Bioorganic & Medicinal Chemistry 20 (2012) 2455–2478

Contents lists available at SciVerse ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

Integrated structure-based activity prediction model of benzothiadiazines on various genotypes of HCV NS5b polymerase (1a, 1b and 4) and its application in the discovery of new derivatives

Mohamed A.H. Ismail *, Dalal A. Abou El Ella, Khaled A.M. Abouzid, Amr H. Mahmoud

Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Ain Shams University, Cairo 11566, Egypt

ARTICLE INFO

Article history:
Received 21 November 2011
Revised 25 December 2011
Accepted 5 January 2012
Available online 28 January 2012

Keywords:
HCV NS5b polymerase inhibitor
Topomer CoMFA
AMANDA
Guided docking
Homology
Library design
Field alignment
Fragment-hopping

ABSTRACT

This work presents the first structure-based activity prediction model for benzothiadiazines against various genotypes of HCV NS5b polymerase (1a, 1b and 4).The model is a comprehensive workflow of structure-based field template followed by guided docking. The field template was used as a pre-filter and a tool to provide hits in good orientation and position. It was created based on detailed molecular interaction field analysis which includes Topomer CoMFA, grid independent analysis and Superstar. On the other hand, Guided docking was used as a refinement and assessment tool. It was actively directed by two scores: Moldock score as an interaction descriptor (r^2 = 0.65) and a template similarity score as a measure for accurate binding-mode compliance. The docking template was based on energy-based pharmacophore analysis. The whole procedure was formulated and tweaked for both screening (ROC of AUC = 0.91) and activity prediction (r^2 of 0.8) for the genotype 1a. In order to widen the model scope, linear interaction energy was used as a tool for predicting activities of other genotypes based on the docked ligand poses while mutation binding energy was used to investigate the effect of each amino acid mutation in genotype 4. The model was applied for structure-based fragment hopping by screening a library designed by reaction enumeration. A top scoring hit was used to generate a focused library such that it has lower TPSA than the original class ligands and thus better pharmacokinetic properties. After that, experimental validation was carried out by the synthesis of this library and its biological evaluation which yielded compounds that exhibit EC_{50} ranging from 1.86 to 23 μM.

1. Introduction

The hepatitis C virus (HCV) is a member of the Flaviviridae family. Chronic infection with this virus is associated with liver cirrhosis that often leads to hepatic failure and hepatocellular carcinoma. Although the number of new infections has been significantly reduced by the introduction of reliable blood testing, more than 170 million people worldwide are chronically infected with HCV, which has become a global health threat and the main cause of adult liver transplants in developed nations. There is no effective therapy for HCV-associated chronic hepatitis up till now. As a result, hepatitis C is considered a major public health threat and there is a growing unmet medical need to discover novel therapies.[1,2] Employing HCV proteins as targets, directly acting antiviral agents have been identified and collectively described as 'specifically targeted antiviral therapy for HCV' (STAT-C).[3–5] Among the nonstructural proteins, NS3–4A protease, NS5B polymerase, NS3 helicase and NS5A have been the object of intense research efforts both by academia and pharmaceutical companies.[6,7] NS5B RNA-dependent RNA polymerase is recognized as a key target for therapeutic intervention mainly because it is not present in mammalian cells and offers a wide range of possibilities for the discovery of new molecular entities as anti-HCV agents.[8–11] Mechanistic and structural studies of this enzyme have revealed the existence of multiple allosteric binding sites, and in particular two thumb sites (thumb I and II) and three palm pockets (palm I, II and III) have been identified to date.[12] According to the target site, the different inhibitors will be referred to as palm site I NNIs (PSI-NNIs), palm site II NNIs (PSII-NNIs), palm site III NNIs (PSIII-NNIs), thumb site I NNIs (TSI-NNIs) and thumb site II NNIs (TSII-NNIs). Out of these different allosteric sites and their corresponding inhibitors, we focused this study on palm I site and in particular on benzothiadiazines as one of the main palm I-NNI[13] (non-nucleoside inhibitors) (see Figs. S1 and S2 in the Supplementary data and see full account on the different allosteric sites in Section 1.1 HCV general information).

The main aims of this study can be outlined in these points:

1. The development for the first time a structure-based activity prediction model for the benzothiadiazine class which has wider applicability domain[14] than that of the ligand based ones.

* Corresponding author. Tel.: +20 224051120; fax: +20 24051107.
E-mail address: mhismaeel@yahoo.com (M.A.H. Ismail).

0968-0896/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2012.01.031

Table 1
Mini review on the CADD attempts for benzothiadiazines and benzothiazines

Hendricks et al.	Ab initio calculations were employed for conformational analysis of a series of compounds[61,62] The importance of internal hydrogen bonding was highlighted as the main method to achieve NS5b-bound conformations
Melagraki et al.	Melagraki et al. built MLR QSAR model using lipophilicity, HOMO energy, Ki2, KiInf0 and KiInf3 as descriptors.[15] This was carried out for Substituted hydroxyquinolone-benzothiadiazine and substituted naphthyridone-benzothiadiazine derivatives
Pourbasheer et al.	Pourbasheer et al.[16] built MLR and SVM models using seven descriptors (hydration energy, PW4, GATS2m, GATS7m, GATS5p, RDF095e and Mor30m) to study pyridazinone-benzothiadiazine
Chen et al.	Chen et al.[17] carried out both 2D-QSAR and 3D-QSAR using MFA coupled with G/PLS to study Benzoisothiazoles dioxide, benzothiophenes dioxide
Wang et al.	Wang et al.[18] developed ligand-based and receptor-based CoMFA and CoMSIA models for large number of benzothiadiazines. Homology model was developed for genotype 1a in this study. It was used to obtain receptor-based alignment for CoMFA analysis via automated docking procedure. Besides, binding mode was studied via Molecular dynamics experiment
Li et al.	Li et al.[20] used MD-simulation to study binding mode of benzothiadiazines and other palm I classes. MM-GB/SA calculations together with free energy decomposition were deployed in this study

This wider scope is attributed to the usage of the knowledge of the three dimensional structure of the biological target obtained through x-ray crystallography. This model is considered a receptor-based QSAR model which was mainly influenced by the fact that previous prediction models by Melagraki et al.[15] Pourbasheer et al.[16] Chen et al.[17] and Wang et al[18] were ligand-based (Table 1) with limited applicability domain.[19] They were likely able to predict only the activities of benzothiadiazines with similar properties to that of the training dataset. Additionally we realized that all the previous structure-based models were used to predict binding mode only.[18,20] Applicability domain method deployed is explained in Section 1.7 in the Supplementary data.

2. Extending the applicability of the model by being able to predict the activities of various genotypes. This is because the palm I site is the least conserved among the different allosteric binding sites of HCV polymerase[21] which affects the activity of the benzothiadiazine class leading to a drastic decrease in potency against genotypes other than genotype 1. The effect is even obvious in the subtypes of genotype 1 itself (10-fold decrease in genotype 1a than 1b). Thus, we wanted to increase the scope of this model to be able to predict the activities of various genotypes. Here we included three genotypes: 1a,1b and 4.
3. Applying the model in virtual screening and identification of new derivatives of this specific class of benzothiadiazines which may have better pharmacokinetic properties especially the topological polar surface area (TPSA). This is mainly due to the fact that this class suffers from poor bioavailability due to the large polarity of the ligands that hinders the permeability.[22]

The most important criterion of this work was the desire to find hits that avoid the common chemotype trap in finding benzothiadinzes derivatives. By this, we are pointing to the fact that most patents depend on simple markush structure similarity. This is very obvious where nearly most of the patents focused on systems having in common a keto-enol system in the ring fused with the benzothiadiazine (e.g. hydroxyquinolinone system). Thus, we shifted to structure-based virtual screening that does not depend on active inhibitors similarity. However, we did not want to leave it unguided such that it may result in a plethora of irrelevant hits that do not have the essential features required for activity. Thus, we decided to support this model by knowledge retrieved from literature concerning known inhibitors of this class. In another words, we decided to generate a guided receptor-based activity prediction model.

2. Design process

Being a workflow, we decided to describe it all over first then give the results of each part. The structure-based activity prediction model presented in this study consists of two elements: structure-based field template[23] followed by guided docking.[24] The field template (first main component of this workflow) consists of field points which describe the van der Waals and electrostatic (both positive and negative) minima and maxima[25] that surround molecules. It is considered a structure-based tool as it takes the bound conformer as a template to construct field points while considering the protein as an excluded volume. In addition, It consists of hotspots which represent nodes or field points of high relevance in ligand–protein interaction. These hotspots are treated in the field template as constraints and are extracted using molecular interaction field analysis as will be described later. The field template is used as a pharmacophore pre-filter and a tool to align ligands in the binding site and providing them in a good orientation before docking. The second main component of the workflow is the guided docking module. It is a type of docking which incorporates chemical information to actively guide the orientation of the ligand into the binding site during the ligand sampling. It is used as an efficient knowledge-based strategy for binding affinity estimation, ligand binding-mode prediction, virtual screening enrichment and further ligand optimization.[24] Herein, we used Moldock that is based on evolutionary algorithm (EA)[26] in sampling that is guided by chemical information represented in the form of pharmacophore template that reflects the essential features for interaction. More specifically, it is directed by a consensus scoring function which is divided into two:

1. A score for ligand–protein interaction that correlates well with activity. This score is used to optimize the ligand in the binding site according to the interactions.
2. A template similarity score which reflects the compliance of the inhibitor with the proper binding-mode (i.e., mapping with the pharmacophore properly). This score applies similarity corrections throughout the entire ligand incremental construction process, thus affecting not only the relative ordering of the solutions but also actively guiding the ligand placement such it complies with the main reference of ligand–protein complex.

Thus, the requirements of the guided docking are a pharmacophore template and a scoring function which correlates well with the activity. These requirements were extracted according to innovative procedures applied in the workflow which is described below.

Systematically, we constructed the aforementioned model and applied it using an innovative workflow which is depicted in Figures 1 and 2 and consists of the following steps:

1. Detailed SAR analysis was done using molecular interaction field analysis methods, molecular dynamics and energy-based pharmacophore[27] to extract hotspots and pharmacophore features. As shown in Figure 1, SAR analysis had three roles. It was used to construct structure-based field template

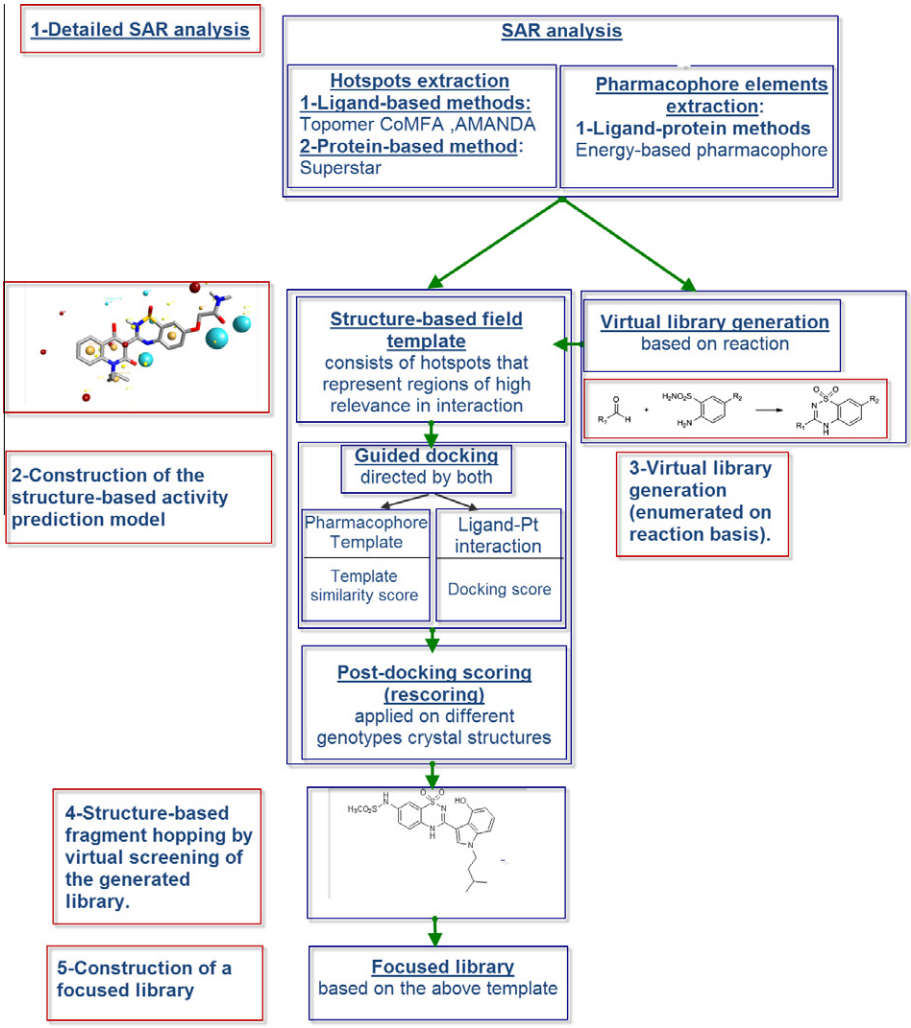


Figure 1. The workflow of the construction and the application of the structure-based activity prediction model for the benzothiadiazine class.

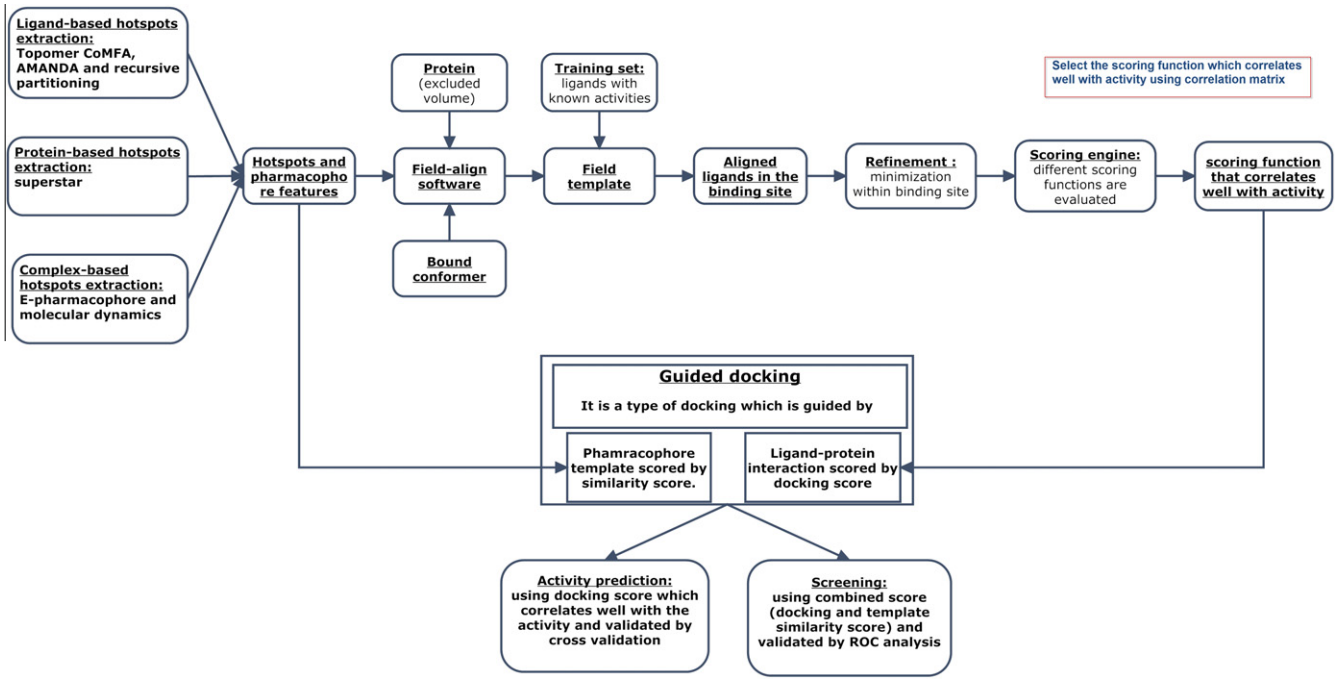


Figure 2. The workflow used to construct and validate the structure-based activity prediction model of benzothiadiazines.

(pharmacophore pre-filter), guided docking pharmacophore template and aided in the designing of the virtual library which was later screened.

2. Construction of the structure-based activity prediction model according to the following steps:
   a. Generating a structure-based field template using the first genotype 1a crystal structure recently released[28] where the bound ligand was used as a template to generate field points while the hotspots (extracted above) were used as constraints.
   b. Setting up the guided docking preferences to direct the ligands properly by:
      i. Constructing a pharmacophore template to be used for guided docking. This was guided by the SAR analysis done in step 1.
      ii. Finding the scoring function that correlates well with the activity of this class of inhibitors. This is depicted in Figure 2. As shown, a group of ligands with reported activities were aligned using the field template in the binding site then refined by energy minimization and scored using different scoring engines to find the best scoring function that correlates well with the activity.
   c. Validation of the model statistically using cross validation (to assess activity predication capability) and ROC analysis (to assess screening capability).[29,30]
   d. Extending the model applicability domain to describe the different activities among various genotypes: linear interaction energy[31] method was used as a post-docking scoring function to calculate binding energies of the ligands against various genotypes (genotype 1a, 1b and 4).The calculations were done using crystal structures of genotypes 1a and 1b and a homological model for genotype 4.
3. A library of benzothiadiazines was enumerated on reaction basis such that it complies with SAR analysis done in step 1.
4. Structure-based fragment hopping was carried out by virtual screening of the generated library using the validated model in order to find new derivatives.
5. Construction of a focused library based on a top scoring hit such that it has lower TPSA followed by its synthesis and biological evaluation to validate model experimentally.

3. Results and discussion

3.1. Molecular modeling

3.1.1. SAR analysis

SAR analysis was carried out in order to extract Hotspots and the complementary pharmacophore features. This is a crucial step in which knowledge related to this class is extracted. As described earlier, this knowledge will be used to construct the structure-based field template, the guided docking pharmacophore template and finally generating a virtual library based on this class of compounds.

3.1.1.1. Hotspots extraction. Hotspots are highly relevant regions for ligand interactions which can be used for accurately placing ligands in the binding site. They were extracted using molecular interaction field (MIF) analysis techniques which can be applied by different methods to study the ligand–receptor interactions. MIF can be computed in the receptor binding site in order to scan regions of importance for interaction, as was originally described by Goodford.[32] This was carried out in the protein-based methods below using Superstar program. It is also possible to proceed the other way around and compute MIF in one or many small compounds in order to characterize them according to their potential to act as ligands, binding a certain receptor. When used in this manner, the MIF can be seen as computationally obtained descriptor variables ('molecular descriptors'), which represent properties of the molecules (e.g., AMANDA[33]) or contour maps (e.g., Topomer CoMFA[34]). Both AMANDA and Topomer CoMFA were carried out in the ligand-based methods. The details of these analyses are listed below:

3.1.1.1.1. Ligand based methods. The methods used here are QSAR methods which deals mainly with the variations among structures (it does not describe constant part) thus they were restricted on two functionalities: studying SAR of the variable substituents and developing virtual screening QSAR models.

Topomer CoMFA and AMANDA analyses were deployed. These are two different field based techniques; the first uses the C.sp3 and H+ probes to represent steric and electrostatic fields while the second use different types of probes (O, N1, probes) which reflect H-bond donor, H-bond acceptor and hydrophobic fields. In this way, the information obtained from the two methods can be complementary and not redundant as it covers different types of fields. The two MIF analysis techniques are non-alignment based and the combination of them, as far as we know, has not been publicly utilized in literature unlike COMFA and COMSIA. These tools were used to avoid alignment step which is the time consuming step of the classic MIF analysis.

Recursive partitioning[35] was used in addition as a fingerprint-based decision tree. It was not used to extract hotspots but simply aided in the rapid understanding of the AMANDA and Topomer CoMFA results as will be described later.

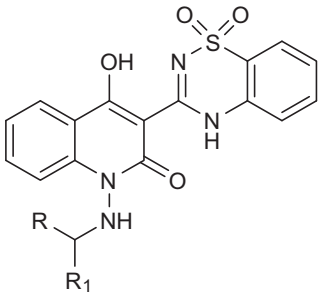
The three aforementioned techniques were carried out using a dataset of 98 benzothiadiazine inhibitors retrieved from literature[36–38] (see Tables 2–6). The activities used were those of genotype 1a. The results of each technique are as following:

3.1.1.1.1.1.*Topomer COMFA.* Topomer CoMFA is a technique introduced by Cramer.[34] It delineates the need for alignment which is mandatory for typical CoMFA analysis. For space saving; only figures with important outcomes are shown. Topomer CoMFA has both graphical and statistical results. Concerning the graphical results, It was used to construct stdev*coeff contour maps to show field effects on the target features. The contour plots are beneficial to identify important regions where some changes in steric or electrostatic fields can affect the biological activity. The maps generated depict regions having scaled coefficients greater than 80% (favored) or less than 20% (disfavored).

CoMFA sterically favorable (green) contours are observed (Fig. 3A) adjacent to the 3-methylbutyl group of compound side chain (the hydrophobic tail substituted at hydroxyquinolinone N) and this indicates the importance of the hydrophobic tail in this site. It is surrounded by sterically unfavorable yellow contours which point out that this site is a narrow pocket and should be occupied by hydrophobic tails that are not too bulky. Another group of CoMFA sterically favored green contours are present at methansulfonamide substituent group which is also surrounded by sterically unfavorable yellow contours indicating that there is a limit for this side chain expandability (Fig. 3B). In the CoMFA electrostatic contour maps, the red contours show favorable electronegative regions, and the blue contours show regions where the electropositive charges are favored. A close inspection of the electrostatic contour plots reveals that, for the tested molecules, electropositive and electronegative pattern which maps perfectly with aminosulfonamide (Fig. 3C) is more preferred surrounding side chain at position-7 of the benzothiadiazine.

Regarding the statistical results of Topomer CoMFA analysis, the model was validated internally yielding a q^2 of 0.777 and r^2 of 0.864 with 3 optimum components (PLS components). This

Table 2
Structures and activities of *N*-1-heteroalkyl-4-hydroxyquinolone-3-yl-benzothiadiazines



Name	FR	Structure		Observed $-\log IC_{50}$		Topomer CoMFA			AMANDA					LIAISON	
		R	R1	Genotype 1a	Genotype1b	PRED	R1	R2	LV1	LV2	LV3	LV4	LV5	Genotype1a pred	Genotype1b pred
1	A1	Ph	H	−0.71	0.88	−0.94	−0.05	−0.69	−0.357	−0.634	−0.776	−0.437	−0.559	0.293282	1.002365
2	A1	2-BrC_6H_4	H	−0.83	0.32	−0.91	−0.05	−0.66	−0.421	−0.656	−0.893	−0.775	−0.811	−0.52737	0.536115
3	A1	3-BrC_6H_4	H	−0.02	1.12	−0.86	−0.05	−0.61	−0.456	−0.364	−0.258	−1.164	−1.296	0.382967	1.064247
4	A1	4BrC_6H_4	H	−1.02	−0.67	−1.02	−0.05	−0.77	−0.305	−0.390	−0.536	−0.368	−0.384	−2.01536	−0.6206
5	A1	2-MeC_6H_4	H	−0.79	0.31	−1.09	−0.05	−0.84	−0.432	−0.587	−0.667	−0.674	−0.867	−0.64675	0.443742
6	T1	3-MeC_6H_4	H	−0.91	0.67	−1.00	−0.05	−0.75	−0.407	−0.294	−0.238	−0.820	−0.879	−0.81434	0.138103
7	A1	2-Thienyl	H	−0.37	0.90	−0.76	−0.05	−0.51	−0.494	−0.470	−0.405	−0.555	−0.586	0.032644	0.722524
8	A1	2-Thiazolyl	H	−1.55	0.17	−0.76	−0.05	−0.51	−0.451	−0.510	−0.537	−0.744	−0.911	−0.35388	0.455823
9	A1	2-Furyl	H	−0.51	0.51	−0.72	−0.05	−0.47	−0.501	−0.403	−0.298	−0.466	−0.488	−0.71322	0.30788
10	A1	3-Furyl	H	−0.19	0.87	−0.48	−0.05	−0.23	−0.467	−0.434	−0.372	−0.527	−0.646	0.109668	0.975671
11	A1	3-Me-thieny-2-yl	H	−1.18	−0.07	−0.79	−0.05	−0.54	−0.419	−0.686	−0.754	−0.712	−0.726	−0.47609	0.471498
12	A1	5-Cl-thienyl-2-yl	H	−0.68	−0.16	−0.84	−0.05	−0.59	−0.410	−0.234	−0.138	−0.328	−0.380	−0.57943	0.370195
13	A1	Pr	H	0.02	0.79	0.08	−0.05	0.33	−0.523	−0.294	−0.025	−0.058	0.006	1.02182	1.595055
14	A1	Bu	H	0.03	1	0.09	−0.05	0.34	−0.435	−0.091	0.220	0.219	0.035	0.332452	0.929392
15	A1	*i*-Bu	H	0.20	0.87	0.03	−0.05	0.28	−0.499	−0.200	0.089	−0.001	−0.111	0.601349	1.114931
16	A1	Neopentyl	H	0.39	0.86	−0.22	−0.05	0.03	−0.496	−0.377	−0.106	−0.013	0.265	−0.61384	0.376449
17	A1	*i*-Pr	H	0.03	0.99	−0.09	−0.05	0.16	−0.462	−0.660	−0.619	−0.497	−0.293	0.16641	1.014823
18	A1	Cyclopropyl	H	0.55	0.83	−0.03	−0.05	0.22	−0.482	−0.327	−0.106	−0.127	−0.004	0.945155	1.452157
19	A1	Cyclohexyl	H	−0.40	0.82	−0.47	−0.05	−0.23	−0.437	−0.264	−0.068	−0.107	−0.113	0.797223	1.320084
20	A1	Me	Me	−0.80	0.19	0.05	−0.05	0.30	−0.524	−0.204	0.130	0.054	−0.037	−1.00277	0.198086
21	A1	Et	Et	−0.39	0.92	0.12	−0.05	0.37	−0.414	−0.232	−0.072	−0.135	−0.206	−0.09094	0.637255
22	A1	Et	Pr	−0.28	0.41	0.14	−0.05	0.39	−0.495	−0.426	−0.296	−0.357	−0.268	0.423538	0.992241
23	A1	Pr	Pr	−0.35	−0.25	0.06	−0.05	0.31	−0.381	−0.296	−0.254	−0.418	−0.442	−0.25025	0.627329
24	A1	Pr	*i*-Pr	−0.81	−0.05	−0.16	−0.05	0.09	−0.442	−0.569	−0.497	−0.395	−0.345	0.19382	1.033736
25	A1	Me	Ph	−1.22	−0.17	−0.89	−0.05	−0.64	−0.455	−0.496	−0.474	−0.617	−0.726	−0.91748	0.166939
26	A1	Cyclobutyl		0.56	0.96	0.25	−0.05	0.50	−0.511	−0.399	−0.198	−0.258	−0.101	0.955955	1.429609
27	A1	Cyclopentyl		0.13	0.82	0.29	−0.05	0.54	−0.468	−0.350	−0.170	−0.182	−0.015	−0.87332	0.287409
28	A1	Cyclohexyl		0.45	1.2	0.25	−0.05	0.50	−0.391	−0.148	0.042	0.045	0.099	0.58855	1.1061
29	A1	Cycloheptyl		−0.02	0.66	0.06	−0.05	0.31	−0.477	−0.370	−0.216	−0.286	−0.225	0.382967	0.964247
30	A1	4-Pyrnayl		−0.67	0.35	−0.09	−0.05	0.16	−0.420	−0.424	−0.426	−0.625	−0.821	0.534419	1.168749

number of components identified in the LOO (leave one out) cross-validation is used in the final non-cross-validated PLS run (see results in Tables 2–6).

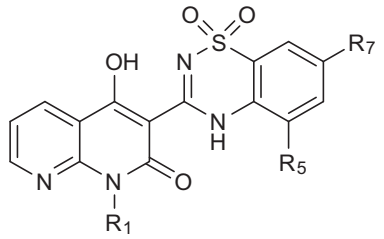
This model can be used for screening, especially R-group screening, as the compounds were divided into 2 across the bond joining ring B and C into R1 and R2 (see Section 5). However, this was not the aim of Topomer CoMFA here; instead, it was used for SAR analysis and extracting the hotspots represented in some field points of highly relevance in interactions.

3.1.1.1.1.2. *Pentacle GRIND*. As the name implies, this MIF analysis depends on GRIND which is short for Grid-independent descriptors.[39–41] It is described in details in Section 1.2 of the Supplementary data. The data set of compounds (98 compounds) was used as an input for Pentacle (see Tables 2–4 and 6). Default probes (DRY, O, and N1) were used obtaining three auto-correlograms (DRY–DRY, O–O, and N1–N1) and three cross-correlograms (DRY–O; DRY–N1; O–N1). The DRY probe represents hydrophobic interaction, O is sp2 carbonyl oxygen and has hydrogen-bond acceptor properties, and N1 is a neutral flat NH and has hydrogen-bond donor properties. The 6 types of GRIND variables were used in PLS (partial leas square) to give a model.

In general, a Pentacle model gives graphical (Fig. 4) as well statistical results (Table 7). Simply we used the numerical values of the PLS coefficients to select the GRIND able to correlate with biological activity (Table 8), whereas the distances were used to visualize the pair of nodes that have been used to assign a value to the GRIND variable (Fig. 4). Inspection of this table confirms the importance of the hotspots at the regions of the methansulfonamide at 7-position, hydrophobic tail and the enolic oxygen. It is obvious that pentacle did not add much to SAR analysis regarding the substituents (same information as Topomer CoMFA. However, the model can be used for rapid virtual screening of the whole ligands which can add to the R-screening in Topomer CoMFA. Practically, the optimum number of PLS components (latent variables, LV) was chosen by monitoring changes in the model's predictivity index evaluated by applying the cross-validation procedure available in Pentacle (see Fig. S3 in Supplementary data). It was three just like Topomer CoMFA.

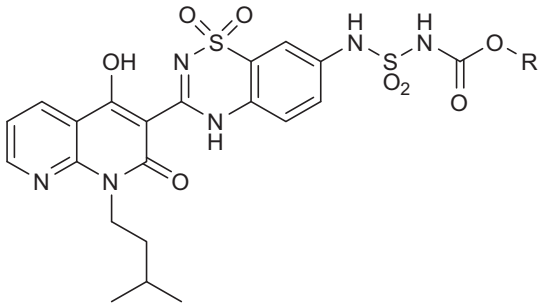
3.1.1.1.1.3. *Recursive partitioning*. Recursive partitioning is a technique to classify members of the population (the training set here) according to several dichotomous dependent variables (fingerprints) in a decision tree. A total of 4 tree(s) were created.

Table 3
Structures and activities of *N*-1-benzyl and *N*-1-(3-methylbutyl)-4-hydroxy-1,8-naphthyridon-3-yl benzothiadiazine analogs containing substituents on the aromatic ring



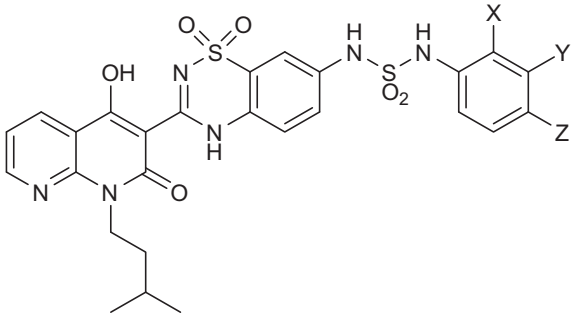
Name	FR	Structure			Observed $-\log IC_{50}$		Topomer CoMFA			AMANDA					LIAISON	
		R1	R5	R7	Genotype 1a	Genotype1b	PRED	R1	R2	LV1	LV2	LV3	LV4	LV5	Genotype1a pred	Genotype1b pred
31	A1	Benzyl	H	H	−0.76	1.0	−0.94	−0.05	−0.69	−0.412	−0.752	−0.901	−0.872	−0.824	−0.96343	0.235235
32	A1	Benzyl	OMe	H	−1.26	−0.32	−1.37	−0.47	−0.69	−0.438	−0.932	−1.187	−1.138	−1.118	−0.95527	0.140864
33	A1	Benzyl	H	OMe	−0.92	1.11	−0.63	0.26	−0.69	−0.194	−0.705	−0.968	−0.856	−0.771	−0.2196	0.618475
34	A1	Benzyl	OH	H	−0.92	1.26	−1.16	−0.26	−0.69	−0.008	−0.071	−0.388	−0.537	−0.611	−0.82273	0.32232
35	A1	Benzyl	H	Me	−1.41	0.63	−0.63	0.27	−0.69	−0.395	−0.843	−1.061	−1.102	−1.237	−0.40993	0.417148
36	A1	Benzyl	Me	H	−0.83	1.01	−1.20	−0.30	−0.69	−0.451	−0.886	−1.095	−1.082	−1.165	−0.52608	0.337008
37	A1	Benzyl	H	Br	−1.22	0.12	−0.76	0.14	−0.69	−0.388	−0.766	−0.980	−1.018	−1.073	−0.82401	0.231433
38	A1	Benzyl	Br	H	−0.79	−0.06	−1.12	−0.22	−0.69	−0.416	−0.853	−1.084	−1.057	−1.000	−1.79239	−0.33675
39	A1	3-Methylbutyl	H	H	0.09	1.38	−0.08	−0.05	0.17	−0.477	−0.828	−0.883	−0.708	−0.353	0.231515	0.959745
40	A1	3-Methylbutyl	OMe	H	−1.25	−0.06	−0.51	−0.47	0.17	−0.455	−0.726	−0.782	−0.754	−0.827	−0.84502	0.186936
41	A1	3-Methylbutyl	H	OMe	−0.05	1.18	0.23	0.26	0.17	−0.242	−0.719	−0.876	−0.689	−0.503	1.146922	1.681376
42	A1	3-Methylbutyl	H	OH	0.41	1.37	−0.14	−0.11	0.17	0.774	0.911	0.593	0.670	0.765	0.208935	0.844165
43	A1	3-Methylbutyl	H	$-OCH_2CH_2CH_3$	−0.21	0.031	0.49	0.52	0.17	0.025	−0.440	−0.505	−0.143	0.129	0.090485	0.762435
44	A1	3-Methylbutyl	H	$-OCH_2CO_2t$-Bu	−0.70	0.07	−0.24	−0.21	0.17	0.670	−0.174	−0.314	−0.089	−0.262	0.00103	0.800711
45	A1	3-Methylbutyl	H	$-OCH_2COOH$	0.44	1.24	0.22	0.25	0.17	1.350	0.562	0.400	0.288	0.471	0.535334	1.26938
46	A1	3-Methylbutyl	H	$-OCH_2CONMe_2$	0.03	1.15	0.38	0.41	0.17	0.412	−0.173	−0.314	0.034	−0.019	1.029653	1.510461
47	A1	3-Methylbutyl	H	$-OCH_2CONHMe$	0.74	1.096	0.36	0.39	0.17	0.969	0.648	0.284	0.400	0.168	1.044727	1.490862
48	A1	3-Methylbutyl	H	$-OCH_2CONH_2$	1.34	1.79	0.44	0.47	0.17	1.471	1.356	1.303	1.141	1.213	1.73724	2.088696
49	A1	3-Methylbutyl	H	$--OCH_2CH_2NH_2$	0.20	1.29	0.44	0.48	0.17	1.308	0.934	0.823	0.461	0.739	−0.80414	0.145144
50	A1	3-Methylbutyl	H	$-OCH_2CN$	0.85	1.60	0.30	0.33	0.17	0.432	0.080	0.184	0.797	1.049	0.990781	1.383639
51	A1	*i*-Pentyl	H	$-NH_2$	0.51	1.63	0.28	0.31	0.17	0.637	1.536	1.422	1.290	1.165	0.908638	1.42696
52	A1	*i*-Pentyl	H	$-NH_2CH_2CN$	0.92	1.61	0.65	0.69	0.17	0.964	1.143	1.024	1.212	0.939	2.120819	2.363365
53	A1	*i*-Pentyl	H	$-NH_2CH_2CONH_2$	0.33	1.46	0.88	0.91	0.17	1.217	1.384	1.086	1.050	1.034	0.127902	0.888252
54	A1	i-Pentyl	H	$-NHCOCF_3$	0.04	1.15	1.03	1.06	0.17	0.654	0.545	0.289	0.366	0.195	0.342872	1.006582
55	A1	*i*-Pentyl	H	$-NHSO_2Ph$	1.39	2.15	1.73	1.77	0.17	0.689	1.145	1.338	1.665	1.571	2.08722	2.330182
56	A1	*i*-Pentyl	H	$-NHSO_2iPr$	2.10	2.52	1.94	1.97	0.17	0.519	0.789	0.885	1.195	1.536	2.19691	2.215868
57	A1	*i*-Pentyl	H	$-NHSO_2(CH_2)_3CH_3$	1.66	2.30	1.56	1.59	0.17	0.750	1.327	1.410	1.635	1.679	2.65758	2.53373
58	A1	*i*-Pentyl	H	$-NHSO_2Me$	2.70	2.22	1.56	1.59	0.17	0.817	1.551	1.520	1.725	2.005	2.99897	2.869289

Table 4
Structures and activities of *N*-alkyl-4-hydroxyquinolon-3-yl-benzothiadiazine



Name	Structure	FR	Observed $-\log IC_{50}$		Topomer CoMFA			AMANDA					Liaison	
	R		Genotype1a	Genotype1b	PRED	R1	R2	LV1	LV2	LV3	LV4	LV5	Genotype1a pred	Genotype1b pred
59	CH_3	A1	1.20		1.18	1.21	0.17	0.817	1.551	1.520	1.725	2.005		
60	CH_2CH_2Cl	A1	0.72		1.10	1.13	0.17	0.938	1.182	0.796	0.778	1.057		
61	CH_2CHCH_2	A1	1.06		1.13	1.16	0.17	1.305	1.219	1.081	1.028	0.962		
62	CH_2CCH	A1	1.14		1.15	1.18	0.17	1.311	1.111	0.989	1.038	0.740		
63	CH_2CH_2CN	A1	1.46		1.10	1.13	0.17	1.115	1.407	1.259	1.319	1.078		
64	CH_2Ph	A1	1.24		1.10	1.13	0.17	1.520	1.255	1.334	1.414	1.617		
65	$CH_2CH_2NH_2$	A1	1.68		1.10	1.14	0.17	1.388	1.387	1.335	1.408	0.948		
66	$CH_2CO_2CH_2CH_3$	A1	1.21		1.10	1.14	0.17	1.777	2.111	1.850	1.475	1.613		
67	$CH_2CH_2OCH_3$	A1	1.06		1.11	1.14	0.17	1.470	1.033	0.689	0.633	0.703		
68	$CH_2CH_2OCH_2Ph$	A1	1.02		1.11	1.14	0.17	1.496	0.743	0.779	0.717	1.017		
69	CH_2COOH	A1	1.30		1.09	1.12	0.17	1.288	1.107	0.839	0.984	0.777		

Table 5
Structures and activities of N-alkyl-4-hydroxyquinolon-3-yl-benzothiadiazine



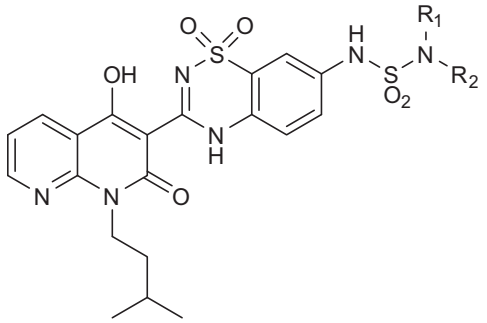
Name	Structure			FR	Observed $-\log IC_{50}$		Topomer CoMFA			AMANDA					Liaison	
	X	Y	Z		Genotype1a	Genotype1b	PRED	R1	R2	LV1	LV2	LV3	LV4	LV5	Genotype1a pred	Genotype1b pred
70	H	H	COOMe	A1	1.32		1.45	1.49	0.17	1.773	0.705	0.964	0.837	1.265		
71	H	H	COOH	A1	1.70		1.42	1.45	0.17	1.327	1.462	1.331	1.386	1.457		
72	H	H	CONHMe	A1	1.37		1.48	1.51	0.17	1.648	0.984	1.628	1.663	1.620		
73	H	H	$CONH_2$	A1	1.74		1.47	1.50	0.17	1.470	1.410	1.550	1.486	1.128		
74	H	COOEt	H	A1	0.88		1.17	1.20	0.17	1.836	1.659	1.964	1.858	1.649		
75	H	COOH	H	A1	0.94		1.23	1.26	0.17	1.066	1.517	1.348	1.359	1.122		
76	H	$CONH_2$	H	A1	1.37		1.24	1.27	0.17	1.651	1.323	1.198	0.856	0.667		
77	H	$CONHCH_2CONH_2$	H	A1	1.37		1.19	1.22	0.17	1.744	1.460	1.666	1.478	1.334		
78	COOMe	H	H	A1	0.89		1.19	1.22	0.17	2.079	1.023	1.826	1.574	1.209		

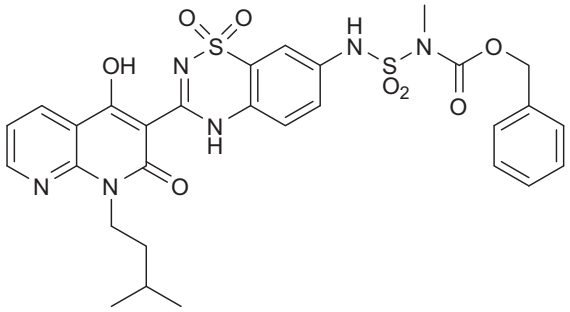
The best tree was the first tree with a cross-validated ROC score of 0.746 (see Fig. S4 in Supplementary data).

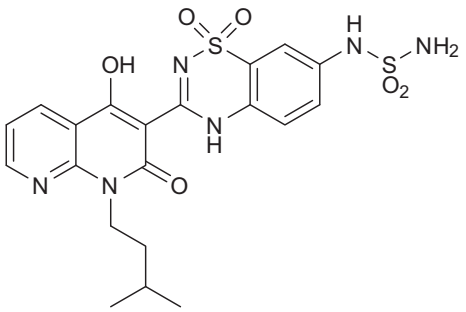
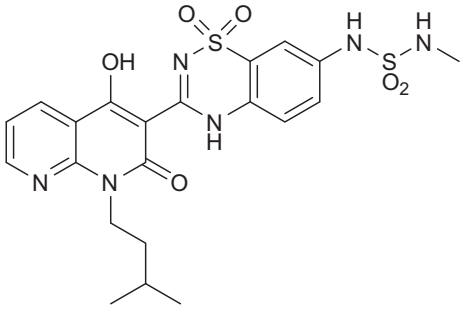
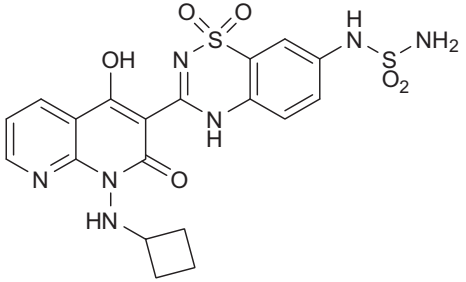
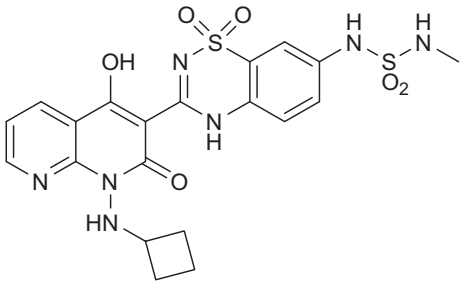
The compounds were divided according to their activity into bins ranging from 1 to 10. The tree begins by dividing the whole series into two major categories: those with amino sulfonamide and those without. Compounds with amino sulfonamide are more active: Activity-bin was 6 and above. The tree also indicates that the isoamyl moiety is the best hydrophobic tail in this class. In addition, it points out that ring A, as pyridine, is more active than benzene. Regarding aminosulfonamide, it should be short in general indicating that it binds in a narrow pocket as increasing bulkiness decreases activity relatively. The pocket hydrophilic nature prefers polar expandability than those hydrophobic substituents with aromatic ring.

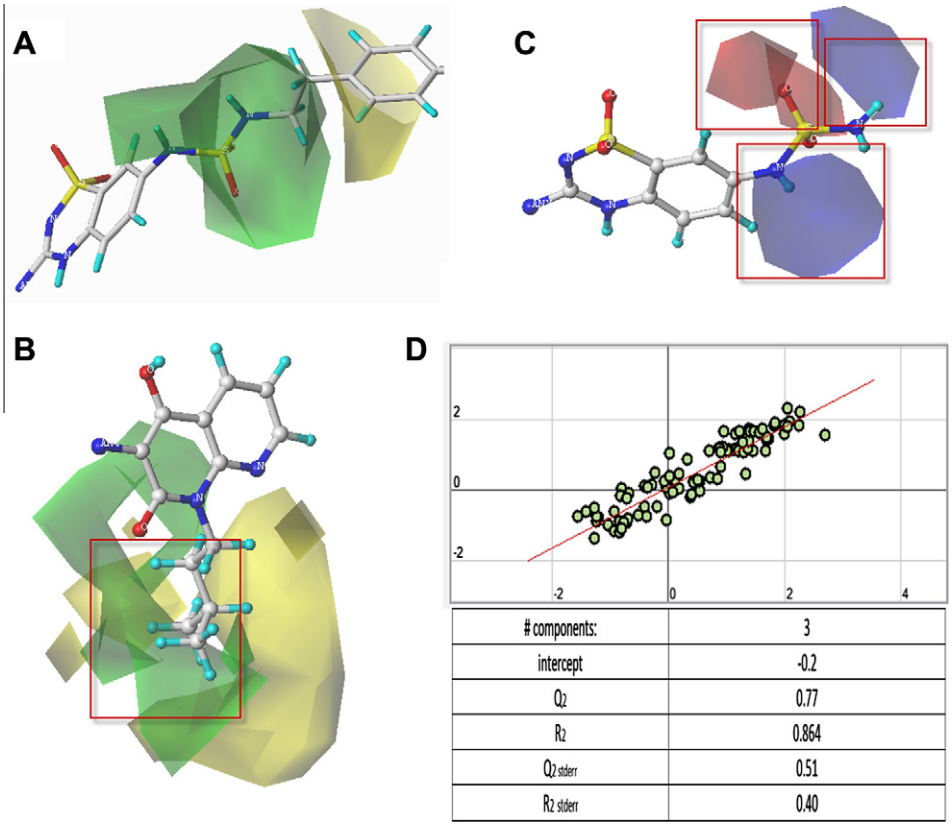
Specifically here, the recursive partitioning was used in order to aid the interpretation of MIF Analysis by pointing out to those

Table 6
Structures and activities of *N*-alkyl-4-hydroxyquinolon-3-yl-benzothiadiazine



Name	Structure		FR	Observed $-\log IC_{50}$		Topomer CoMFA			AMANDA					Liaison	
	R1	R2		Genotype 1a	Genotype1b	PRED	R1	R2	LV1	LV2	LV3	LV4	LV5	Genotype1a pred	Genotype1b pred
79	H	CH_2CH_2Ph	A1	1.74		1.47	1.50	0.17	1.550	1.280	1.060	1.142	1.031		
80	H	Ch_2Ph	A1	2.26		1.83	1.86	0.17	1.111	1.782	1.862	1.975	1.809		
81	H	Ph	A1	1.39		1.40	1.43	0.17	1.172	1.811	1.938	2.027	1.722		
82	H	CH_2CH_2OH	A1	0.97		1.65	1.68	0.17	0.862	1.414	1.454	1.603	1.456		
83	H	Cyclohexyl	A1	1.17		1.61	1.65	0.17	1.040	1.327	1.218	1.196	1.363		
84	H	Cyclopentyl	A1	1.41		1.68	1.71	0.17	0.831	1.400	1.405	1.497	1.499		
85	H	$CH_2CH_2NH_2$	A1	2.00		1.61	1.65	0.17	1.197	1.504	1.518	1.561	1.630		
86	H	4-piperidinyl	A1	1.49		1.65	1.68	0.17	1.654	2.176	2.000	1.688	1.945		
87	H	$CH_2CH_2CONH_2$	A1	1.70		1.49	1.53	0.17	1.259	1.155	1.669	1.591	1.914		
88	H	4-$MeOC_6H_4CH_2$	A1	2.00		1.85	1.88	0.17	1.914	1.319	1.612	1.615	2.056		
89	H	3-$MeOC_6H_4CH_2$	A1	1.82		1.84	1.87	0.17	1.215	1.428	1.475	1.703	1.576		
90	H	2-$MeOC_6H_4CH_2$	A1	1.24		1.67	1.71	0.17	1.465	1.713	1.651	1.741	1.701		
91	Piperidinyl	A1	1.29	1.34679	1.35	1.39	0.17	1.319	1.297	1.575	1.673	1.484	1.69243	2.067777	
92	Pyrrolidinyl	A1	1.57		1.62	1.65	0.17	0.736	1.081	1.252	1.525	1.277			
93	Azetidinyl	A1	1.62		1.75	1.78	0.17	0.881	1.408	1.374	1.597	1.702			
94		A1	0.92		1.06	1.09	0.17	0.929	1.505	1.662	2.037	2.004			

95		A1	2.05	2.55	1.81	1.85	0.17	1.081	1.033	0.957	1.154	0.821	1.04576	1.521574
96		A1	1.85		1.92	1.96	0.17	1.202	1.925	1.870	1.780	2.304		
97		A1	2.28	3.4	2.22	1.85	0.57	1.267	1.703	1.624	1.510	1.618	2.424	2.44256
98		A1	2.06		2.33	1.96	0.57	1.210	2.149	2.036	1.769	2.155		



# components:	3
intercept	-0.2
Q_2	0.77
R_2	0.864
$Q_{2\ stderr}$	0.51
$R_{2\ stderr}$	0.40

Figure 3. Topomer CoMFA results. (A) Green contours favor steric bulk. Adding substituents that reach into these areas can improve activity. It is generally well illustrated above that increasing bulkiness in this position increase activity but to certain limit indicated by the yellow contour, (B) Contour maps at the hydrophobic tail indicate that the sub pocket is relatively narrow as the yellow contour limits the extending of the substituent, (C) Blue contours indicate favorable electropositive groups represented idealistically by NH groups, while red contours represent favorable electronegative regions (sulfonyl oxygen), (D) The linear relationship between experimental (*x*-axis) and predicted activity (*y*-axis) yielding Q2 of 0.777 and R2 of 0.864.

important features that should be taken into consideration. In addition, it facilitates picking up those compounds which include these features. The way it is used to aid interpretation is exemplified in Figure S5 of the Supplementary data. Finally, the model validation using confusion matrix and ROC analysis was done in Supplementary data (1.3 recursive partitioning).

3.1.1.1.2. Protein-based method. SuperStar[42] technique was used here. It is given in Section 1.4 in the Supplementary data. In this work, it was used to investigate the protein binding site in order to assign the position of the hotspots contributing in the favorable interactions that comply with those obtained from the ligand based MIF analysis. Actually, SuperStar (see Section 1.4 in Supplementary data for full analysis) was used together with 2D interaction diagram (Fig. 5) (generated after passing the complex through a MD production run) to highlight the main pockets of the binding site (pockets I, II and III) and the amino acids which form them (Table 9): pocket I consists of Phe415, Met414 and Gly410, pocket II consists of Pro197 and Leu384 and finally pocket III consists of Ser288, Ser556, Asn291 and Asp318. Besides, it was used mainly to solve the issue of structural water molecules in the binding site which are essential for binding. This is shown by the pharmacophore points (Figs. S6, S7 and S10 in Supplementary data) that indicate the importance of two structural water molecules that should be included in the docking study (Table 9 and Fig. 5). The first water molecule interacts with both the ligand sulfonyl oxygen and Ser556 (acts as a bridge) while the second water interacts with both NH of Gly449 and the ligand enolic oxygen and similarly acts as another bridge. Superstar analysis, as mentioned before, is given in details in the Supplementary data in Section 1.4. In addition, we provided the total 2D interaction diagrams of all the complexes of benzothiadiazine and benzothiazines in Section 1.8 in the Supplementary data to confirm the importance of the features corresponding to the hotspots and pharmacophore features.

3.1.1.2. Pharmacophore elements extraction. Pharmacophore features were extracted using molecular dynamics and energy-based pharmacophore and this is described in the ligand–protein approach below.

3.1.1.2.1. Ligand–protein complex method. Energy-optimized pharmacophores (E-pharmacophores)[27] are based on mapping of the energetic terms of the Glide XP scoring function onto atom centers. Ligand-receptor complex was refined using Glide XP. After that, the energy terms computed were mapped onto the molecule atoms to guide the pharmacophore generation. The features are energy weighed because each feature is the sum of the Glide XP energies of the atoms forming it (Fig. 6). The energies of each site are used for ranking where the most favorable sites are selected to generate the hypothesis. GLide XP was chosen to be energetically dissected as it accounts for more complex energy terms than traditional molecular mechanics or empirical scoring functions such as ChemScore.[43] These extra features (e.g., hydrophobic enclosure term and hydrophobic interactions motifs detection) are well illustrated in our case. For example, hydrophobic enclosure is detected when lipophilic ligand atoms are enclosed on opposite sides by lipophilic protein atoms, producing a large energetic reward as a consequence of the release of unfavorable water molecules that were previously present in the hydrophobic environment. Pockets I and II represent examples of this enclosure. Furthermore, Glide XP recognizes common hydrophobic interaction motifs such as π···π and π···cation interactions.[44] This was also

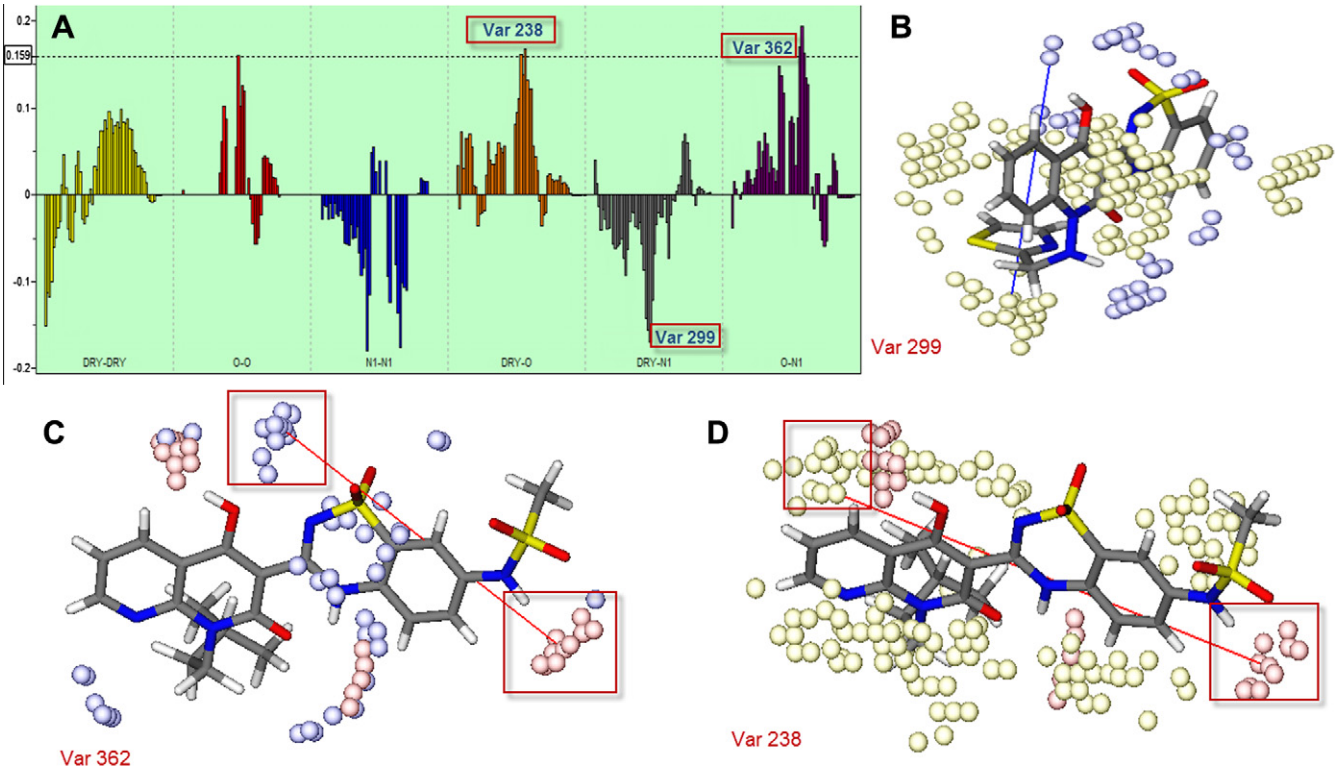


Figure 4. AMANDA MIF analysis for the benzothiadiaizines as HCV polymerase inhibitors: (A) PLS pseudo-coefficients plot showing the importance of single descriptors to explain Y-variable: positive values of a coefficient indicates direct correlation to the Y and negative ones indicate an inverse correlation to the Y. Large values mean a strong impact and low values mean a marginal impact. In this special plot, variables 238 and 362 are highlighted showing a high positive value while variable 299 shows high negative value, (B) Var 299 (Dry-N1 grind descriptor) with negative coefficient shows the negative impact of aromatic ring as hydrophobic tail on the activity, (C) Var 362 (O-N1 grind descriptor) showing importance of H-bond donor NH group of methansulfonamide and H-bond acceptor feature of the enolic group, (D) Var 238 (DRY-O grind descriptor) with positive coefficient showing importance of the hydrophobic aromatic group (ring A) and the H-bond donor of the NH of the methansulfonamide.

Table 7
Statistical results of the AMANDA model

LV	SSX	SSXacc	SDEC	SDEP	R2	R2acc	Q2acc
1	57.11	57.11	0.62	0.64	0.65	0.65	0.63
2	5.71	62.83	0.47	0.52	0.15	0.80	0.76
3	4.53	67.36	0.42	0.50	0.04	0.84	0.77
4	4.07	71.43	0.38	0.51	0.03	0.87	0.76
5	2.22	73.66	0.32	0.54	0.04	0.90	0.74

(a) SSX: percentage of the X sum of squares explained by this LV; (b) SSXacc: accumulative percentage of the X sum of squares explained by the model; (c) SDEC: standard deviation error of the calculations. It is an index of model fitting on the training set; the lower the better; (d) SDEP: standard deviation error of the predictions. It is an index of the model predictive ability obtained by cross-validation; the nearer to SDEC the better. (e) R^2: contribution of the current LV to the coefficient of determination (r^2) of the model; (f) R^2acc: coefficient of determination (r^2) of the model. It is an index of model fitting on the training set; the nearer to 1.00 (theoretical maximum) the better; (g) Q^2acc: equivalent to r^2 but obtained from cross-validation. It is an index of the model predictive power. It is obvious that the best descriptive model is the one with three latent variables as q^2 reached a ceiling level where further increase in latent variables decrease q^2.

depicted in the interaction diagram of 3HHK complex which shows the π···cation interaction between MET414 and ring A and that between PHE193 and ring D (Fig. 5).

In this study E-pharmacophore together with 2D interaction diagrams were deployed to analyze molecular dynamics simulation snapshots along production trajectory and the outcome can be summarized as following:

1. The important information extracted using this method is related to the main scaffold (constant part). This information cannot be obtained via ligand-based methods which describe difference in activity as a function of the structure variability.
2. The enolic oxygen proved to be essential feature as it forms H-bond with Tyr448 directly and indirectly via water bridge with Gly449 (Fig. 6).
3. Met414 is important in the formation of sigma–pi interaction with ring A.
4. Phe193 is important in the formation of sigma–pi interaction with ring D.

Table 8
AMANDA important grind variables which are illustrated in Figure 5

Grind variable	Fig.	Importance
Var 238 (DRY-O grind descriptor).The distance range of this variable is 14.8–15.2 A	5D	NH of the Methanesulfonamide group and hydrophobic aromatic feature (Ring A) are important hotspots. Postive PLS coefficient indicates direct correlation with activity
Var 362 (O-N1 grind descriptor) The distance range of this variable is 10.8–11.2 A	5C	NH of the methansulfonamide group and the enolic group are important features which correlate positively with activity
Var 299 (dry-N1 grind descriptor) The distance range of this variable is 12.4–12.8 A	5B	The bulky groups as a hydrophobic tail have negative impact on the activity. The effect of this variable is negative one

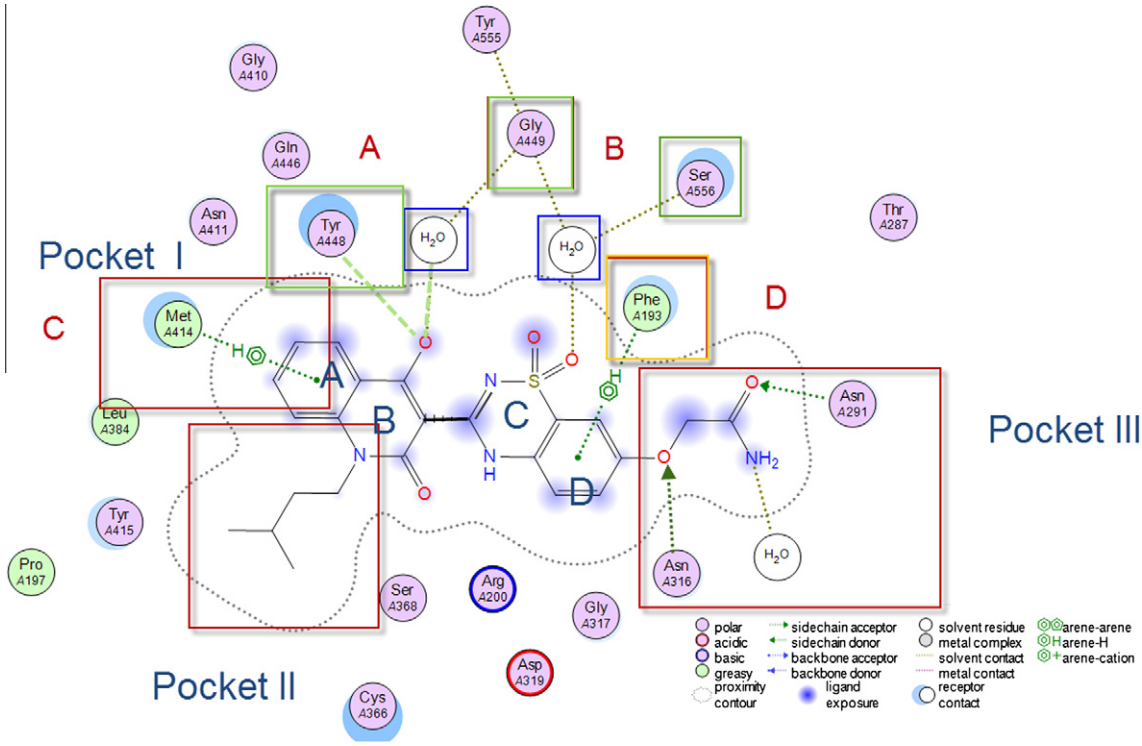


Figure 5. 2D ligand–protein interaction diagram showing detailed analysis of the binding site. The importance of water molecules in the interaction is shown: structural water acts as a bridge with Gly449 while the other acts as a bridge with Ser 556. Hydrophobic interaction of the isoamyl tail with Pro 197 and Leu 384 (greasy green residues) is obvious. Two important pi sigma interactions are depicted: Met 414 with ring A and Phe193 with ring D. Three hydrophobic pockets are red boxed. Polar amino acids are green-boxed. Structural water molecules are blue-boxed.

Table 9
Binding pockets of palm site I

Region in protein	Components	Description	Complementary feature in drug	Method of detection
Pocket I	Consists of residues Phe415, Met414, Asn411 and Gly410, close to the surface of the enzyme	A hydrophobic shallow and wide sub-pocket	The benzo moiety (ring-A) of the benzothiadiazine	E-pharmacophore ring aromatic hydrophobic feature and Superstar contours
Pocket II	Consists of residues Pro197 and Leu384	Deep and narrow hydrophobic sub-pocket	The hydrophobic tail at attached to the ring A	E-pharmacophore hydrophobic feature,Ligand-based methods and Superstar
Pocket III	Consists of Ser288, Ser556, Asn291 and Asp318.	Relatively narrow hydrophilic subpocket	The methansulfonamide at position 7 of ring D	Ligand based methods and 2D-interaction diagram
Polar amino acid residues	Tyr448,Gly449 and Ser 556	Involved in polar interactions with inhibitor	The enol moiety forms Hbond directly with Tyr 448 and indirectly via water bridge with Gly449 The sulfonyl oxygen and ser 556 are bridged by water molecule	Superstar, 2D interaction-map and E-pharmacophore.
Structural water molecules	Two water molecules Figure 5	Forms water bridge between ligand and the protein	The interaction between sulfonyl oxygen and serine 556 is bridged by a structural water molecule The oxygen atom of the enol moiety forms Hbond with a structural water molecule, which itself forms H-bond to the backbone NH of Gly449	Superstar and interaction-map

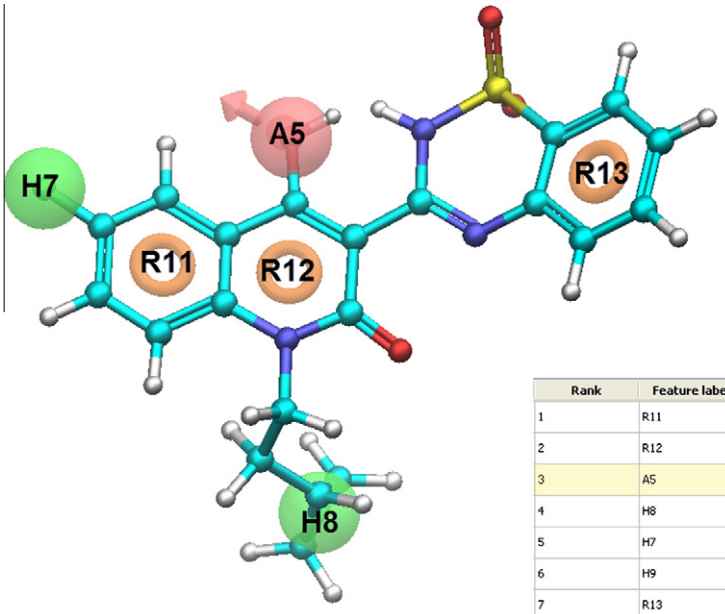
5. Sulfonyl oxygen proved to be important in the formation of H-bond with Ser556 via water bridge.

The energy-based pharmacophore method was preferred to contact-based methods[45] where accuracy can be achieved by passing the complex through a nanosecond MD production. The E-pharm was extracted along the trajectory and thus, each pharmacophore feature represents the average energy across trajectory. Molecular dynamics was performed by replicating literature results while shortening production time to 1 ns instead of 10.[20] To sum up, we summarized the results obtained in Table 9.

3.1.2. Generation of the structure-based activity prediction model

The structure-based activity prediction model is composed of two components: field template followed by guided docking. This model was validated after that and its scope was extended to be able to predict the activity of various genotypes.

3.1.2.1. Field template. Field template is a structure-based pharmacophore that is used before docking as shown in the workflow. It uses field points which describe the van der Waals and electrostatic (both positive and negative) minima and maxima that



Rank	Feature label	
1	R11	-0.83
2	R12	-0.42
3	A5	-0.27
4	H8	-0.17
5	H7	0.00
6	H9	0.00
7	R13	-0.57

Figure 6. Energy optimized pharmacophore. It is clear that hydrophobic tail and the enolic oxygen of ring B are essential features. Green spheres represent hydrophobic features; the red sphere represents hydrogen-bond acceptor feature and the orange circle represents ring aromatic feature. The table shows the energy weight of each feature.

surround molecules. The field points are generated using the extended electron distribution (XED) force field[46] which explicitly represents electron anisotropy as an expansion of point charges around each atom, whereas in classical molecular mechanics charges are placed on the centers of atoms. They are calculated by moving different probes on a grid of points placed above the Van der Waals molecular surface. As depicted in Figure 7, field points are color coded (see Supplementary data for the key in Section 1.5) and their radius represents the depth of the energy well.

Field template has two main roles in the workflow:

First, it was used to align ligands in the binding site with respect to the main bound conformer of the reference ligand such that they are placed in a good orientation and position. This is like providing initial good poses to the docking engine to speed up convergence (i.e., reaching a good solution very fast). In other words, docking

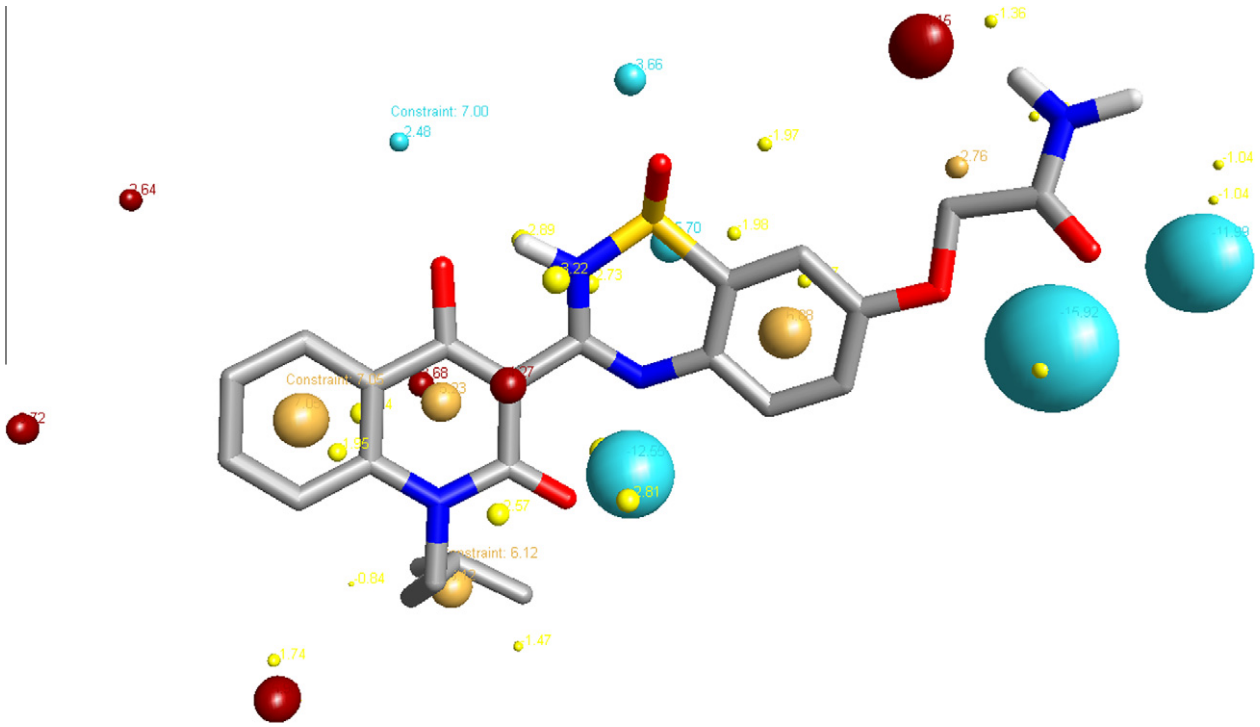


Figure 7. Field template was used as a guide for ligand placement. Small blue spheres indicate H-bond acceptor; Yellow spheres represent hydrophobic fields while red spheres indicate positive electrostatic field. Constraints were applied on some important hotspots as indicated in the figure.

will be restricted on just pose refinement and assessment. Field template was chosen as a best choice for this task and was preferred to any geometrical features structural-based pharmacophore for the following reasons:

a. Success of MIF analysis in describing the activity with good correlation suggested the suitability of using field points in alignment.
b. The fact that structure-based pharmacophore (based on geometrical features) cannot keep the RMSD between ligands to a minimum. On the other hand, the ideal tool for accurate ligand alignment, when the compounds share the same scaffold, is the field alignment. It is one of the best choices to supply aligned ligands for COMFA analysis.
c. Molecules are aligned based on their molecular fields, not on their structure. The interaction between a ligand and a protein involves electrostatic fields and surface properties (e.g., hydrogen bonding, hydrophobic surfaces and so on). Two molecules which both bind to a common active site tend to make similar interactions with the protein and hence have highly similar field properties.

Second, it was used in finding the best scoring function which correlates well with the activity.

In order to carry out these two functionalities, we had to build the field template first. This template needs the reference bound ligand to extract field points and a protein to be used as an excluded volume. The process of building the template was carried out in two steps:

a The bound conformer and the protein were prepared as following: the crystal structure of genotype 1a was used because the MIF analysis showed good correlation with the reported activities for this genotype. Previously, Homological model of Genotype 1a[18] was used in studying of benzothiadiazines to inspect binding mode. However, the first 3D structure of the subtype 1a NS5B polymerase in complex with an inhibitor (a novel sulfone-benzodiazepine derivative) was publicly disclosed (PDB ID 3HKW) recently.[28] In order to use it in this study, we had to insert the benzothiadiaizine in the palm I site of this deposited crystal. Thus, alignment was done using 3HHK (genotype 1b with benzothiadiazine ligand) as a template using 3DMA program to align the sequences of the two protein structures according to their 3D structural similarity. 3HHK bound conformer was then copied to 3HKW binding site and clashing waters were removed. It was obvious that water molecules responsible for forming water bridges (with SER556 and GLY449) were present in genotype 1a crystal. The complex was then optimized using LigX algorithm.
b The field template was generated using 3HHK bound conformer and 3HKW protein as excluded volume while using the extracted hotspots as constraints (Fig. 7).

After we generated the field template, we used it to find the suitable scoring function. The template was used to align the ligands of the dataset (98 compounds). This type of alignment takes place on a ligand level and avoids clashing with the protein using a soft potential but it does not take interactions with protein into consideration. Therefore; an optimization step was needed to refine the ligands after alignment. After that, the ligands were scored to assess ligand binding. The scoring functions used were:

1. LigScore1,[47] LigScore2,[47] PLP1,[48] PLP2,[48] Jain,[49] PMF,[50] LUDI1,[51] LUDI2[51] and LUDI 3.[51] These were evaluated after in situ ligand minimization using CHARMm (Accelrys Discovery Studio).
2. Goldscore, Chemscore ASP, CHEMPLP.[52] This was done using Gold 5 suite.
3. Moldock score[53] using Moldock optimizer in Molegro Virtual Docker.
4. Glide score[44] using Schrodinger Glide XP refinement.

Correlation matrix was generated and Moldock score proved to correlate well with activity yielding r^2 of 0.62. This result is not commonly found in literature as scoring functions are normally not measured in chemical units and are not likely to correlate with the activity. They are mainly used to achieve good enrichment or ranking in screening. Despite this fact, this is sometimes achieved in some cases.[54]

3.1.2.2. Knowledge-based docking. Guided docking is a knowledge-based strategy which is guided by both pharmacophore features and ligand–protein interaction. Moldock was used in this study as an example for this type of docking. The inputs of the docking engine can be listed as follows:

1. Pharmacophore template and this was obtained as a function of the bound conformer using the energy-based pharmacophore and molecular dynamics results. It is used to guide docking depending on the template similarity score which evaluates compliance of the docked poses with the pharmacophore template.
2. Interaction score and we choose Moldock score (variant of PLP) since it correlates well with activity.
3. Aligned ligands which emerge from the field template in a good orientation and position thus restricting docking role on refinement and assessment.

Guided docking was used here for general and specific reasons. In general, guided docking is used because it applies similarity corrections (according to the pharmacophore supplied) throughout the entire ligand incremental construction process, thus affecting not only the relative ordering of solutions but also actively guiding the ligand docking. In the end, the use of a similarity-weighted docking score aims at promoting target ligand orientations after the binding mode observed for the reference structure and penalize those diverging from the observed binding mode. The final best docking solutions should result in docking orientations of the ligand that provide a balance between maintaining favorable interactions (guided by docking score) with the receptor without deviating too much from the observed binding mode of the reference structure used (guided by template similarity score).

Guided docking was used specifically in this case because:

1. Palm I site is wide and can accept the ligand in many different orientations which differ drastically from the ideal pattern.
2. The rotatable bond joining ring B and C allows flipping of the hydrophobic tail, especially when it is bulky, from the narrow pocket II to the wider pocket I.

The algorithm used in this study for the guided docking is based on an evolutionary algorithm (EA).[26] (See Supplementary data for details concerning the algorithm and why it was used in Section 1.6). The EA used by the Moldock optimizer was guided by the pose score which consists of: Moldock score and a Template similarity score which is normalized to −500.

The uses of guided docking can be tailored here either for virtual screening or activity prediction. Virtual screening can be carried out using the combined score (pose score) which is the sum of the interaction and similarity scores. This is to obtain final best docking solutions with a balance between maintaining favorable

interactions with the receptor without deviating too much from the binding mode described by the template. On the other hand, activity can be predicted using the interaction score as it is the one which correlates with the activity. This is carried out by docking using the combined score as a guide during sampling and finally predicting the activity using the interaction score.

3.1.2.3. Validation. Validation was carried for both prediction and screening capabilities. For screening validation, ROC analysis[29,30] was done using 1000 diverse ligands from zinc database as decoys (supplied with Sybyl). It was repeated twice: once using the combined score and the other using the interaction score. The combined score proved superiority in screening (AUC = 0.914 if compared to Interaction score which have AUC = 0.748) as it guarantees that ligands which lack essential chemical features will be penalized and will have lowered overall score (see Fig. S12 in the Supplementary data).

For prediction capability validation, correlation study was done using the Moldock score and the activity in the form of $-\log(IC_{50})$ after increasing the number of iterations of the EA to 10,000. The analysis yielded r^2 of 0.8 (see Fig. S13 in the Supplementary data) which is superior to the previous correlation using energy minimization for the aligned ligands ($r^2 = 0.62$).This is a strong indication that multiple parameter optimization deployed in the EA algorithm can yield better results.

Model extrapolative capability was tested using ligands which do not belong to the benzothiadiazine class to evaluate the model extrapolative capabilities and compare it with ligand-based MIF analysis described early in the study. 1,5 Benzodiazepine was used as a test (pdb 3HKW). It acts by the same mechanism on the allosteric site (palm I)[55] and its activity is around 10 nM (i.e., $-\log(IC_{50}) = 2$). Using AMANDA as an example for ligand-based MIF analysis, the activity predicted for this compound was −0.218 using 3 latent variables. Using Moldock optimizer, on the other hand, as an example for guided docking; the activity predicted was 1.8 with RMSD of 1.5. As a result, one can conclude that the receptor-based model can be extrapolated to other series without losing predictability or pose accuracy. This proved that structure-based activity prediction models have wider application than that of ligand based ones.

3.1.2.4. Extending model applicability. In this workflow, it was important to have a valid procedure to predict the activity difference among various genotypes.[21] This is because palm I site is the least conserved among the other allosteric sites of HCV and thus will have effect on the activity of its inhibitors. For instance, the model was used to describe genotype 1a; however, in biological assay genotype 1b is used as a first choice because the activity level is generally higher for the same ligand.[21] Another example is the genotype 4 which is another variant of HCV. It shows high mutation in the binding site amino acids and is more likely to exhibit dramatic change in the activity. In addition, it is the most prevalent genotype in our country (Egypt). Generally, activity variation can be studied by modeling when using more accurate methods for binding energy prediction like FEP (free energy perturbation).[56] The FEP is the ideal method to be used in this case but it needs many nanoseconds of explicit solvation molecular dynamics production to achieve a reasonable result. Thus, we shifted to a less exhaustive computational method which is the linear interaction energy.[57] It was used to calculate binding energy of the poses retrieved from docking thus it was used as a rescoring or a post-docking scoring engine. It is a QSAR method that depends on finding the correlation between the binding energy and several energy terms which includes electrostatic and van der waals interactions (see Section 1.9 in the Supplementary data for detailed description of the method).

3.1.2.4.1. Genotype 1a/1b profiling. First, we investigated the actual correlation between the observed activities of both genotypes using a group of 61 compounds with reported activities on both genotypes[36–38] (see Tables 2, 3 and 6). The correlation between the activity of the benzothiadiazines of genotype 1a and 1b is shown in Figure S15 in the Supplementary data. It is clear from that graph that the impact of this difference is almost constant (about 0.7–1.2 log unit for $-\log(IC_{50})$ of 1b more than 1a)[21]and the activity of 1b is generally greater than 1a.This triggered us to investigate genotype 1b using Topomer CoMFA analysis but the results were not good as genotype 1a ($r^2 = 0.722$, $q^2 = 0.467$).This may explain why CoMFA studies in literature were never carried out using genotype 1b.Thus we shifted to linear interaction energy method to predict the activity difference between genotype 1a and 1b. These two genotypes have only two different amino acids (415 and 446) in the palm I allosteric site (Table 10). Genotype 1b has Glutamine 446 instead of glutamic acid (in genotype 1a) and Tyrosine 415 instead of phenylalanine (in genotype 1a). Due to this small difference in the binding site between the two genotypes, we carried out point mutations in the binding site of genotype 1a crystal (as a way of expressing genotype natural polymorphism) and applied LIA equation trained using genotype 1a on the genotype 1b. This was done in order to minimize variance between the two genotypes and decrease time required for convergence in minimization simulation employed during LIA optimization step.

Practically, this method was done using Schrodinger 2009[58] Glide for refinement of the ligand poses in the binding site of genotype 1a crystal employing XP mode. Following this, the LIE energy terms were calculated using Liasion. Finally, Strike was used to make a multiple linear regression (MLR yielding a q^2 of 0.7 using leave one out (LOO) for cross validation while neural network regression yielded q^2 of 0.75. The equation was employed using 3HKW (genotype1a). After that, the equation was applied by carrying out the two point mutations on the two amino acids 415 and 446 and the activities were predicted. It was generally observed that the difference between the activities of one ligand in the two genotypes ranges from 0.7 to 1.2 log units which were consistent with the general finding above (see Fig. S16 in the Supplementary data). Liaison results are also given in the tables of the training data set.

3.1.2.4.2. Genotype 4. Another approach was followed to study genotype 4 due to the big variation in the binding site region (Table 10). The approach is divided into two main aspects:

1. Studying the effect of each amino acid variation by calculating mutation binding-energy and detect the destabilizing effect if exists.

Table 10
Palm I allosteric site amino acids variation among different genotypes

Residue position	410	411	412	413	414	415	446	447	448	449	450	451	551	555	556
1b	G	N	I	I	M	Y	Q	I	Y	G	A	C	F	Y	S
1a_2	–	–	–	–	–	F	E	–	–	–	–	–	–	–	–
4a_26	–	–	–	–	V	–	D	M	–	–	V	T	–	A	G

Table 11
Mutation binding energy calculated for interacting amino acids of palm I site for genotype 4

Residue	Mutated to	Mutation energy(kcal/mol)	Effect
MET414	VAL	0.91	Destabilizing
GLN446	ASP	0.42	Neutral
ILE447	MET	0.07	Neutral
TYR555	ALA	−0.16	Neutral
SER556	GLY	0.11	Neutral

The destabilizing effect of MET414 is shown.

The main variations in genotype 4 (M414V, Q446D, I447M, A450V, C451T, Y555A, S556G) were studied using an approach implemented in Accelrys discovery studio which depends on calculating free energy of binding of AB complex either mutated or not. The protocol used was described in the Supplementary data in Section 1.11.

The method was applied on the amino acids which interact directly with the benzothiadiazines (Table 11). The results comply well with the literature data about the effect of mutation on the activity of the benzothiadiazines: It is obvious that MET414 plays an important role interacting with pocket I and its mutation, in natural polymorphism of genotypes, is deleterious on the activity.[21]

2. Creating a homological model using genotype 1b (3H98) as a template and applying the LIA equation derived from genotype 1a on it. The Homological model was based on the alignment of the sequence of genotype 4 with that of genotype 1b (see Fig. S17 in the Supplementary data). The template chosen was 3H98. Details are mentioned in Section 5.

The homological model of genotype 4 was aligned with genotype 1a crystal structure in order to copy the docked poses from 1a to 4. Refinement of these poses was done using Glide XP. After that, Liaison parameters were calculated and substituted in the general equation created using genotype 1a. After that, an equation was derived by applying MLR (multiple linear regressions) relating the predicted activity on genotype 4 (by applying LIA model on the homology model) and the genotype 1b reported activity. The difference in the activities can be illustrated as follows:

Predicted genotype 4 = 0.994659 $*$ $^{neg1b-1.70595}$ where neg1b is the binding energy of genotype 1b

It is obvious that the general trend observed is the decrease in activity by about 2 log units (it means 100-fold decrease in activity).This is somehow consistent with the results published.[21]

Further analysis was carried out on the effect of the mutation of genotype 4 (Met414 to valine) using molecular dynamics and radial distribution function of the sigma-pi interaction that can be formed between Met414 and ring A (see Section 1.13 in the Supplementary data).

Up till now, we have created two main components which are the detailed SAR analysis and the development of the virtual screening and activity prediction models. The integrity of the workflow and the non-redundancy of its components are discussed in details in Section 1.14 in the Supplementary data. What is left is the way that this model can be applied for new inhibitors design or discovery. This model can be used in designing new ligands by many methods:

1. Virtual screening of commercial databases (not used in this study but was validated using decoys sets).This can be used for discovery of benzothiadiaizine analogues.
2. Virtual screening of virtual libraries that are built while taking into consideration the essential structural requirements (this takes into consideration the detailed SAR analysis provided by the workflow).
3. De-novo ligand design directed by the correlating scoring function (PLP).

Herein, we will give a simple experimental example based on method-2 where we will try to find a new template that can be further exploited by SAR analysis.

3.1.3. Virtual library generation

As mentioned above, we have chosen virtual library screening method to apply model on. This approach of designing new derivatives of benzothiadiazine is based on fragment hopping. The designed ligands can be prioritized after that by evaluating their activity using the generated model. Initially, a combinatorial virtual library was built using library design module in discovery studio that is based on reaction enumeration. The reaction was selected according to a literature method in which aldehydes react with orthanilamides in the presence of sodium bisulfite at 150 °C (Fig. 8).[59]

The reactants were selected such that they are synthetically feasible and meet the essential SAR analysis. They were searched using Scifinder due to its wide scope where it includes compounds synthesized in literature and thus avoiding limitations of using the commercially available substances which cover a narrow range of synthetically feasible compounds. Initially, the aldehydes were searched such that: the N-substituted alkyl group at position 3 with respect to the aldehyde group should exist (Fig. 8B), usage of dashed bonds in Scifinder to avoid specific bond type and searching is by substructure search to get diverse solutions. The results were retrieved in the form of pdf (portable document format) file which was converted into structure data (sd) file using a chemical OCR (Clide Pro).[60]

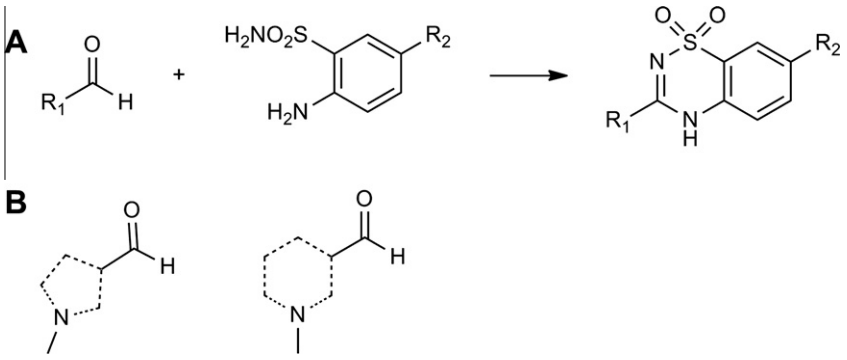


Figure 8. Reaction A is used to build a combinatorial virtual library. It is a one pot reaction between an aldehyde and an orthanilamide. The aldehyde was searched by query B using Scifinder to build the virtual library. Dashed lines are used to indicate unspecified bonds and the aldehyde group is placed at 3-position with respect to N.

Additional constraints were applied on the aldehydes retrieved by applying post-filters. The first filter applied was a one which removes aldehydes having less than two hydrogen-bond acceptors. This was based on the fact that the hydrogen bond acceptor feature which exists in the quinolinone (oxygen of the keto-enol) system is essential for interaction with Tyr448 and Gly449 residues. This was added to the second acceptor which exists by default taking into consideration that the retrieved hits are aldehydes (i.e., carbonyl group of the aldehyde). The filtered hits were further modified by substituting the alkyl group at N by isoamyl group in order unify the hydrophobic tail among the candidates.

The second filter applied was a pharmacophore filter which keeps only aldehydes which have the proper shape and orientation of hydrogen bond acceptor. This was done using Accerlys Discovery studio catalyst module. First, the quinolinone substructure was used as a template for creating the shape query and assigning the proper position to the other essential features. Based on the reaction selected to make the focused library, the benzothiadizine and quinolinone were dissected along the bond joining the ring B and C such that aldehyde group was added to the quinolinone. After that, an aldehyde custom feature was created such that it maps C and O atoms of the aldehyde where the number of hydrogen atoms suitable for substitution at C was assigned one. Finally, the query was created as shown in the Figure 9a manually such that A represents the hydrogen-bond acceptor feature, B represents the aldehyde feature, C represents the hydrophobic tail and D represents shape of the ligand. The filtered aldehydes were used to construct the virtual library by reacting them with orthanilamides that were selected from HCV literature such that they have a short polar substituent at position 7 (e.g., methansulfonamide).

3.1.4. Virtual screening

The Virtual library was screened using the model and one of the top hits (an indole based hit) was selected to be further explored through experimental validation (another hits were found but we have chosen to explore the hit which is considered in itself a new scaffold if we considered that we are rescaffolding the quinolinone). The hits retrieved and methods deployed are exemplified in Section 1.15 in the Supplementary data.

This mentioned hit was selected because of the following reasons (Fig. 10):

I. Hydroxyl group at the 4 position of the indole makes H-bond directly with Tyr448 and indirectly with both Gly449 and Asp 446 via water molecules.

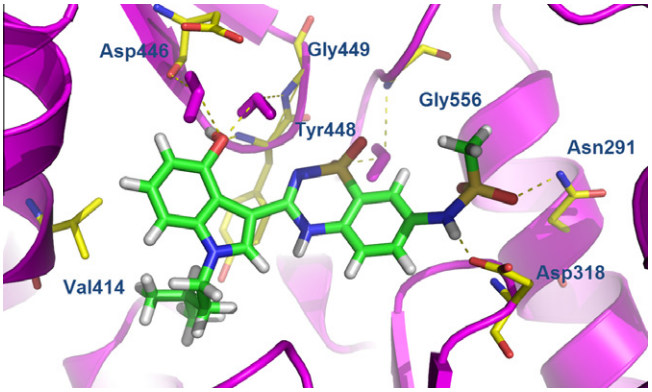


Figure 10. Docking of the hit found using virtual screening in the palm I site. Docking was performed on the homological model of genotype 4 which shows a mutation of Met414 to Val414. The methansulfonamide can form H-bond with Asn291 and Asp318.

II. Simple molecular dynamics simulation pointed out to the formation of an internal hydrogen bond between indole OH and benzothiadiazine N2. This gives some rigidity to the structure and increase propensity of the desired conformation.
III. Indole benzene ring is aromatic and can form sigma–pi interaction with Met414.
IV. The methansulfonamide group forms a network of H-bonds with Asn291 and Asp318 that can compensate loss of the activity due to Met414 mutation in genotype 4.
V. Topological polar surface area (TPSA) of the ligand is 146 which is near the desired range. This may improve the pharmacokinetic properties of this class which is known to suffer from poor bioavailability due to the large PSA.
VI. The aldehyde fragment, which formed the hit, showed perfect mapping with the pharmacophore and perfect alignment with the reference as shown in the Figure 9b.

Further assessment of this compound was specifically carried out by synthesis and biological evaluation which proved its activity (IC_{50} against polymerase enzyme was 0.3 μM).

3.1.5. Focused library generation

In order to validate the model experimentally, a small focused library of ligands was synthesized based on the hit retrieved and was further biologically evaluated. The library members are listed

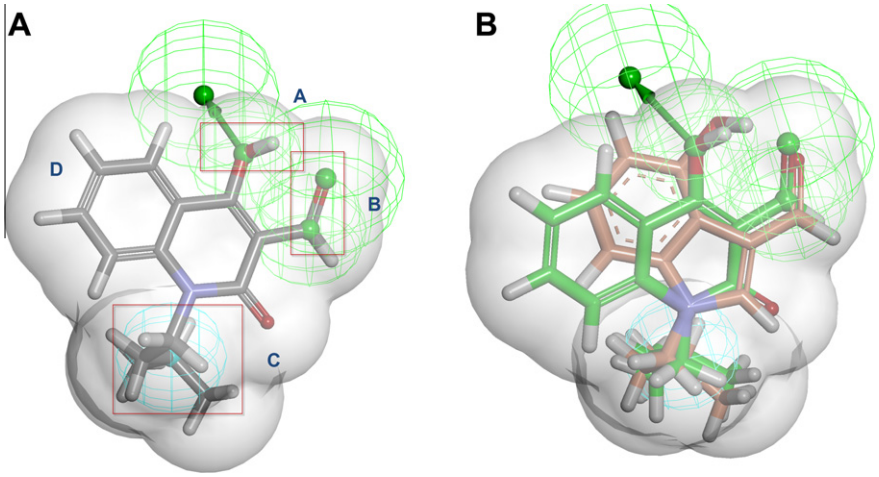


Figure 9. (A) The manual pharmacophore used to filter aldehydes which were used to construct the virtual library. Feature A represents H-bond acceptor, feature B represents aldehyde, Feature C represents hydrophobic feature which corresponds to the tail and Feature D represents the shape feature; (B) mapping of the hit with the quionlinone shows good superposition of the main features.

in Scheme 3 as compounds **III (a–o)** and Scheme 4 as compounds **IV (a–f).** The library was created while aiming at further decreasing the topological polar surface (TPSA) in order to increase the bioavailability, therefore, the hydroxyl group on the 4-position of the indole ring was omitted depending on the following theories. We will deal with classic keto-enol system of 4-hydroxyquinolin-2(1*H*)-one as a reference. The keto-enol system seems to be necessary for the formation of internal hydrogen bonding with the two nitrogen atoms of the 1,1 dioxo-4*H*-1,2,4-benzothiadiazine (Fig. S22 in Supplementary data). This was validated in literature using ab-initio technique.[61,62] The internal hydrogen bonding resulted in a 20° tilt between the two systems (hydroxyquinolinone and benzothiadiazine) (Fig. 5) that leads to the positioning of the hydrophobic tail (e.g., isoamyl tail) in the correct way necessary for interaction with the hydrophobic pocket (pocket II) amino acid residues. Despite the above facts, the ligands designed below were based on bypassing the internal hydrogen bonding depending on a Hartree–Fock conformational sampling of the ligand. It proved that the ideal conformer, which can achieve proper binding, is a local minimum that only differs by 0.9 Kcal/mol from the global minimum which can interact similarly but less efficient (due to the flipping pattern) (Fig. 11) (see Section 1.16 in the Supplementary data for complete details).

Further validation was done by embedding the synthesized compounds in a library comprising of 1000 zinc decoys and 98 active compounds used in the study. The library was screened using the combined workflow and it was found that the synthesized compounds are in the top 10% of the ranked library together with the actives.

3.2. Synthesis

The synthesis of the target compounds was carried out using a one pot reaction of orthanilamide[59] (or one of its derivatives) and the corresponding aldehyde. This reaction is efficient for easy building up of combinatorial libraries of this class of compounds. Orthanilamides were synthesized according to Schemes 1 and 4. 2-Aminobenzenesulfonamide (**I-a**) was prepared starting from the commercially available *o*-nitrobenzensulfonylchloride which was reacted with ammonia to give the amide that was further reduced by catalytic hydrogenation. Compound (**I-b**) was synthesized via direct bromination of the orthanilamide (**I-a**) using bromine/acetic acid while compound (**I-d**) was synthesized via direct iodination using NIS (*N*-iodosuccinimide). Compound (**I-c**) was prepared according to a modified procedure starting from the compound (**I-a**). In this procedure, orthanilamide was protected by fusion with urea to give 2*H*-1,2,4-benzothiadiazin-3(4*H*)-one 1,1-dioxide which was nitrated, further de-protected and reduced to give 2,5-diaminobenzenesulfonamide.This was alkylated using methansulfonyl chloride to give 2-amino-5-methanesulfonylamino-benzenesulfonamide. This pathway has many advantages to the other procedures mainly used to synthesize this important intermediate (see Supplementary data for details of other procedures of **I-c** synthesis in Section 1.17): It is very economical, yet effective modified method, it avoids drastic reagents like chlorosulfonyl isocayante, harsh conditions like Friedel–Crafts acylation at −40 °C. In addition; it avoids sulfonation reaction of *p*-nitroanline which can lead to polymerization reaction if conditions are slightly uncontrolled.[63]

2-Amino-5-bromo-pyridine-3-sulfonamide, an analogue to orthanilamide in Scheme 4, was prepared starting from 2-aminopyridine. This was carried out by brominating 2-aminopyridine to give 2-amino-5-bromopyridine which was chlorosulfonated and further reacted with ammonia to give 2-amino-5-bromo-pyridine-3-sulfonamide.

Substituted indole carbaldehydes were synthesized according Scheme 2. The reaction starts by indoles formylation via Vilsmeier–Haack reaction using DMF and phosphorus oxychloride mixture for in situ formation of the formylating agent. This was followed by N-alkylation using various alkyl halides. Substituted 5-bromo indole carbaldehydes were synthesized starting by indoles via bromination in three steps. First, indole was converted into Sodium Indoline-2-sulfonate using sodium bisulfite. Second, it was acetylated using acetic anhydride to give sodium 1-acetylindoline-2-sulfonate.Third; bromination was carried out on this protected indole followed by deprotection using sodium hydroxide. The resulting 5-bromo indole was further formylated and substituted as above.

The synthesis of the target compounds **III (a–o)** was carried out using the one pot reaction of orthanilamide (or one of its derivatives) and the corresponding aldehyde. This was done by heating the mixture in dimethylacetamide in the presence of sodium bisulfite at 150 °C. On the other hand, the target compounds **IV (a–f)** were synthesized via modified procedure which uses pressure tube in addition to the aforementioned procedure.

3.3. Biological evaluation

Biological evaluation was carried out for the synthesized compounds using replicon assay against genotype 1b (Table 12).[64] It was found that biological activities of the above mentioned

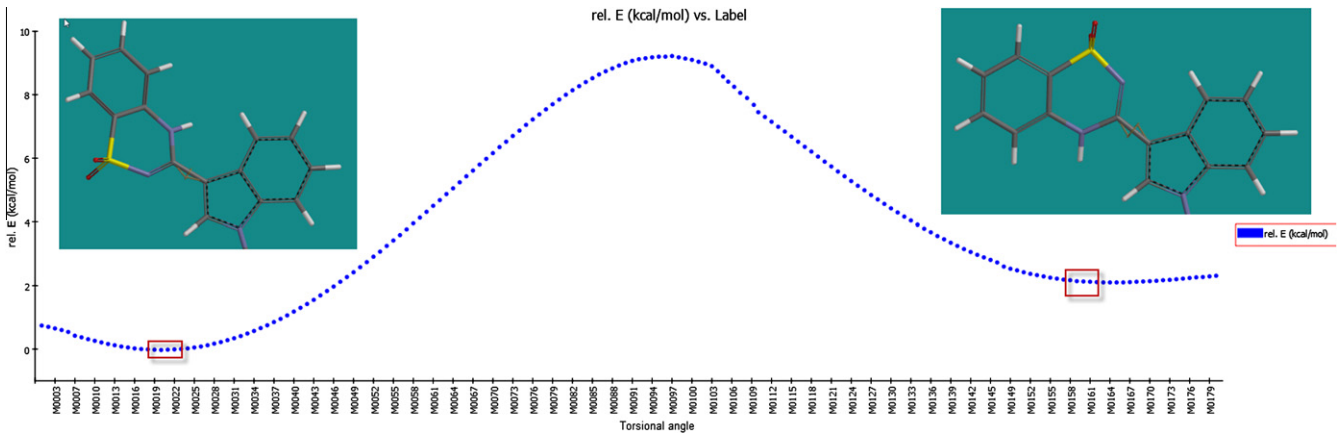
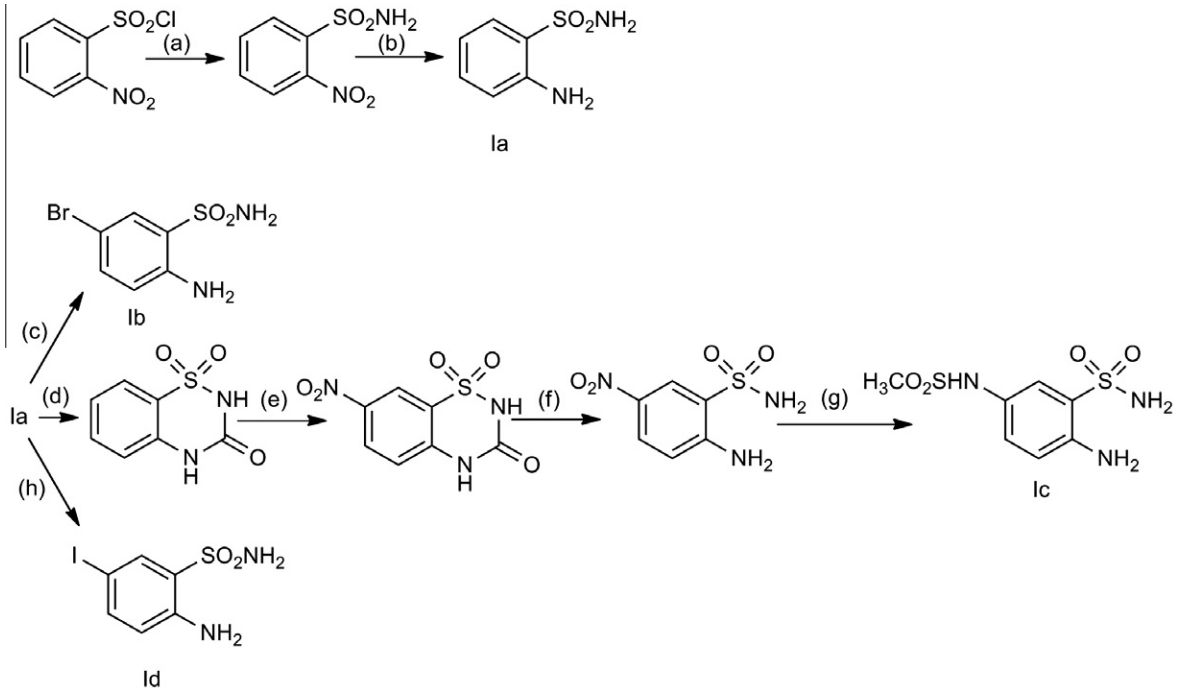
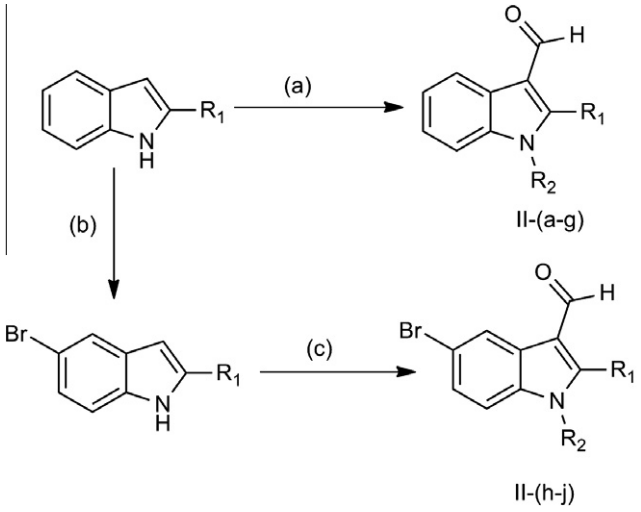


Figure 11. Conformational analysis of the indole derivative revealing local and global minima. Local minimum is at angle 159 while global minimum is at angle 22. The difference between local and global minimum is 0.9 Kcal/mol. Analysis was carried out by Spartan 2010 using Hartee-Fock 3-21G in vacuum while letting the torsional angle to span 180°.



Scheme 1. Reagents and conditions (a) NH_3; (b) Pd/C, NH_2NH_2, reflux,2 h; (c) Br_2, glacial acetic acid; (d)NH_2CONH_2,180 °C; (e) $HN0_3/H_2SO_4$, 20 h; (f) 50% H_2S0_4,130 °C; (g) (i) Pd/C, H_2, THF, 5 h, (ii) CH_3SO_2Cl, acetonitrile, pyridine, 20 h; (h) NIS, $CHCl_3$, reflux, 24 h.

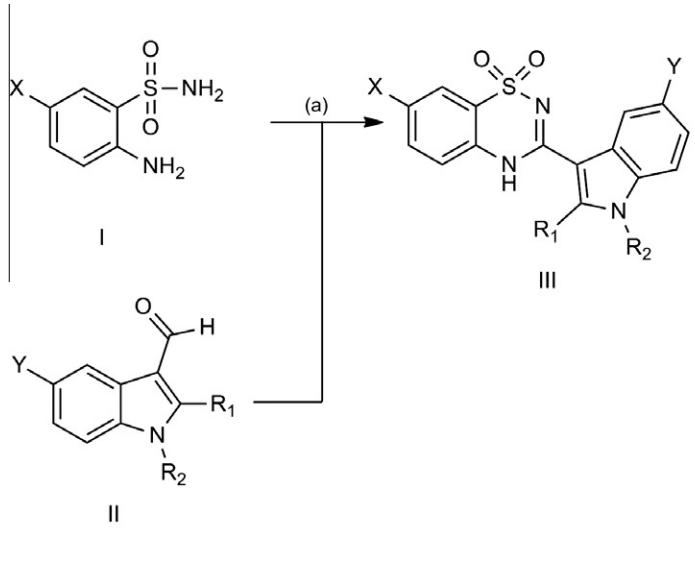


CPD	Y	R1	R2
II-a	H	H	allyl
II-b	H	H	n-propyl
II-c	H	H	isopropyl
II-d	H	H	n-butyl
II-e	H	H	benzyl
II-f	H	H	isoamyl
II-g	H	CH3	benzyl
II-h	Br	H	benzyl
II-i	Br	H	isoamyl
II-j	Br	H	butyl

Scheme 2. Reagents and conditions (a) (i) $POCl_3$, DMF, 1 h, 0 °C; (ii) alkyl halide, NaH, DMF,4 h; (b) (i) $NaHSO_3$, (ii) acetic anhydride, (iii) Br_2; (c) (i) $POCl_3$, DMF, 1 h, 0 °C; (ii) alkyl halide, NaH, DMF, 4 h.

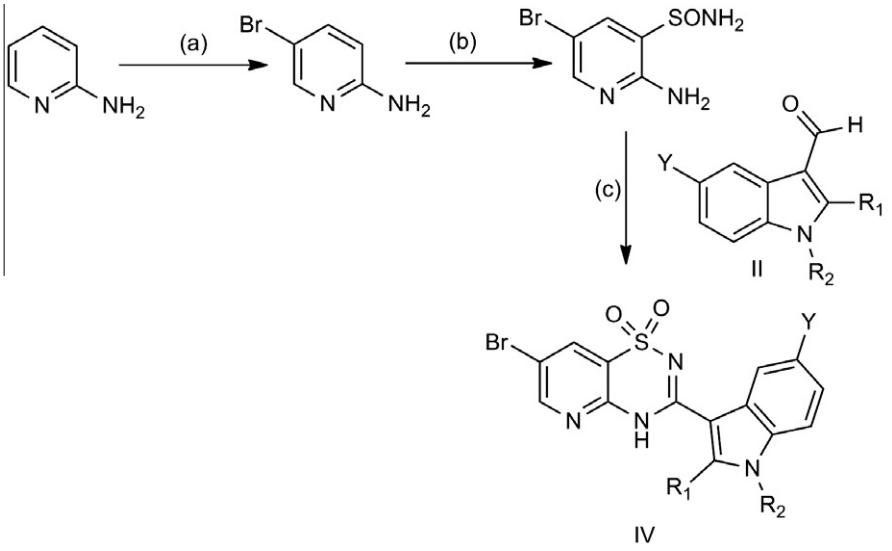
compounds comply with the model described. The most active ligands were having isoamyl substituent as a hydrophobic tail and methansulfonamide in the pocket III. Methansulfonamide capability to form extensive hydrogen bonding network with pocket III amino acids is illustrated in Figure S26 in the Supplementary data. Additionally, in this investigation we explored the effect of substituting methansulfonamide with halogen and see the effect of halogen bonding (see Figs. S27 and S28 in the Supplementary data). It was obvious from the biological data that these compounds are of equivalent potency but with higher toxicity. To confirm the activity mechanistically, ligand **III-m** was selected for polymerase enzymatic inhibition assay and its IC_{50} was found to be 1.2 μM. In our opinion, the most important conclusion is that indole proved to be a good alternative for the quionolinone system and that 4-hydroxy substituted indole was more potent due to the importance of a hydrogen bond acceptor at this position. This concluded one of the main aims of the article which is drifting away from the chemotype trap using guided structure-based models.

The assay was performed, in addition, on genotype 4 and all compounds showed almost no activity except those with methansulfonamide group that showed moderate inhibition (60%) at 10 μM. The reason can be due to the hydrogen bonding network which is formed by the methansulfonamide group that compensates the loss of the activity due to Met414 mutation. This phenomenon actually confirms the computational studies carried out on the mutation effect where the molecular dynamics (using radial distribution function on the sigma–pi interaction between



CPD	X	R1	R2	Y
III-a	H	H	allyl	H
III-b	H	H	n-propyl	H
III-c	H	H	isopropyl	H
III-d	H	H	n-butyl	H
III-e	H	H	benzyl	H
III-f	H	H	isoamyl	H
III-g	H	CH3	benzyl	H
III-h	H	H	benzyl	Br
III-i	H	H	isoamyl	Br
III-j	H	H	butyl	Br
III-k	Br	H	n-propyl	H
III-l	I	H	n-propyl	H
III-m	NHSO2CH3	H	isoamyl	H
III-n	NHSO2CH3	H	benzyl	H
III-o	NHSO2CH3	H	butyl	Br

Scheme 3. Reagents and conditions (a) $NaHSO_3$, DMA, 150 °C, 2 h.



CPD	R1	R2	Y
IV-a	H	allyl	H
IV-b	H	n-propyl	H
IV-c	H	isopropyl	H
IV-d	H	n-butyl	H
IV-e	H	benzyl	H
IV-f	H	isoamyl	H

Scheme 4. (a) Br_2, acetic acid; (b) (i) $ClSO_3H$, 160 °C, 3 h, (ii) NH_3; (C) $NaHSO_3$, DMA, 150 °C, pressure vessel.

Table 12
Biological data of various synthesized compounds using replicon assay against genotype 1b

Compound	CC_{50} (μM)	EC_{50} (μM)
III-a	>50	ND
III-b	>50	4.15
III-d	>50	8.01
III-e	>50	1.86
III-f	>50	19.7
III-i	>50	18.4
III-j	37.2	9.9
III-o	49.2	7.7
V-a	>50	25.4
V-b	16.6	2.96
V-d	21.9	5.89
V-e	>50	20.2
V-f	>50	7.07

ring A and Val414) proved the less frequent interaction (see Section 1.13 in the Supplementary data). Additionally, the mutation binding energy method which was carried out using Abagyan and Tortov algorithm proved the same concept.

Allover, one can deduce that the model was used in this study to retrieve ligands which are likely to be active because of the pharmacophoric and binding pattern similarity they share with the main active compounds.

4. Conclusion

In summary, we provided a comprehensive workflow to study benzothiadiazines as important HCV NS5b polymerase inhibitors acting on palm I allosteric site. The workflow is a multipurpose one that can be used for:

1. SAR analysis of this class using a combined approach by applying Topomer CoMFA (q^2 = 0.77), AMANDA (same q2) and recursive partitioning (ROC = 0.746).
2. Binding mode prediction using Guided docking.

3. Activity estimation using PLP scoring function which was proved to correlate well with activity (q^2 = 0.65 and after optimization of the whole protocol gave q^2 = 0.8).
4. Virtual screening enrichment by using consensus of both interaction and template similarity scores which was validated by ROC analysis (AUC = 0.91).
5. Genotypes profiling using linear interaction energy to calculate binding energy of the ligand with different genotypes and mutation binding energy calculation.
6. R-screening using validated Topomer CoMFA and general ligand screening using pentacle model or recursive partitioning.

Extensive validation was carried out for this model using both statistical and experimental methods. Statistically, cross validation was used throughout this study to validate different correlations. Experimentally, model was validated by using it as a tool for fragment-hopping. A virtual library, enumerated on reaction basis, was created such that it complies with SAR analysis. It was further screened and one of the top scoring hits (indole-based) was further investigated by considering it a template for focused library creation. Synthesis of this focused library was followed by biological evaluation which proved the activity of the indole-based derivatives (most active ligand showed activity of 1.86 μM).

Regarding synthesis, one step reaction was adopted to easy synthesize this class of compounds and a facile synthesis of the 2-amino-5-methanesulfonylamino-benzenesulfonamide was provided in details due to its importance as an intermediate in synthesis of HCV inhibitors.

In our opinion, this workflow can be used generally to study congeneric series of drugs. It uses field-alignment followed by guided docking. The protocol described provides a mean to generate a receptor-based QSAR model which relies on the field alignment in decreasing the RMSD between ligands in the binding site to the degree that permits finding a correlating scoring function. It is worth to validate this workflow on other datasets as its applicability domain is generally wider than that of the ligand-based approaches.

5. Experimental data

Molecular modeling and biological testing experimental procedures are mentioned in Sections 1.20.1 and 1.20.3 in the Supplementary data.

5.1. Synthesis

General methods. Starting materials were either commercially available or prepared as reported previously in the literature, unless otherwise noted. Solvents and reagents were used without further purification. Reactions were monitored by TLC, performed on silica gel glass plates containing F-254 indicator (Merck).Visualization on TLC was achieved by UV. Proton and carbon NMR spectra were recorded on a Bruker ARX-300. Chemical shifts (ä) are reported in ppm downfield from internal TMS standard or from solvent references. Mass spectra were recorded on API-SCIEX 2000. HRMS were recorded using an Agilent MSD-TOF (G1969A) connected to an Agilent 1100 HPLC system. Melting points were determined on a Stuart melting point apparatus (Stuart Scientific, Redhill, UK) and are uncorrected.

5.1.1. Intermediates synthesis

5.1.1.1. Compounds I (a–o). Procedures of synthesis are mentioned in Supplementary data in Section 1.20.2.1 according to Scheme I.

5.1.1.2. Compounds II (a–j). Procedures of synthesis are mentioned in Supplementary data in Section 1.20.2.2 according to Scheme II.

5.1.1.3. 2-Amino-5-bromopyridine. 28.2 g (0.3 moles) of 2-aminopyridine were dissolved in 50 ml of acetic acid. The solution is cooled to below 20° by immersion in an ice bath, and 48 g (15.4 ml, 0.3 moles) of bromine dissolved in 30 ml of acetic acid is added dropwise with vigorous stirring over a period of 1 h. Initially the temperature is maintained below 20°. After half the bromine solution has been added, it is allowed to rise to 50° to delay as long as possible the separation of the hydrobromide of 2-amino-5-bromopyridine. At 50° the hydrobromide usually begins to crystallize when about three-quarters of the bromine has been added. When addition of bromine is completed, the mixture is stirred for 1 h and is then diluted with 75 ml of water to dissolve the hydrobromide. The contents of the flask are transferred to a 500 ml beaker and are neutralized, with stirring and cooling, by the addition of 120 ml of 40% sodium hydroxide solution. 2-Amino-5-bromopyridine, contaminated with some 2-amino-3,5-dibromopyridine,was filtered and dried. The 2-amino-3,5-dibromopyridine is removed from the product by washing with three 500-ml. portions of hot petroleum ether. The yield of 2-amino-5-bromopyridine, is 32–34.7 g (62–67%). mp 134 °C (literature[65] mp 132–135°C)

5.1.1.4. 2-Amino-5-bromo-pyridine-3-sulfonamide. It was prepared according to the literature method.[66]

5.1.2. Target compounds synthesis

5.1.2.1. Compounds III (a–o). They were synthesized according to Scheme III. This was carried out using the general procedure proposed by Imai et al.[59]

General procedure: sodium hydrogen sulfite (0390 g 3.75 mmol) is added to a solution of *o*-aminobenzenesulfonamide (0.430 g, 2.5 mmol) and the aldehyde (2.5 mmol) in dimethyl acetamide (3 ml). The mixture is heated with stirring at 150 °C for 2 h and then poured into water (500 ml). A precipitate forms which is collected and dried (recrystallized is generally from ethanol or any appropriate alcohol).

*5.1.2.1.1. 3-(1-Allyl-1H-indol-3-yl)-7-bromo-4H-pyrido[2,3-e] [1,1270 2,4] thiadiazine 1,1-dioxide (**IV-a**).* Yield: 50%. ^{1}H NMR (300 MHz, DMSO) δ = 11.58 (s, 1H), 8.48 (d, J = 3.2, 1H), 8.39 (dt, J = 6.6, 2.7, 1H), 7.82 (dt, J = 7.8,2,1H), 7.76–7.20 (m, 6H), 6.21–6.00 (m,1H), 5.35–5.12 (m,2H), 4.98(dq, J = 5.5, 2.3, 1.9, 2H). ^{13}C NMR (75 MHz, DMSO) δ = 141.9, 136.1, 134.6, 133.3, 129.2, 127.3, 124.2, 122.8, 120.5, 119.8, 117.6, 114.5, 110.6, 100.2, 49.3. HRMS calcd for $C_{18}H_{15}N_3O_2S$: 337.0885; found: 337.0873. Anal. Calcd for $C_{18}H_{15}N_3O_2S$: C, 64.08; H, 4.48; N, 12.45. Found: C, 64.12; H, 4.46; N, 12.43.

*5.1.2.1.2. 3-(1-Propyl-1H-indol-3-yl)-4H-1,2,4-benzothiadiazine 1,1-dioxide (**III-b**).*Yield: 60%. ^{1}H NMR (300 MHz, DMSO) δ = 11.70 (s, 1H), 8.51 (d, J = 2.5, 1H), 8.39–8.30 (m, 1H), 7.98–7.83 (m, 3H), 7.67 (d, J = 7.6, 1H), 7.51 (dd, J = 8.8,2.4,1H), 7.32 (ddd, J = 7.2,5, 1.9,2H), 4.28 (t, J = 7.3, 2H), 1.89 (q, J = 7.5, 2H), 0.91 (t, J = 7.4,3H). ^{13}C NMR (75 MHz, DMSO) δ = 141.7, 136.7, 134.4, 127.3, 127.9, 125.2, 124.9, 122.3, 120.9, 119.4, 117.8, 110.1, 98.1, 52.0, 21.0, 11.9.HRMS calcd for $C_{18}H_{17}N_3O_2S$: 339.1041; found: 339.1053. Anal. Calcd for $C_{18}H_{17}N_3O_2S$: C, 63.70; H, 5.05; N, 12.38. Found: C, 63.73; H, 5.02; N, 12.35

*5.1.2.1.3. 3-(1-Isopropyl-1H-indol-3-yl)-4H-1,2,4-benzothiadiazine 1,1-dioxide (**III-c**).* Yield: 65%. ^{1}H NMR (300 MHz, DMSO) δ = 11.40 (s, 1H), 8.54 (d, J = 1.4, 1H), 8.35 (m, 1H), 7.95–7.62 (m, 3H), 7.54–7.23 (m, 4H),4.90 (m, 1H), 1.37 (d, J = 6.0, 6H). ^{13}C NMR (75 MHz, DMSO) δ = 142.3, 136.9, 135.8, 135.1, 133.1, 129.2, 127.1, 125.2, 124.1, 120.5, 120.2, 119.3, 116.9, 112.4, 97.8, 53.9,

18.8.HRMS calcd for $C_{18}H_{17}N_3O_2S$: 339.1041; found: 339.1032. Anal. Calcd for $C_{18}H_{17}N_3O_2S$: C, 63.70; H, 5.05; N, 12.38. Found: C, 63.73; H, 5.03; N, 12.36

*5.1.2.1.4. 3-(1-Butyl-1H-indol-3-yl)-4H-1,2,4-benzothiadiazine 1,1-dioxide (**III-d**).* Yield: 56%. ^{1}H NMR (300 MHz, DMSO) δ = 11.58 (s, 1H), 8.52 (d, *J* = 1.4, 1H), 8.36 (dt, *J* = 6.3, 1.7, 1H), 7.87–7.62 (m, 3H), 7.58–7.23(m, 4H), 4.33 (t, *J* = 7.0, 2H), 1.84 (p, *J* = 7.1, 2H), 1.32 (p, *J* = 7.4, 2H), 0.93 (td, *J* = 7.5, 1.3, 3H).^{13}C NMR (75 MHz, DMSO) δ = 142.2, 136.4, 135.6, 127.8, 127.7, 124.8, 124.8, 123.4, 121.6, 115, 118.6, 110.4, 98.1, 49.8, 28.7, 20.7, 14.0. HRMS calcd for $C_{19}H_{19}N_3O_2S$: 353.1198; found: 353.1188. Anal. Calcd for $C_{19}H_{19}N_3O_2S$: C, 64.57; H, 5.42; N, 11.89. Found: C, 64.59; H, 5.40; N, 11.88

*5.1.2.1.5. 3-(1-Benzyl-1H-indol-3-yl)-4H-1,2,4-benzothiadiazine 1,1-dioxide (**III-e**).* Yield: 45%. ^{1}H NMR (300 MHz, DMSO) δ = 11.64 (s, 1H), 8.59 (d, *J* = 3.6, 1H), 8.39 (dd, *J* = 6.0, 3.5, 1H), 7.82 (m,1H), 7.75–7.58(m,2H), 7.57–7.23 (m, 9H) 5.59 (d, *J* = 3.5, 2H). ^{13}C NMR (75 MHz, DMSO) δ = 141.9, 137.1, 136.1, 134.5, 131.4, 128.5, 127.9, 127.1, 124.2, 122.7, 120.4, 119.8, 117.6, 110.8, 100.4, 52.6. HRMS calcd for $C_{22}H_{17}N_3O_2S$: 387.1041; found: 387.1049. Anal. Calcd for $C_{22}H_{17}N_3O_2S$: C, 68.20; H, 4.42; N, 10.85. Found: C, 68.23; H, 4.40; N, 10.83.

*5.1.2.1.6. 3-[1-(3-Methylbutyl)-1H-indol-3-yl]-4H-1,2,4-benzothiadiazine 1,1-dioxide (**III-f**).* Yield: 60%. ^{1}H NMR (300 MHz, DMSO) δ = 11.58 (s, 1H), 8.51 (d, *J* = 1.4, 1H), 8.35 (dt, *J* = 6.3, 1.7, 1H), 7.98 (dd, *J* = 7.5, 1.5, 1H), 7.63–7.36 (m, 4H), 7.24–7.08 (m, 2H), 0.98 (d, *J* = 7, 6H), 1.56 (m, 2H), 1.74 (m, 1H), 4.16 (m, 2H). ^{13}C NMR (75 MHz, DMSO) δ = 143.1, 137.3 (s), 136.1, 129.1, 127.1, 126.3, 125.3, 128.9, 121.2, 118.9, 118.2, 110.4, 97.9, 48.9, 36.9, 27.1, 22.7.HRMS calcd for $C_{20}H_{21}N_3O_2S$: 367.1354 found: 367.1345. Anal. Calcd for $C_{20}H_{21}N_3O_2S$: C, 65.37; H, 5.76; N, 11.44. Found: C, 65.39; H, 5.74; N, 11.42

*5.1.2.1.7. 3-(1-Benzyl-2-methyl-1H-indol-3-yl)-4H-1,2,4-benzothiadiazine 1,1-dioxide (**III-g**).* Yield: 55%. ^{1}H NMR (300 MHz, DMSO) δ = 11.58 (s, 1H), 8.39 (dt, *J* = 6.6, 2.7, 1H), 7.80–7.2(m, 12H), 5.59 (s, 1H), 2.48 (m, 3H). ^{13}C NMR (75 MHz, DMSO) δ = 143.3, 140.9, 136.9, 136.1, 134.6, 129.0, 128.4, 128.0, 127.1, 124.2, 123.8, 120.3, 119.8, 119.4, 117.6, 103.8, 103.0, 47.9, 14.2. HRMS calcd for $C_{23}H_{19}N_3O_2S$:401.1198; found: 401.1185. Anal. Calcd for $C_{23}H_{19}N_3O_2S$: C, 68.81; H, 4.77; N, 10.47. Found: C, 68.84; H, 4.75; N, 10.46.

*5.1.2.1.8. 3-(1-Benzyl-5-bromo-1H-indol-3-yl)-4H-1,2,4-benzothiadiazine 1,1-dioxide (**III-h**).* Yield: 70%. ^{1}H NMR (300 MHz, DMSO) δ = 11.64 (s, 1H), 8.59 (d, *J* = 3.6, 1H), 8.39 (dd, *J* = 6.0, 3.5, 1H), 7.99 (dd, *J* = 7.5, 1.5, 1H), 7.84 (d, *J* = 7.5, 1H), 7.71–7.51 (m,2H), 7.40–7.08 (m,7H), 5.59 (d, *J* = 3.5, 2H). ^{13}C NMR (75 MHz, DMSO) δ = 141.8, 137.3, 136.3, 135.4, 134.6, 132.7, 129.6, 128.9, 127.8, 127.7, 124.3, 124.6, 122.7, 119.5, 117.7, 112.6, 110.9, 99.0, 52.6. HRMS calcd for $C_{22}H_{16}Br(79)N_3O_2S$: 465.0147; found: 465.0135; calcd for $C_{22}H_{16}Br(81)N_3O_2S$: 467.0126; found: 467.0135. Anal. Calcd for $C_{22}H_{16}BrN_3O_2S$: C, 56.66; H, 3.46; N, 9.01. Found: C, 56.68; H, 3.45; N, 9.00.

*5.1.2.1.9. 3-[5-Bromo-1-(3-methylbutyl)-1H-indol-3-yl]-4H-1,2,4-benzothiadiazine 1,1-dioxide (**III-i**).* Yield: 55%. ^{1}H NMR (300 MHz, DMSO) δ = 11.58 (s, 1H), 8.52 (d, *J* = 1.4, 1H), 8.36 (dt, *J* = 6.3, 1.7, 1H), 7.99 (dd, *J* = 7.4,1.5), 7.72–7.48 (m, 3H), 7.23–7.08 (m, 2H), 4.16 (m, 2H), 1.74 (m, 1H),1.56 (m, 2H), 0.98 (d, *J* = 7, 6H). ^{13}C NMR (75 MHz, DMSO) δ = 142.3, 136.5, 135.6, 134.6, 130.1, 127.1, 126.6, 124.2, 122.9, 119.8, 117.6, 111.9, 111.6, 97.4, 49.7, 36.3, 27.1, 22.7.HRMS calcd for $C_{20}H_{20}Br(79)N_3O_2S$: 445.0460; found: 445.3615,calcd for $C_{20}H_{20}Br(81)N_3O_2S$: 447.0439; found: 447.0446. Anal. Calcd for $C_{20}H_{20}BrN_3O_2S$: C, 53.82; H, 4.52; N, 9.41. Found: C, 53.85; H, 4.50; N, 9.40.

*5.1.2.1.10. 3-(5-Bromo-1-butyl-1H-indol-3-yl)-4H-1,2,4-benzothiadiazine 1,1-dioxide (**III-j**).* Yield: 60%. ^{1}H NMR (300 MHz, DMSO) δ = 11.58 (s, 1H), 8.51(d, *J* = 1.5, 1H), 8.34 (dt, *J* = 6.2, 1.6, 1H), 7.81 (dd, *J* = 7.4,1.5),7.62–7.38 (m,3H),7.13–7.09 (m,2H), 4.03 (t, *J* = 7.3, 2H), 1.79 (p, *J* = 7.1, 2H), 1.31 (m, 2H), 0.93 (t, *J* = 7.3, 3H). ^{13}C NMR (75 MHz, DMSO) δ = 143.6, 135.1, 135.4, 134.9, 139.6, 126.5, 127.1, 123.9, 123.1, 118.9, 118.1, 112.3, 112.6, 97.6, 49.8, 28.7, 20.7, 14.0. HRMS calcd for $C_{19}H_{18}Br(79)N_3O_2S$: 431.0303; found: 431.0310; calcd for $C_{19}H_{18}Br$ (81) N3O2S:433.0283 found 433.0273. Anal. Calcd for $C_{19}H_{18}BrN_3O_2S$: C, 52.78; H, 4.20; N, 9.72. Found C, 52.80; H, 4.18; N, 9.70.

*5.1.2.1.11. 7-Bromo-3-(1-propyl-1H-indol-3-yl)-4H-1,2,4-benzothiadiazine 1,1-dioxide (**III-k**).* ^{1}H NMR (300 MHz, DMSO) δ = 11.70 (s, 1H), 8.51 (d, *J* = 2.5, 1H), 8.39–8.30 (m, 1H), 7.71–7.36 (m, 5H), 7.15 (d, *J* = 7.4, 1H), 4.20 (t, *J* = 7.3, 2H), 1.93 (q, *J* = 7.3, 2H), 0.98 (t, *J* = 7.3, 3H). ^{13}C NMR (75 MHz, DMSO) δ = 141.9, 136.4, 134.7, 131.2, 129.2, 127.8, 125.1, 122.9, 121.5, 120.6, 119.2, 110.4, 98.1, 52.0, 21.0, 11.9.HRMS calcd for $C_{18}H_{16}Br(79)N_3O_2S$: 417.0147; found 417.0143, $C_{18}H_{16}Br(81)N_3O_2S$: 419.0126; found: 419.0134. Anal. Calcd for $C_{18}H_{16}BrN_3O_2S$: C, 51.68; H, 3.86; N, 10.05. Found: C, 51.69; H, 3.84; N, 10.03.

*5.1.2.1.12. 7-Iodo-3-(1-propyl-1H-indol-3-yl)-4H-1,2,4-benzothiadiazine 1,1-dioxide (**III-l**).* Yield: 60%. ^{1}H NMR (300 MHz, DMSO) δ = 11.58 (s, 1H), 8.52 (d, *J* = 1.4, 1H), 8.41 (d, *J* = 1.7,1H), 8.36 (dt, *J* = 6.3, 1.7, 1H), 7.78 (dd, *J* = 7.5,1.5,1H), 7.63–7.36 (m, 3H), 6.96 (d, *J* = 7.5,1H), 4.26 (t, *J* = 7.3,2H),1.90 (q, *J* = 7.5,2H), 0.90(td, *J* = 7.4,2.3,3H).^{13}C NMR (75 MHz, DMSO) δ = 148.7, 141.9, 134.7, 134.2, 127.8, 125.2, 124.6, 122.9, 121.2, 120.6, 110.4, 103.3, 98.1, 52.0, 21.0, 11.9. HRMS calcd for $C_{18}H_{16}IN_3O_2S$:465.0008; found: 465.0015. Anal. Calcd for $C_{18}H_{16}IN_3O_2S$: C, 46.46; H, 3.47; N, 9.03. Found: C, 46.43; H, 3.49; N, 9.05.

*5.1.2.1.13. N-{3-[1-(3-Methylbutyl)-1H-indol-3-yl]-1,1-dioxido-4H-1,2,4-benzothiadiazin-7-yl}methansulfonamide (**III-m**).* Yield: 46%. ^{1}H NMR (300 MHz, DMSO) δ = 11.55 (s, 1H), 9. 21 (s, 1H), 8.52 (d, *J* = 1.4, 1H), 8.36 (dt, *J* = 6.3, 1.7, 1H), 7.63–7.29 (m, 4H), 7.21 (d, *J* = 1.5, 1H), 7.11 (d, *J* = 7.5, 1H), 4.30 (m, 2H), 1.89–1.58 (m, 3H), 0.91(d, *J* = 6.2,6H). ^{13}C NMR (75 MHz, DMSO) δ = 141.4, 135.5, 134.7, 128.6, 127.8, 125.2, 123.3, 122.9, 120.9–120.4 (m), 113.8, 110.4, 98.1, 49.7, 42.9, 36.3, 27.1, 22.7.HRMS calcd for $C_{21}H_{24}N_4O_4S_2$: 460.1239; found:460.1224. Anal. Calcd for $C_{21}H_{24}N_4O_4S_2$: C, 54.76; H, 5.25; N, 12.16. Found: C, 54.77; H, 5.23; N, 12.14

*5.1.2.1.14. N-[3-(1-Benzyl-1H-indol-3-yl)-1,1-dioxido-4H-1,2,4-benzothiadiazin-7-yl]methansulfonamide (**III-n**).* Yield: 54%. ^{1}H NMR (300 MHz, DMSO) δ = 11.64 (s, 1H), 9.13 (s,1H), 8.59 (d, *J* = 3.6, 1H), 8.39 (dd, *J* = 6.0, 3.5, 1H), 7.69–7.55 (m,1H), 7.43–7.18 (m, 9H), 7.10(d, *J* = 7.5, 1H), 5.56 (s, 2H), 3.22 (s, 3H). ^{13}C NMR (75 MHz, DMSO) δ = 141.9, 137.1, 135.5, 134.4, 131. 3, 128.7–128.3(m), 127.9, 127.0, 123.3, 122.7, 120.6–120.2 (m), 113.8, 110.8, 100.4, 52.6, 42.9. HRMS calcd for $C_{23}H_{20}N_4O_4S_2$: 480.0926; found: 480.0943. Anal. Calcd for $C_{23}H_{20}N_4O_4S_2$: C, 57.48; H, 4.19; N, 11.66. Found: C, 57.49; H, 4.17; N, 11.65

*5.1.2.1.15. N-[3-(5-Bromo-1-butyl-1H-indol-3-yl)-1,1-dioxido-4H-1,2,4-benzothiadiazin-7-yl]methansulfonamide (**III-o**).* Yield: 60%. ^{1}H NMR (300 MHz, DMSO) δ = 11.58 (s, 1H), 9.21 (s,1H), 8.52 (d, *J* = 1.4, 1H), 8.36 (dt, *J* = 6.3, 1.7, 1H), 7.67 (d, *J* = 1.2, 2H), 7.46–7.34 (m, 1H), 7.28 (d, *J* = 1.5, 1H), 7.12 (d, *J* = 7.5, 1H), 4.31 (t, *J* = 7.7, 2H), 1.82 (p, *J* = 7.8, 2H), 1.31 (tq, *J* = 7.7, 6.5, 2H), 0.89 (t, *J* = 6.6, 3H). ^{13}C NMR (75 MHz, DMSO) δ = 141.3, 136.1, 136.4, 130.1, 128.9, 126.4, 123.9, 123.1, 121.4, 121.5, 112.9, 110.9, 111.6, 97.4, 49.8, 42.9, 28.7, 20.7, 14.0. HRMS calcd for: $C_{20}H_{21}Br(79)N_4O_4S_2$:524.0188; found:524.0174;calcd for $C_{20}H_{21}Br(81)N_4O_4S_2$:526.0167; found: 526.0175. Anal. Calcd for $C_{20}H_{21}BrN_4O_4S_2$: C, 45.72; H, 4.03; N, 10.66. Found: C, 45.74; H, 4.01; N, 10.64

5.1.2.2. Compounds IV (a–j). They were synthesized according to the following general procedure.

Sodium hydrogen sulfite (0.390 g, 3.75 mmol) is added to a solution of 2-amino-5-bromo-pyridine-3-sulfonamide (0.59 g,

2.5 mmol) and the aldehyde (2.5 mmol) in dimethyl acetamide (3 ml). The mixture is heated with stirring at 150° for 2 h and then poured into water (500 ml). A precipitate forms which is collected and dried (recrystallization is generally from ethanol or any appropriate alcohol).

5.1.2.2.1 3-(1-Allyl-1*H*-indol-3-yl)-7-bromo-4*H*-pyrido[2,3-*e*][1, 2, 4]thiadiazine 1,1-dioxide (***IV-a***).

Yield: 50%. ^{1}H NMR (300 MHz, DMSO) δ = 11.69 (s, 1H), 8.54 (s, 1H), 8.35 (m, 1H), 8.12 (m, 1H), 7.72 (m, 1H), 7.42–7.25 (m, 3H), 6.21–6.00 (m,1H), 5.35–5.12 (m,2H), 4.98 (dq, *J* = 5.5, 2.3, 1.9, 2H). ^{13}C NMR (75 MHz, DMSO) δ = 152.0, 141.7, 141.3, 140.5, 134.5, 133.3, 129.2, 127.4, 122.8, 121.2, 120.5, 118.7, 114.5, 110.6, 100.2, 49.3.HRMS calcd for $C_{17}H_{13}Br(79)N_4O_2S$: 415.9943; found: 415.9934, calcd for $C_{17}H_{13}Br(81)N_4O_2S$: 417.9922; found 417.9934. Anal. Calcd for $C_{17}H_{13}BrN_4O_2S$: C, 48.93; H, 3.14; N, 13.43. Found: C, 48.95; H, 3.12; N, 13.41.

*5.1.2.2.2. 7-Bromo-3-(1-propyl-1H-indol-3-yl)-4H-pyrido[2,3-e][1,2,4]thiadiazine 1,1-dioxide (**IV-b**).* Yield: 56%. ^{1}H NMR (300 MHz, DMSO) δ = 11.58 (s, 1H), 8.48 (s, 1H), 8.39 (m, 1H), 8.02 (m, 1H), 7.59 (m, 1H), 7.51–7.23 (m, 3H), 4.15 (t,*J* = 7.1, 2H), 1.93 (q, *J* = 7.3,2H), 0.96 (td, *J* = 7.3,2.3,3H). ^{13}C NMR (75 MHz, DMSO) δ = 151.4, 141.8, 142.3, 140.5, 134.7, 127.8, 125.2, 122.9, 121.2, 120.6, 118.7, 110.4, 98.1, 52.0, 21.0, 11.9.HRMS calcd for $C_{17}H_{15}Br(79)N_4O_2S$: 418.0099; found: 418.0080, calcd for $C_{17}H_{15}Br(81)N_4O_2S$: 420.0079;found: 420.0092. Anal. Calcd for $C_{17}H_{15}BrN_4O_2S$: C, 48.70; H, 3.61; N, 13.36; Found: C, 48.72; H, 3.60; N, 13.34.

*5.1.2.2.3. 7-Bromo-3-(1-isopropyl-1H-indol-3-yl)-4H-pyrido[2,3-e][1,2,4]thiadiazine 1,1-dioxide (**IV-c**).* Yield: 75%. ^{1}H NMR (300 MHz, DMSO) δ = 11.64 (s, 1H), 8.52 (s, 1H), 8.27 (m, 1H), 8.13 (m, 1H), 7.63 (m, 1H), 7.46–7.22 (m, 3H), 4.90 (m, 1H), 1.37 (d, *J* = 6.0, 6H). ^{13}C NMR (75 MHz, DMSO) δ = 152.0, 141.7, 141.3, 140.5, 137.3, 132.5, 128.7, 123.9, 121.2, 120.5, 118.6, 111.9, 97.8, 54.1 18.8. HRMS calcd for $C_{17}H_{15}Br(79)N_4O_2S$: 418.0099; found 418.0080, calcd for $C_{17}H_{15}Br(81)N_4O_2S$: 420.0079; found:420.0065. Anal. Calcd for $C_{17}H_{15}BrN_4O_2S$: C, 48.70; H, 3.61; N, 13.36. Found: C, 48.71; H, 3.60; N, 13.35.

*5.1.2.2.4. 7-Bromo-3-(1-butyl-1H-indol-3-yl)-4H-pyrido[2,3-e][1,2,4]thiadiazine 1,1-dioxide (**IV-d**).* Yield: 60%. ^{1}H NMR (300 MHz, DMSO) δ = 11.70 (s, 1H), 8.49 (s, 1H), 8.34 (m, 1H), 8.14 (m, 1H), 7.62 (m, 1H), 7.63–7.36 (m, 3H), 4.12 (t, *J* = 7.3, 2H), 1.78 (p, *J* = 7.1, 2H), 1.31 (p, *J* = 7.4, 2H), 0.93 (t, *J* = 7.3,3H). ^{13}C NMR (75 MHz, DMSO) δ = 152.2, 141.9, 141.5, 140.7, 134.9, 127.6, 124.9, 123.1, 122.2, 120.7, 118.8, 110.4, 98.1, 49.8, 28.7, 20.7, 14.0.HRMS calcd for $C_{18}H_{17}Br(79)N_4O_2S$: 432.0256 found: 432.0243 calcd for $C_{18}H_{17}Br(81)N_4O_2S$: 434.0235;found: 434.0225. Anal. Calcd for $C_{18}H_{17}BrN_4O_2S$: C, 49.89; H, 3.95N, 12.93. Found: C, 49.92; H, 3.92; N, 12.91.

*5.1.2.2.5. 3-(1-Benzyl-1H-indol-3-yl)-7-bromo-4H-pyrido[2,3-e][1,2,4]thiadiazine 1,1-dioxide (**IV-e**).* Yield: 66%. ^{1}H NMR (300 MHz, DMSO) δ = 11.52 (s, 1H), 8.53 (s, 1H), 8.35(m, 1H), 8.12 (m, 1H), 7.72 (m, 1H), 7.69–7.55 (m, 1H), 7.43–7.20 (m, 7H), 5.56 (s, 2H). ^{13}C NMR (75 MHz, DMSO) δ = 152.0, 141.7, 141.3, 140.5, 137.1, 134.4, 131.4, 128.5, 127.9, 127.0, 122.7, 121.2, 120.4, 118.7, 110.8, 100.4, 52.6. HRMS calcd for $C_{21}H_{15}Br(79)N_4O_2S$:: 466.0099; found: 466.0088, calcd for $C_{21}H_{15}Br(81)N_4O_2S$: 468.0079; found: 468.0098. Anal. Calcd for $C_{21}H_{15}BrN_4O_2S$: C, 53.97; H, 3.24; N, 11.99. Found: C, 53.99; H, 3.23; N, 11.97.

*5.1.2.2.6. 3-(1-Isopentyl-1H-indol-3-yl)-7-bromo-4H-pyrido[2,3-e][1,2,4]thiadiazine 1,1-dioxide (**IV-f**).* Yield: 43%. ^{1}H NMR (300 MHz, DMSO) δ = 11.69 (s, 1H), 8.50 (s, 1H), 8.35 (m, 1H), 8.21 (m, 1H), 7.65 (m, 1H), 7.42–7.22 (m, 3H), 4.21 (m, 2H), 1.72 (m, 1H), 1.54 (m, 2H),0.96 (d, *J* = 7,6H). ^{13}C NMR (75 MHz, DMSO) δ = 152.0, 141.7, 141.3, 140.5, 134.7, 127.8, 125.2, 122.9, 121.2, 120.6, 118.7, 110.4, 98.1, 49.7, 36.3, 27.1, 22.7.HRMS calcd for $C_{19}H_{19}Br(79)N_4O_2S$: 446.0412; found: 446.0408, calcd for $C_{19}H_{19}Br(81)N_4O_2S$: 448.0392; found: 448.0388. Anal. Calcd for $C_{19}H_{19}BrN_4O_2S$: C, 51.01; H, 4.28; N, 12.52. Found: C, 51.03; H, 4.26; N, 12.50.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2012.01.031.

References and notes

1. Ikegashira, K.; Oka, T.; Hirashima, S.; Noji, S.; Yamanaka, H.; Hara, Y.; Adachi, T.; Tsuruha, J.-I.; Doi, S.; Hase, Y.; Noguchi, T.; Ando, I.; Ogura, N.; Ikeda, S.; Hashimoto, H. *J. Med. Chem.* **2006**, *49*, 6950.
2. Gopalsamy, A.; Chopra, R.; Lim, K.; Ciszewski, G.; Shi, M.; Curran, K. J.; Sukits, S. F.; Svenson, K.; Bard, J.; Ellingboe, J. W.; Agarwal, A.; Krishnamurthy, G.; Howe, A. Y. M.; Orlowski, M.; Feld, B.; O'Connell, J.; Mansour, T. S. *J. Med. Chem.* **2006**, *49*, 3052.
3. Lange, C. M.; Sarrazin, C.; Zeuzem, S. *Aliment Pharmacol. Ther.* **2010**, *32*, 14.
4. Parfieniuk, A.; Jaroszewicz, J.; Flisiak, R. *World J. Gastroenterol.* **2007**, *13*, 5673.
5. Thompson, A. J.; McHutchison, J. G. *J. Viral Hepat.* **2009**, *16*, 377.
6. Schinazi, R. F.; Coats, S. J.; Bassit, L. C.; Lennerstrand, J.; Nettles, J. H.; Hurwitz, S. *J. Handb. Exp. Pharmacol.* **2009**, *189*, 25.
7. Flisiak, R.; Parfieniuk, A. *Expert Opin. Investig. Drugs* **2010**, *19*, 63.
8. Legrand-Abravanel, F.; Nicot, F.; Izopet, J. *Exp. Opin. Investig. Drugs* **2010**, *19*, 963.
9. Li, H.; Shi, S. T. *Future Med. Chem.* **2010**, *2*, 121.
10. Deore, R. R.; Chern, J. W. *Curr. Med. Chem.* **2010**, *17*, 3806.
11. Watkins, W. J.; Ray, A. S.; Chong, L. S. *Curr. Opin. Drug Discov. Devel.* **2010**, *13*, 441.
12. Talele, T. T. *Curr. Bioact. Compd.* **2008**, *4*, 86.
13. Barreca, M. L.; Iraci, N.; Manfroni, G.; Cecchetti, V. *Future Med. Chem.* **2011**, *3*, 1027.
14. Gombar, V. K. 6036349; March 14, **2000**.
15. Melagraki, G.; Afantitis, A.; Sarimveis, H.; Koutentis, P. A.; Markopoulos, J.; Igglessi-Markopoulou, O. *Bioorg. Med. Chem.* **2007**, *15*, 7237.
16. Pourbasheer, E.; Riahi, S.; Ganjali, M.; Norouzi, P. *Mol. Divers.* **2011**, *15*, 645.
17. Chen, K.-X.; Xie, H.-Y.; Li, Z.-G. *Acta Chim. Slov.* **2009**, *56*, 684.
18. Wang, X.; Yang, W.; Xu, X.; Zhang, H.; Li, Y.; Wang, Y. *Curr. Med. Chem.* **2010**, *17*, 2788.
19. Weaver, S.; Gleeson, M. P. *J. Mol. Graph Model* **2008**, *26*, 1315.
20. Li, T.; Froeyen, M.; Herdewijn, P. *J. Mol. Model.* **2010**, *16*, 49.
21. Pauwels, F.; Mostmans, W.; Quirynen, L. M.; van der Helm, L.; Boutton, C. W.; Rueff, A. S.; Cleiren, E.; Raboisson, P.; Surleraux, D.; Nyanguile, O.; Simmen, K. A. *J. Virol.* **2007**, *81*, 6909.
22. Palm, K.; Stenberg, P.; Luthman, K.; Artursson, P. *Pharm. Res.* **1997**, *14*, 568.
23. Cheeseright, T.; Mackey, M.; Rose, S.; Vinter, A. *Exp. Opin. Drug Disc.* **2007**, *2*, 131.
24. Fradera, X.; Mestres, J. *Curr. Top. Med. Chem.* **2004**, *4*, 687.
25. Cheeseright, T.; Mackey, M.; Rose, S.; Vinter, A. *J. Chem. Inf. Model.* **2006**, *46*, 665.
26. Thomsen, R.; Christensen, M. H. *J. Med. Chem.* **2006**, *49*, 3315.
27. Salam, N. K.; Nuti, R.; Sherman, W. *J. Chem. Inf. Model.* **2009**, *49*, 2356.
28. Nyanguile, O.; Devogelaere, B.; Vijgen, L.; Van den Broeck, W.; Pauwels, F.; Cummings, M. D.; De Bondt, H. L.; Vos, A. M.; Berke, J. M.; Lenz, O.; Vandercruyssen, G.; Vermeiren, K.; Mostmans, W.; Dehertogh, P.; Delouvroy, F.; Vendeville, S.; VanDyck, K.; Dockx, K.; Cleiren, E.; Raboisson, P.; Simmen, K. A.; Fanning, G. C. *J. Virol.* **2010**, *84*, 2923.
29. Kirchmair, J.; Markt, P.; Distinto, S.; Wolber, G.; Langer, T. *J. Comput. Aided Mol. Des.* **2008**, *22*, 213.
30. Triballeau, N.; Acher, F.; Brabet, I.; Pin, J. P.; Bertrand, H. O. *J. Med. Chem.* **2005**, *48*, 2534.
31. Aqvist, J.; Medina, C.; Samuelsson, J. E. *Protein Eng.* **1994**, *7*, 385.
32. Goodford, P. J. *J. Med. Chem.* **1985**, *28*, 849.
33. Duran, A.; Martinez, G. C.; Pastor, M. *J. Chem. Inf. Model.* **1813**, *2008*, 48.
34. Cramer, R. D. *J. Med. Chem.* **2003**, *46*, 374.
35. Rusinko, A.; Farmen, M. W.; Lambert, C. G.; Brown, P. L.; Young, S. S. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1017.
36. Pratt, J. K.; Donner, P.; McDaniel, K. F.; Maring, C. J.; Kati, W. M.; Mo, H.; Middleton, T.; Liu, Y.; Ng, T.; Xie, Q.; Zhang, R.; Montgomery, D.; Molla, A.; Kempf, D. J.; Kohlbrenner, W. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1577.
37. Krueger, A. C.; Madigan, D. L.; Jiang, W. W.; Kati, W. M.; Liu, D.; Liu, Y.; Maring, C. J.; Masse, S.; McDaniel, K. F.; Middleton, T.; Mo, H.; Molla, A.; Montgomery, D.; Pratt, J. K.; Rockway, T. W.; Zhang, R.; Kempf, D. J. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 3367.
38. Rockway, T. W.; Zhang, R.; Liu, D.; Betebenner, D. A.; McDaniel, K. F.; Pratt, J. K.; Beno, D.; Montgomery, D.; Jiang, W. W.; Masse, S.; Kati, W. M.; Middleton, T.; Molla, A.; Maring, C. J.; Kempf, D. J. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 3833.
39. Cruciani, G.; Ed. *Molecular Interaction Fields: Applications in Drug Discovery and ADME Prediction.; 2006, 27*: Wiley-VCH GmbH & Co. KGaA, 2006.

40. Kastenholz, M. A.; Pastor, M.; Cruciani, G.; Haaksma, E. E.; Fox, T. *J. Med. Chem.* **2000**, *43*, 3033.
41. Fontaine, F.; Pastor, M.; Sanz, F. *J. Med. Chem.* **2004**, *47*, 2805.
42. Boer, D. R.; Kroon, J.; Cole, J. C.; Smith, B.; Verdonk, M. L. *J. Mol. Biol.* **2001**, *312*, 275.
43. Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. *J. Comput. Aided Mol. Des.* **1997**, *11*, 425.
44. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. *J. Med. Chem.* **2006**, *49*, 6177.
45. Wolber, G.; Langer, T. *J. Chem. Inf. Model.* **2004**, *45*, 160.
46. Vinter, J. G. *J. Comput. Aided Mol. Des.* **1994**, *8*, 653.
47. Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94*, 8897.
48. Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. *Chem. Biol.* **1995**, *2*, 317.
49. Jain, A. N. *J. Comput. Aided Mol. Des.* **1996**, *10*, 427.
50. Muegge, I. *J. Med. Chem.* **2005**, *49*, 5895.
51. Böhm, H.-J. *J. Comput. Aided Mol. Des.* **1994**, *8*, 243.
52. Korb, O.; Stutzle, T.; Exner, T. E. *J. Chem. Inf. Model.* **2009**, *49*, 84.
53. Yang, J. M.; Chen, C. C. *Proteins* **2004**, *55*, 288.
54. Guo, J.; Hurley, M. M.; Wright, J. B.; Lushington, G. H. *J. Med. Chem.* **2004**, *47*, 5492.
55. McGowan, D.; Nyanguile, O.; Cummings, M. D.; Vendeville, S.; Vandyck, K.; Van den Broeck, W.; Boutton, C. W.; De Bondt, H.; Quirynen, L.; Amssoms, K.; Bonfanti, J. F.; Last, S.; Rombauts, K.; Tahri, A.; Hu, L.; Delouvroy, F.; Vermeiren, K.; Vandercruyssen, G.; Van der Helm, L.; Cleiren, E.; Mostmans, W.; Lory, P.; Pille, G.; Van Emelen, K.; Fanning, G.; Pauwels, F.; Lin, T. I.; Simmen, K.; Raboisson, P. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 2492.
56. Guerois, R.; Mendes, J.; Serrano, L. *Protein Science Encyclopedia*; Wiley-VCH GmbH & Co: KGaA, 2008.
57. Zhou, R. H.; Friesner, R. A.; Ghosh, A.; Rizzo, R. C.; Jorgensen, W. L.; Levy, R. M. *J. Phys. Chem. B* **2001**, *105*, 10388.
58. Owsley, D. C.; Bloomfield, J. J. *Synthesis* **1977**, 118.
59. Imai, Y.; Sato, S.; Takasawa, R.; Ueda, M. *Synthesis* **1981**, 35.
60. Valko, A. T.; Johnson, A. P. *J. Chem. Inf. Model.* **2009**, *49*, 780.
61. Hendricks, R. T.; Fell, J. B.; Blake, J. F.; Fischer, J. P.; Robinson, J. E.; Spencer, S. R.; Stengel, P. J.; Bernacki, A. L.; Leveque, V. J. P.; Pogam, S. L.; Rajyaguru, S.; Najera, I.; Josey, J. A.; Harris, J. R.; Swallow, S. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3637.
62. Hendricks, R. T.; Spencer, S. R.; Blake, J. F.; Fell, J. B.; Fischer, J. P.; Stengel, P. J.; Leveque, V. J. P.; LePogam, S.; Rajyaguru, S.; Najera, I.; Josey, J. A.; Swallow, S. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 410.
63. Poshkus, A. C. *Product R&D* **1971**, *10*, 72.
64. Paeshuyse, J.; Leyssen, P.; Mabery, E.; Boddeker, N.; Vrancken, R.; Froeyen, M.; Ansari, I. H.; Dutartre, H.; Rozenski, J.; Gil, L. H.; Letellier, C.; Lanford, R.; Canard, B.; Koenen, F.; Kerkhofs, P.; Donis, R. O.; Herdewijn, P.; Watson, J.; De Clercq, E.; Puerstinger, G.; Neyts, J. *J. Virol.* **2006**, *80*, 149.
65. Fox, B. A.; Threlfall, T. L. *Org. Synth.* **1964**, *44*, 34.
66. Johnson, J. A.; Lloyd, J.; Finlay, H.; Jiang, J.; Neels, J.; Dhondi, N. K.; Gunaga, P.; Banerjee, A.; Adisechan, A. WO2011028741A1; **2011**, 495pp.