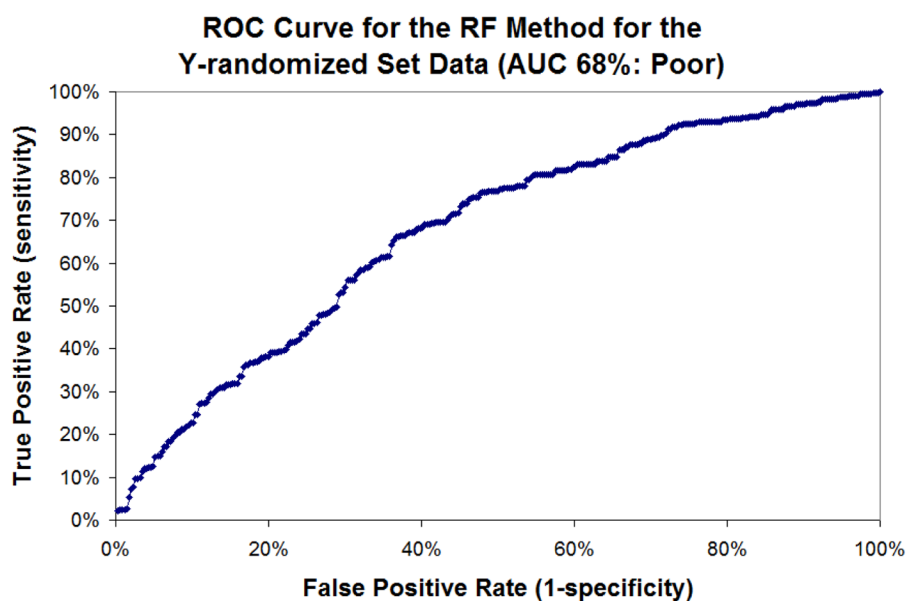


Fig. 3.
ROC curve for the RF method for the Y-randomized set data with the outcome randomized with a 50% chance of being active. The AUC for this model applied to its training set is only 68% which shows that our approach does not overfit the data.

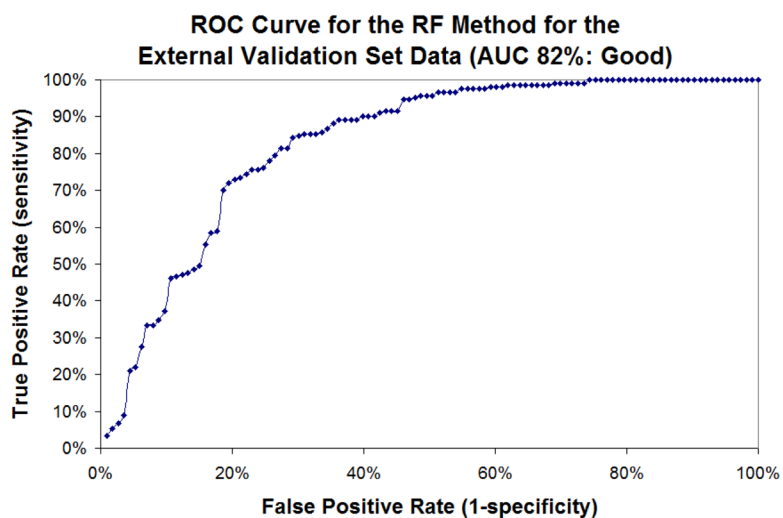


Fig. 4. ROC curve for the RF method for the external validation set data (323 molecules from ChEMBL⁴¹). The AUC for this test set is 82%, which shows good predictivity.

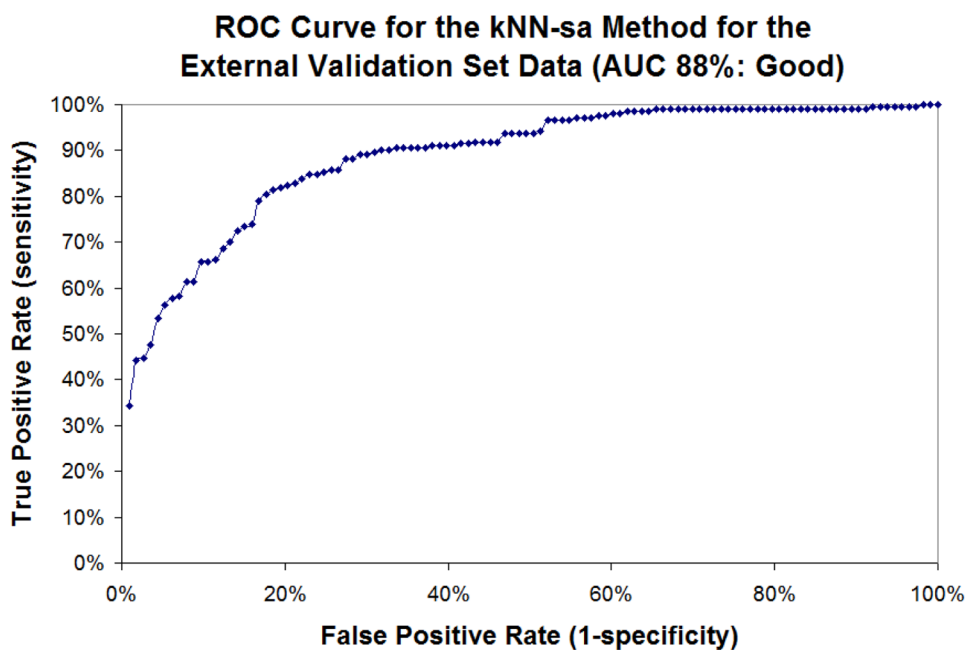


Fig. 5. ROC curve for the kNN-sa method for the external validation set data (323 molecules from ChEMBL⁴¹ with known IC₅₀ values). The AUC for this test set is 88%, which shows good predictivity.

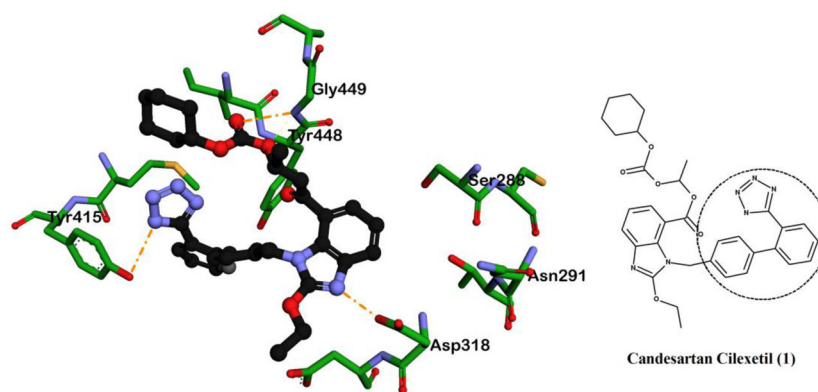


Fig. 6. (Left) Key binding site interactions for the conformation of the best pose of Candesartan cilexetil **1** (black ball and stick representation). Hydrogen bonds (GLY 449, ASP 318 and TYR 415) are represented as orange dashed lines. (Right) The molecular structure of Candesartan cilexetil **1** showing the biphenyl-tetrazole substituent circled in dashed lines. Hydrogens are not shown for clarity.

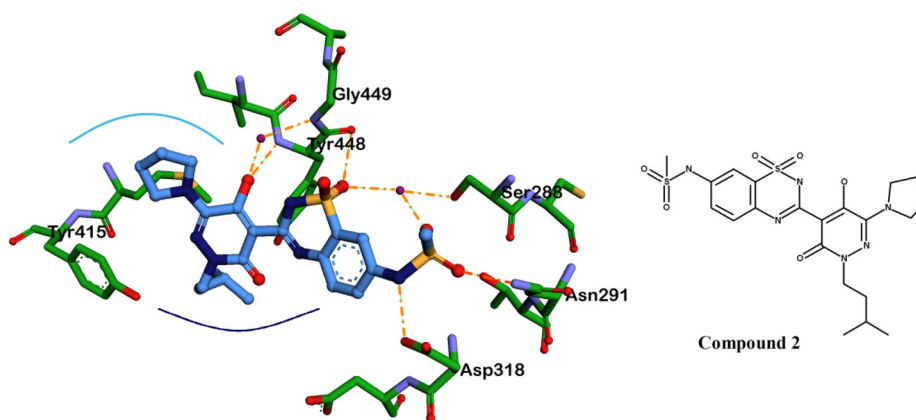


Fig. 7. (Left) X-ray crystal structure of compound **2** bound to the Palm domain of NS5B (PDB ID 3E51, Resolution 1.9 Å). Compound **2** is colored in blue and represented as ball-and-stick model (also shown in Fig. 1). It forms hydrogen bonds (represented as orange dashed lines) with GLY 449, TYR 448, ASP 318, ASN 291 and SER 288 as well as two water molecules HOH 617 and HOH 590 (shown in violet). The deep hydrophobic pocket comprised of NS5B residues PRO 197, ARG 200, LEU 384, CYS 366, MET 414, TYR 415, and TYR 448 is schematically displayed as a dark blue line. In contrast, the shallow cleft near the surface of the enzyme consisting of residues GLY 410, MET 414, and GLN 446 is shown schematically as a light blue line. (Right) The molecular structure of compound **2**. Hydrogens and hydrophobic interactions are not shown for clarity.

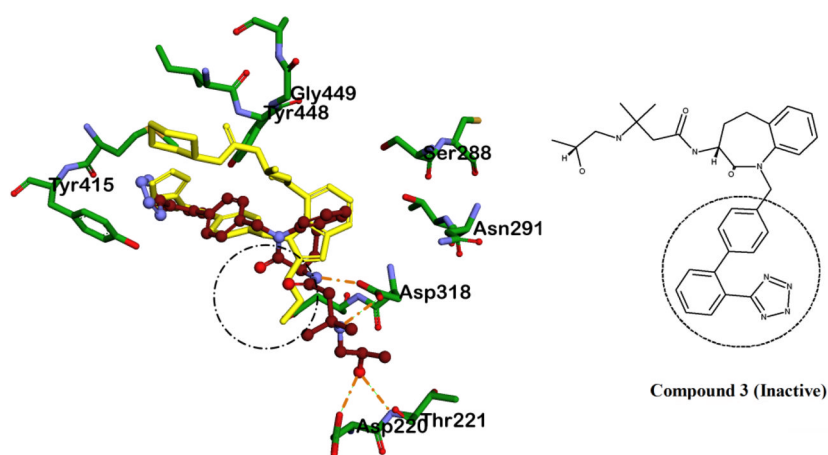


Fig. 8. (Left) Conformation of the best pose of inactive compound **3** (brown) superimposed with the best pose of Candesartan cilexetil (**1**) (yellow) from crystal structure 3E51 (Palm). Compound **3** forms hydrogen bonds (represented as orange dashed lines) with ASP 318, ASP 220 and THR 221. Two polar carbonyl groups present in compound **3** and discussed in the text are circled in dashed lines. Hydrogens are not shown for clarity. (Right) The molecular structure of compound **3** showing the biphenyl-tetrazole substituent circled in dashed lines.

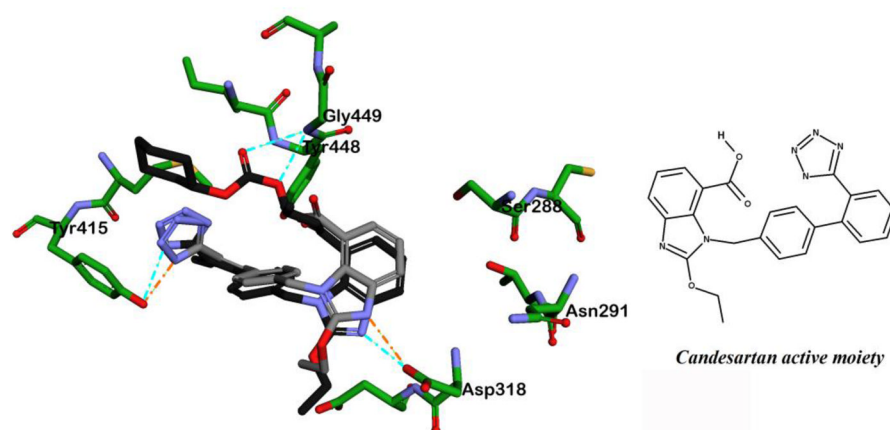


Fig. 9.

Conformation of the best pose of Candesartan cilexetil **1** (black) superimposed with the best pose of the Candesartan active moiety (grey). Hydrogen bonds are represented as dashed lines in orange and light blue for compound **1** and the Candesartan active moiety, respectively. The active moiety forms hydrogen bonds with TYR 415 and ASP 318. Hydrogens are not shown for clarity.

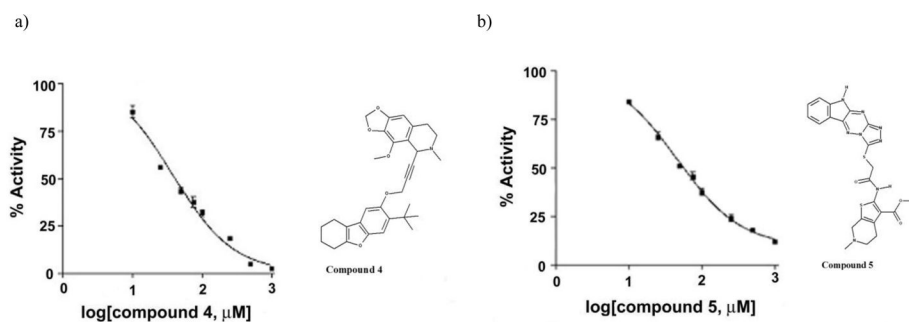


Fig. 10. Dose-response curves for Compound **4** ($\text{IC}_{50}=38.0 \pm 2.2 \mu\text{M}$, panel a) and Compound **5** ($\text{IC}_{50}=46.3 \pm 4.4 \mu\text{M}$, panel b). The molecular structures of compounds are as indicated.

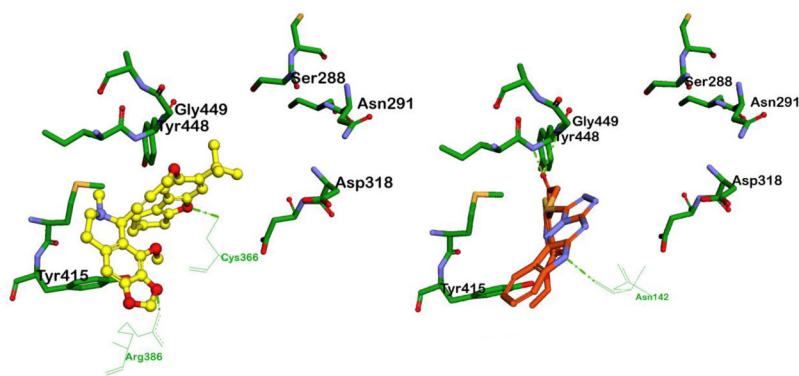


Fig. 11.

Best docking poses of compound **4** (left, yellow ball and stick representation) and compound **5** (right, orange tube representation). Compound **4** forms hydrogen bonds to CYS 366 and ARG 386 while compound **5** forms hydrogen bonds to ASN 142, TYR 448 and GLY 449.

Table 1

The performance of the RF and kNN-sa SAR models. The Area Under the Curve (AUC) shows the performance of the model in percentage scale.

Dataset	AUC
Self-prediction (training set) (RF) (Fig. 2)	90%
Cross validation (RF)	86%
Y-randomization (RF) (Fig. 3)	68%
External validation (ChEMBL (323) -k-NN-sa (Fig. 4)	88%
External validation (ChEMBL (323) – RF (Fig. 5)	82%

Table 2

Databases, number of molecules predicted as active by both models, and number of commercially available compounds among predicted structures.

DATABASES	Total number of structures	Total number of structures predicted as active by both SAR models	Commercially available structures
Beilstein Natural Products	72920	412	19
Comprehensive Herbal Medicine Information System for Cancer (CHMIS-C)	8006	34	1
ComGenex (now AMRI Hungary) database of drug-like small molecules	184265	130	11
DrugBank	6269	15	5
MDL Drug Data Report (MDDR)	164420	1718	101
MDL Patent Chemistry	30580	315	54
Traditional Chinese Medicine (TCM)	135334	124	24
IBM NIH Donation	2407585	2508	3

Table 3Calculated physicochemical properties of compound **4** and compound **5**.

Compound	Molecular Solubility ⁸⁹	S+Sw ⁹⁰	LogP (o/w) ⁹¹	AlogP ⁹²
4	-10.170	$1.08 \cdot 10^{-4}$	6.725	2.834
5	-4.977	$1.06 \cdot 10^{-2}$	2.458	3.810