

Original Article

COMBINATORIAL PHARMACOPHORE MODELING AND ATOM BASED 3D QSAR STUDIES OF BENZOTHIADIAZINES AS HCV-NS5B INHIBITORS

PRASANTHI POLAMREDDY^{1,2}, VINITA VISHWAKARMA^{1*}, MANOJ KUMAR MAHTO²

¹Centre for Nanoscience and Nanotechnology, International Research Centre, Sathyabama Institute of Science and Technology, Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai 600119, India, ²Excelra Knowledge Solutions Pvt Ltd, NSL-SEZ, Uppal, Hyderabad 500039, India
Email: vinitavishwakarma@sathyabama.ac.in

Received: 18 Nov 2017 Revised and Accepted: 11 Jan 2018

ABSTRACT

Objective: The objective of the current study was to elucidate the 3D pharmacophoric features of benzothiadiazine derivatives that are crucial for inhibiting Hepatitis C virus (HCV) Non-structural protein 5B (NS5B) and quantifying the features by building an atom based 3D quantitative structure-activity relationship (3D QSAR) model.

Methods: Generation of QSAR model was carried out using PHASE 3.3.

Results: A five-point pharmacophore model with two hydrogen bond acceptors, one negative ionization potential and two aromatic rings (AANRR) was found to be common among a maximum number of benzothiadiazine based NS5B inhibitors. A statistically significant 3D QSAR model was obtained from AANRR.6 which had correlation-coefficient (R^2) value of 0.924, cross-validated correlation-coefficient (Q^2) of 0.774, high Fisher ratio of 138 and low root mean square standard error (RMSE=0.29). There is another parameter, Pearson's R, its value emphasizes correlation between predicted and observed activities of the test set. For the current model, Pearson's R-value is 0.90, hence underlining the good quality of the model. The present study suggests that nitrogen atom of benzothiadiazine sulfamide ring, oxyacetamide group attached to C7 carbon of benzothiadiazine and sulfonamide oxygens are crucial for NS5B inhibitory activity. Prediction of activities of hit drugs generated in earlier research suggests that Aprepitant (Phase predicted activity: 6.9) could be a potential NS5B inhibitor.

Conclusion: This 3D QSAR model developed was statistically good and can be used to predict the activities of newly designed NS5B inhibitors and virtual screening as well. Predict the activities of newly designed NS5B inhibitors and virtual screening as well.

Keywords: Pharmacophore, QSAR, NS5B, HCV, Benzothiadiazine, Inhibitor

© 2018 The Authors. Published by Innovare Academic Sciences Pvt Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)
DOI: <http://dx.doi.org/10.22159/ijpps.2018v10i3.23734>

INTRODUCTION

Hepatitis C virus (HCV) Non-structural protein 5B (NS5B) is an RNA dependent RNA polymerase enzyme essential for viral genome replication. NS5B is an intensely pursued target for the development of Hepatitis C therapies, as it is the key enzyme in viral replication conserved among all the genotypes [1, 2]. NS5B is 591 amino acids long last 21 amino acids at C-terminal end function as membrane anchor and hence NS5B is also classified under tail anchor proteins. NS5B can be available in a full-length form and truncated form. Truncated form NS5BA55: lacks 55 amino acids extending beyond the C-terminal 21 amino acids (a portion of the active site). Truncated form NS5BA21: lacks C-terminal 21 amino acid tail [3, 4]. NS5B synthesises complementary-Ve stranded RNA from +Ve stranded RNA template. RNA replication occurs through either primer-dependent elongation or de novo initiation [5].

From the NS5B crystal structure, it was found that the structure of the enzyme resembles the shape of an encircled right hand and has 3 domains, namely thumb, finger and palm. The thumb domain and finger domain are bridged by two loops namely $\Delta 1$ loop and $\Delta 2$ loop. NS5B has four allosteric sites and one active site [6, 7]. The thumb domain has two allosteric sites, one located at the apex of $\Delta 1$ loop and other is adjacent to $\Delta 1$ loop. Third allosteric site is located in the palm domain in close proximity to active site [7, 8].

There are several classes of HCV NS5B inhibitors that target different binding sites of the enzyme and inhibit HCV replication: (i) nucleoside analogues, (ii) pyrophosphate (PPi) analogues, (iii) non-nucleoside inhibitors. Nucleoside analogues bind to the active site of the enzyme by competing with the natural nucleotide triphosphate (NTP) counterparts for incorporation. In contrast, the non-nucleoside inhibitors bind to allosteric sites and inhibit RNA synthesis at initiation stage. The pyrophosphate analogues mimic

the PPi released during nucleotidyl transfer [9]. Many compounds targeting active and allosteric sites (Direct Acting Anti-virals) are in clinical development and till date, only active site inhibitors were launched in the market. Sofosbuvir is the breakthrough drug in the HCV research with pan-genotypic activity and good safety profile and is effective against cirrhosis as well [10]. Since then many Sofosbuvir based combination therapies were launched in the market. However, still there is a need for cost-effective therapies with short treatment duration (<8 w) and should be beneficial for niche patient population resistant to earlier therapies and those with renal failure as a comorbidity [11].

In the present study, a pharmacophore modelling and 3D-QSAR analysis was carried out using currently available NS5B inhibitors binding to Palm I site of the enzyme which could facilitate the activity prediction of newly identified Palm I inhibitors, to identify novel NS5B inhibitors through virtual screening and help in optimizing identified hits. QSAR technique is widely used in ligand-based drug designing especially for lead optimization and identification of variable and potent compounds through virtual screening. QSAR quantifies the relationship between structural features of the compounds and biological activities. Herein an atom based QSAR model was built based on pharmacophore features generated using a group of benzothiadiazine derivatives binding to Palm I site of NS5B and tested under same experimental conditions. PHASE module of Schrodinger, New York was used for this analysis [12].

MATERIALS AND METHODS

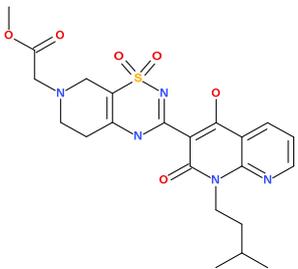
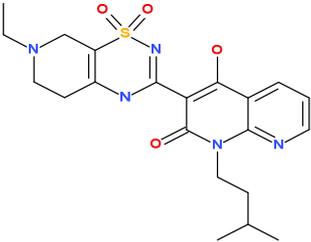
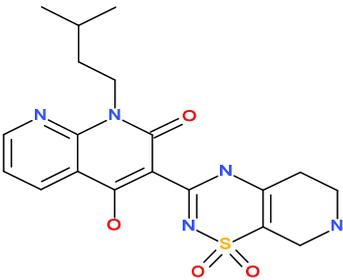
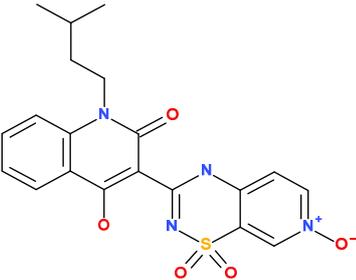
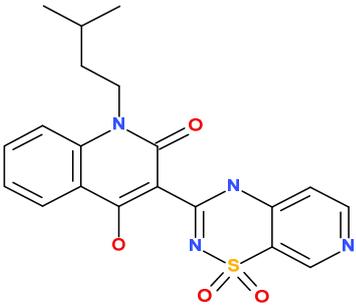
Dataset

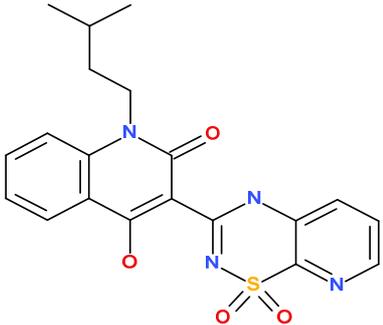
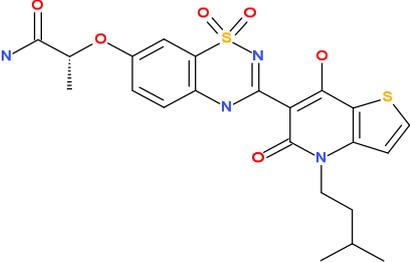
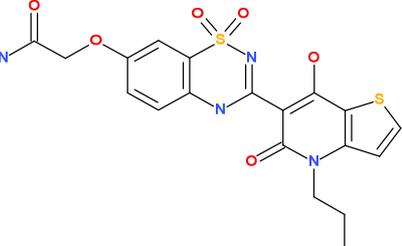
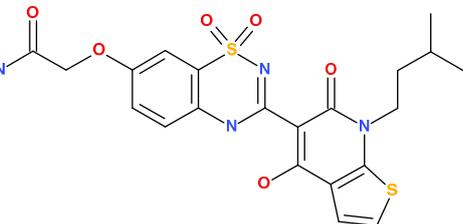
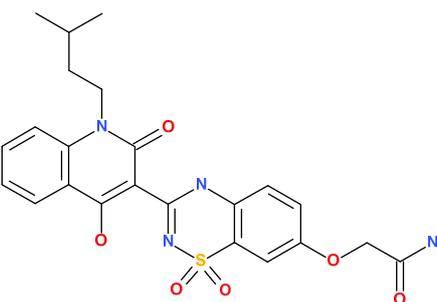
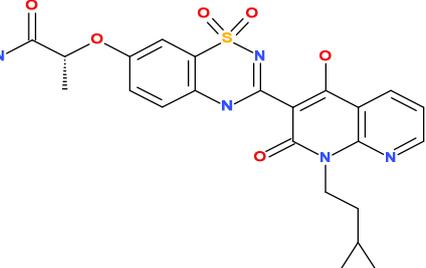
For the generation of pharmacophore based 3D-QSAR hypothesis, a set of 89 unique benzothiadiazine derivatives tested under same experimental conditions and with well-defined NS5B inhibitory

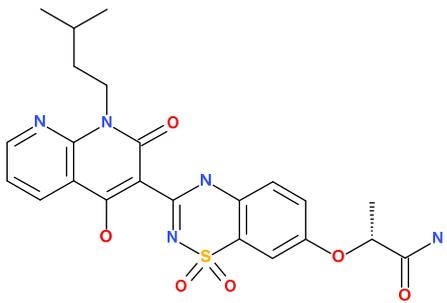
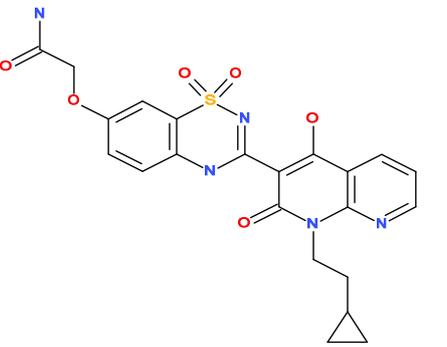
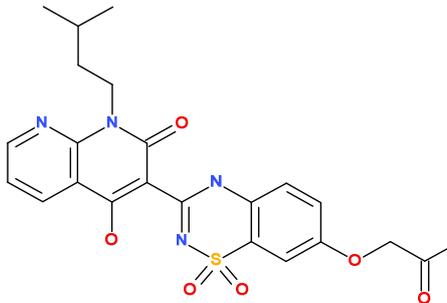
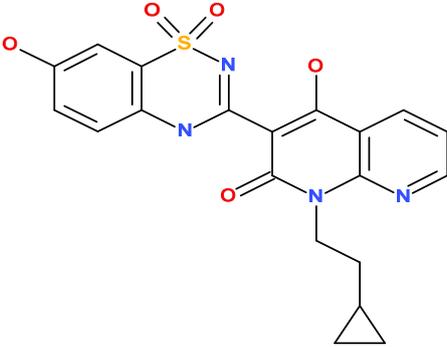
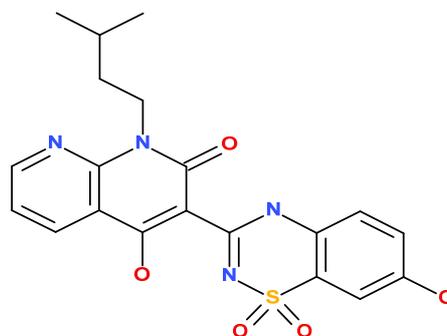
activity were selected [13-15] (table 1). All the 89 molecules were optimized and energy minimized using LigPrep module of Schrödinger. Ligand preparation involves the addition of hydrogens, adjusting bond orders, generation of stereoisomers, ionization states and tautomers, calculation of the energy of the molecule using OPLS_2005 force field, conversion of 2D to low energy 3D structure.

The prepared ligands were subjected to conformation analysis where conformers were generated using ConfGen module implemented in Schrödinger which uses Monte Carlo simulations. Conformer with the lowest energy was selected for each of the 89 molecules and loaded to Phase 3.3 model for the development of pharmacophore model.

Table 1: Dataset of benzothiadiazine derivatives used for building 3D QSAR model

S. No.	Compound NO	Compound	IC ₅₀ (nM)	pIC ₅₀
1	CMPD-1		879	6.056
2	CMPD-2		679	6.168
3	CMPD-3		604	6.218
4	CMPD-4		18	7.744
5	CMPD-5		31	7.508

6	CMPD-6		20	7.698
7	CMPD-7		< 5	8.301
8	CMPD-8		< 5	8.301
9	CMPD-9		< 5	8.301
10	CMPD-10		6	8.221
11	CMPD-11		< 5	8.301

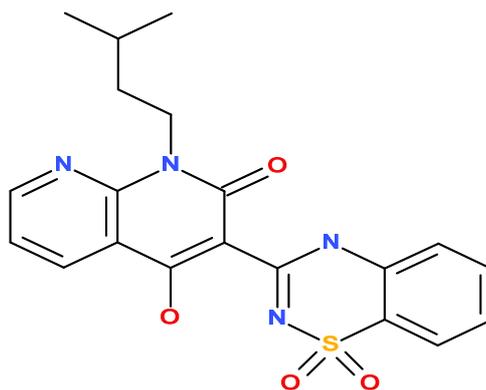
12	CMPD-12		< 5	8.301
13	CMPD-13		< 5	8.301
14	CMPD-14		< 5	8.301
15	CMPD-15		12	7.920
16	CMPD-16		16	7.795

17

CMPD-17

212

6.673

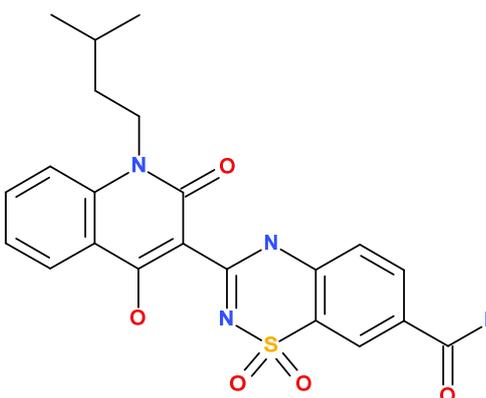


18

CMPD-18

6

8.221

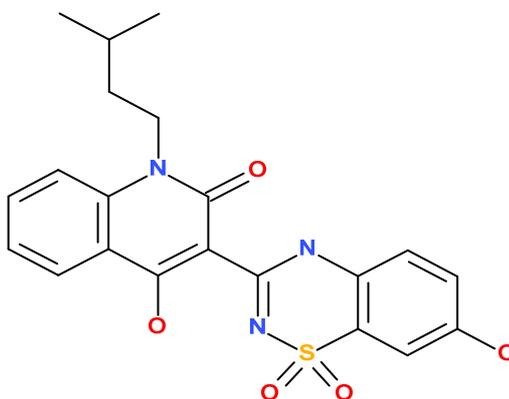


19

CMPD-19

30

7.522

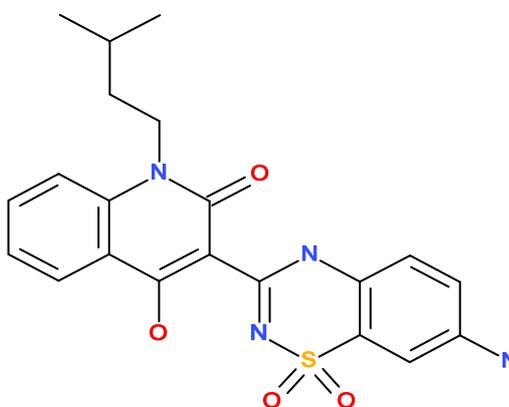


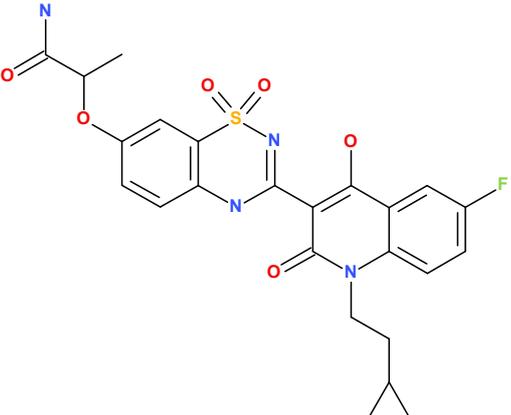
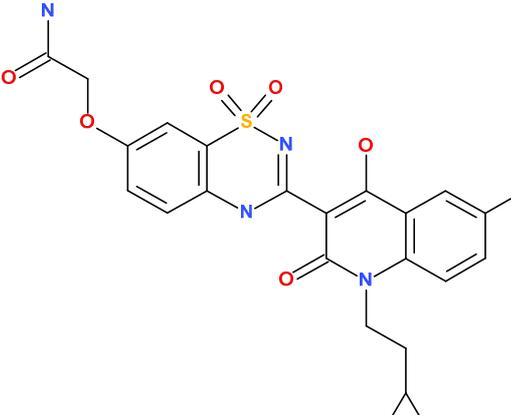
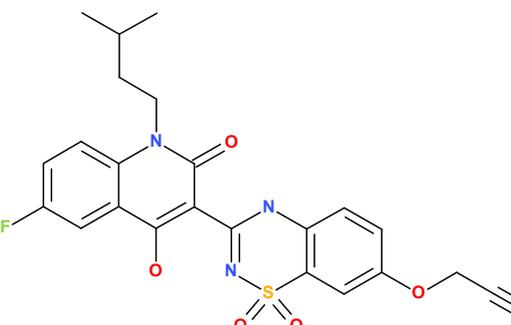
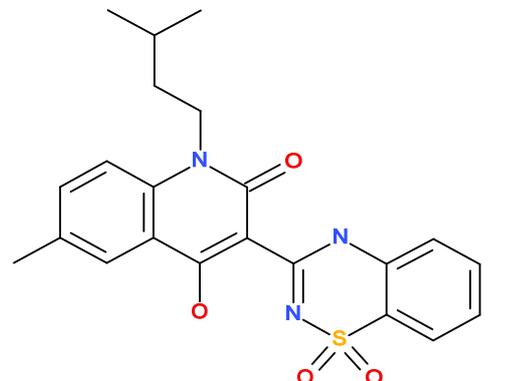
20

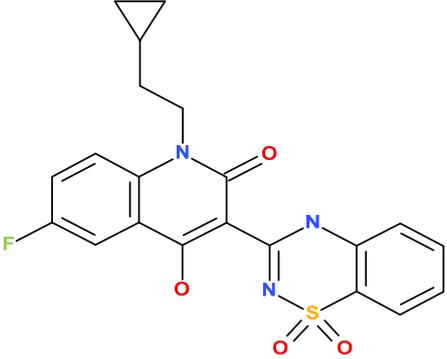
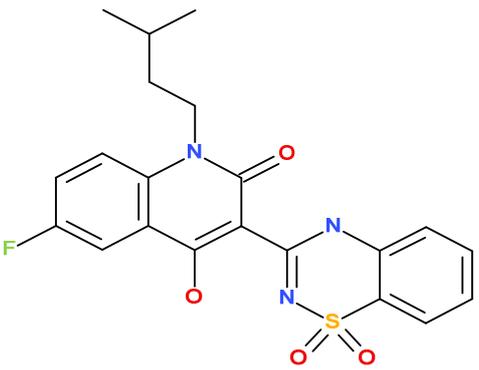
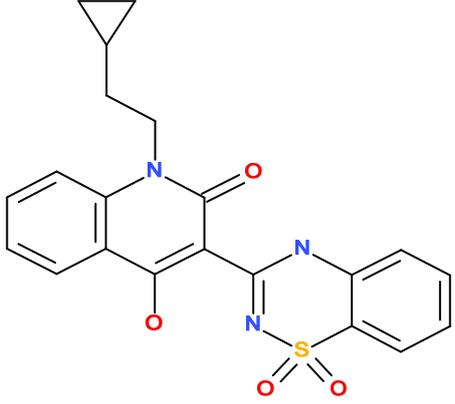
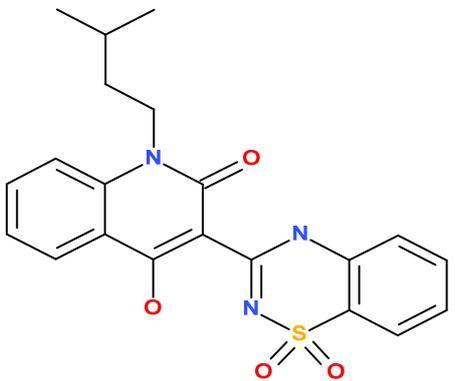
CMPD-20

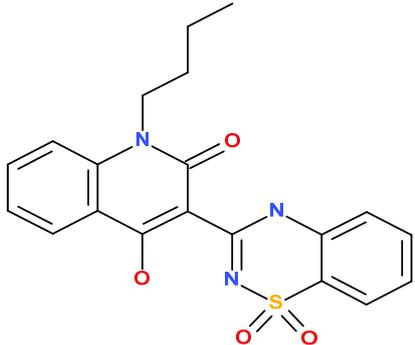
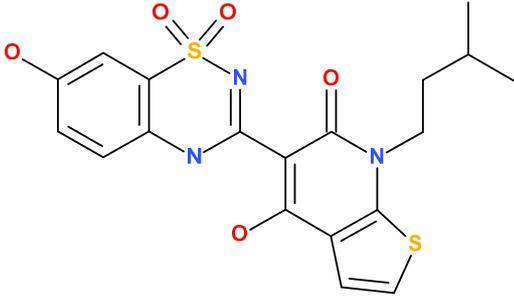
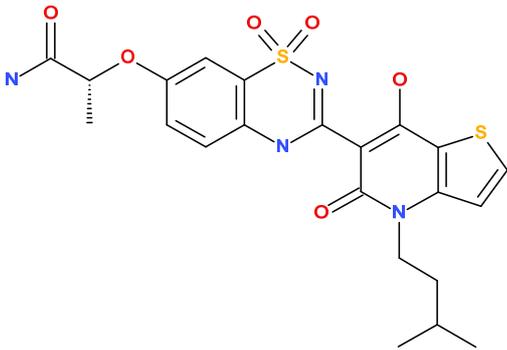
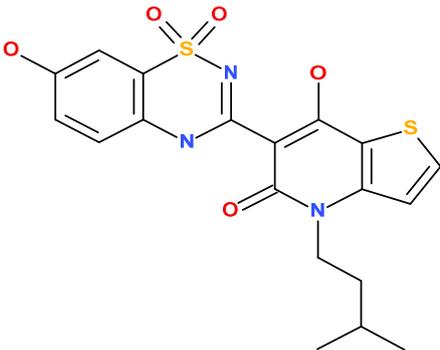
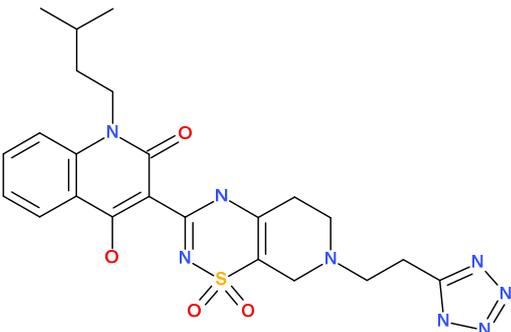
13

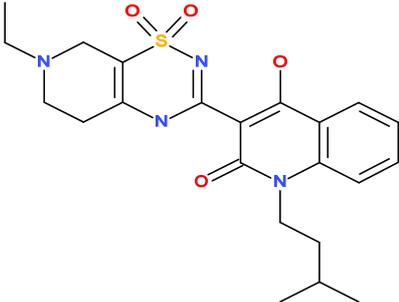
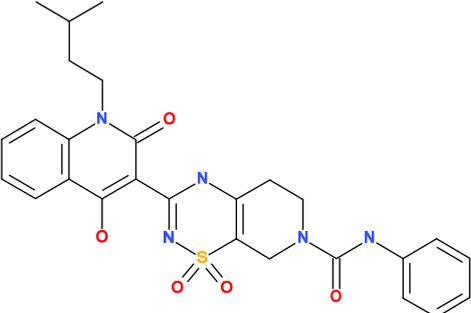
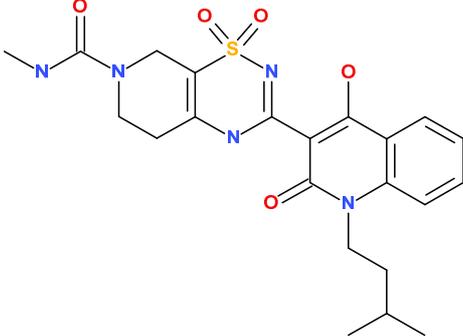
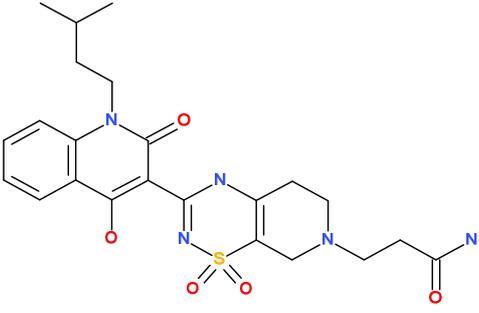
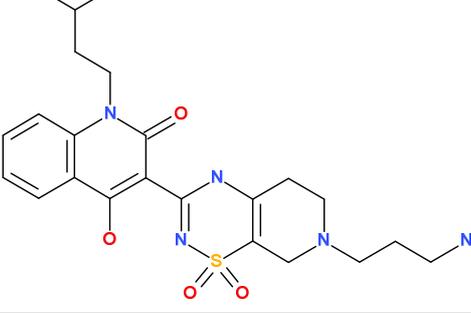
7.886

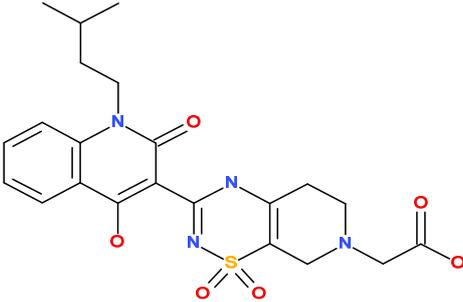
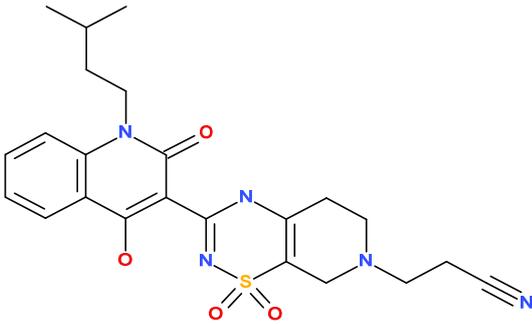
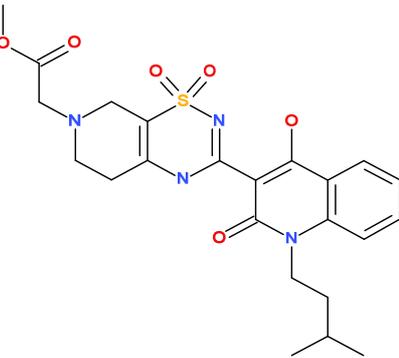
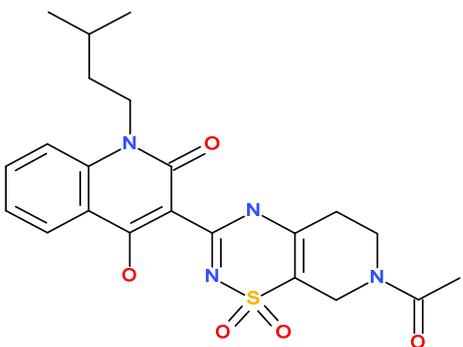
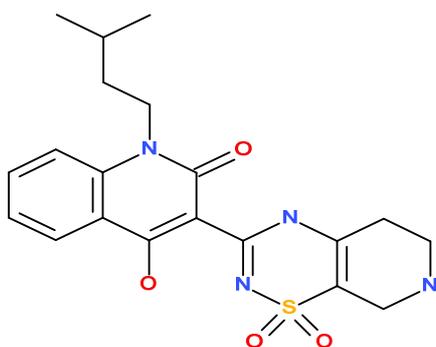


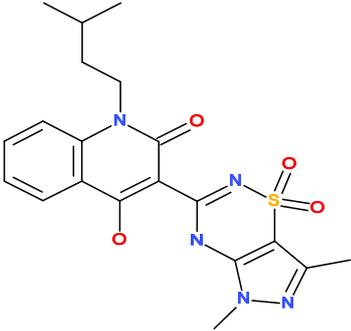
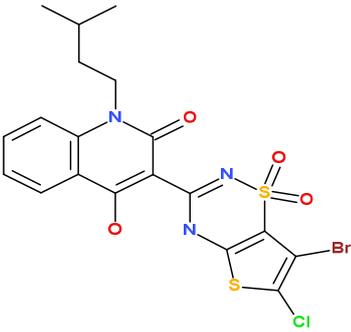
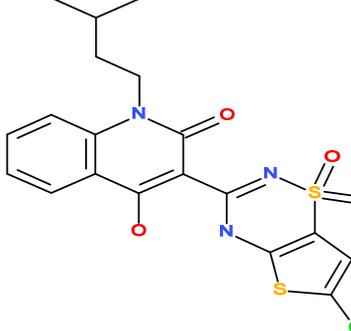
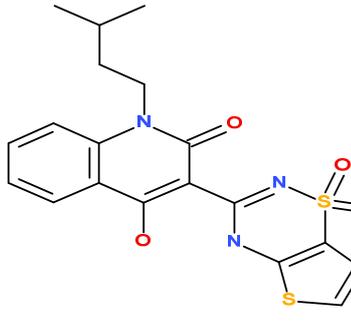
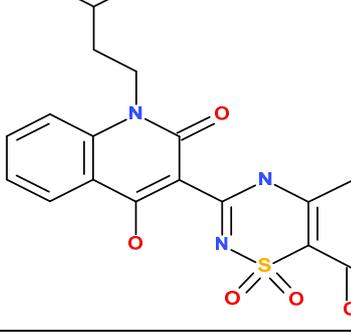
21	CMPD-21		< 5	8.301
22	CMPD-22		< 5	8.301
23	CMPD-23		7	8.154
24	CMPD-24		47	7.327

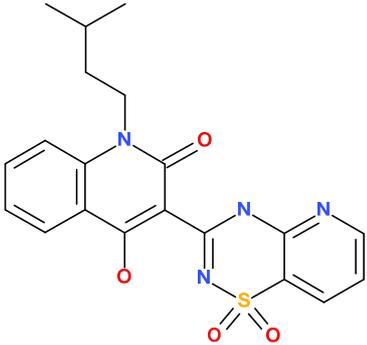
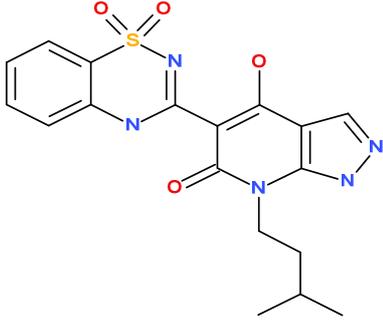
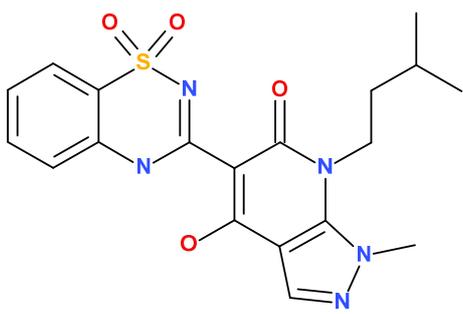
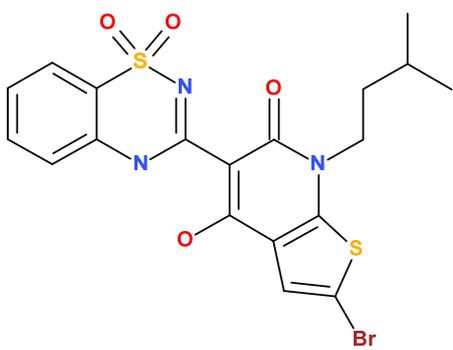
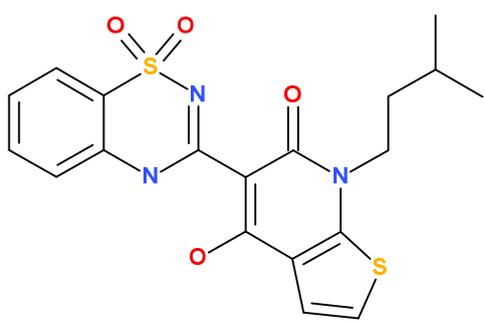
25	CMPD-25		10	8.000
26	CMPD-26		20	7.698
27	CMPD-27		13	7.886
28	CMPD-28		32	7.494

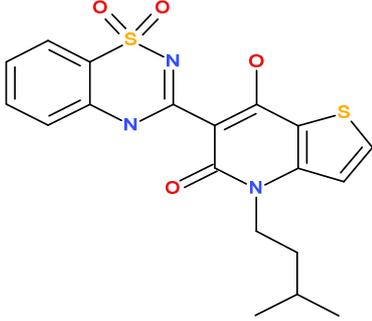
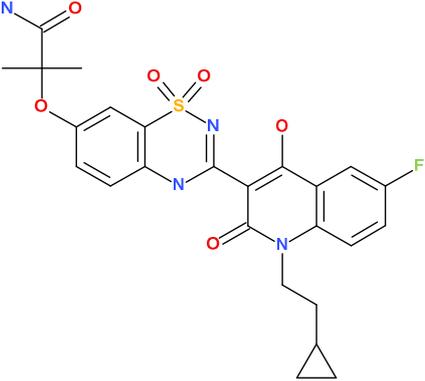
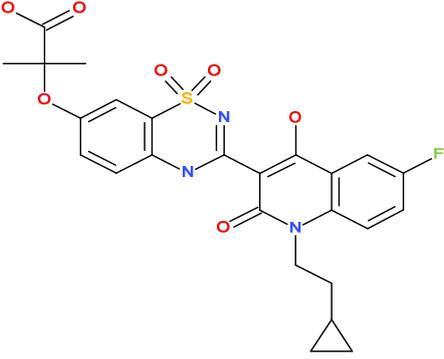
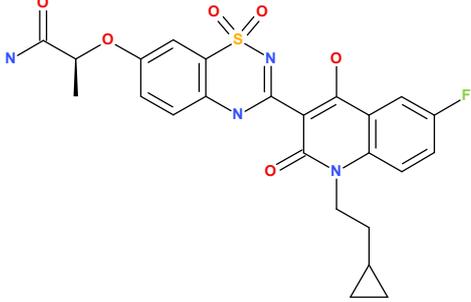
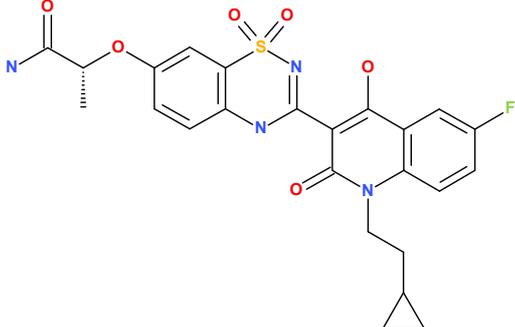
29	CMPD-29	200	6.698
			
30	CMPD-30	25	7.602
			
31	CMPD-31	< 5	8.301
			
32	CMPD-32	16	7.795
			
33	CMPD-33	289	6.539
			

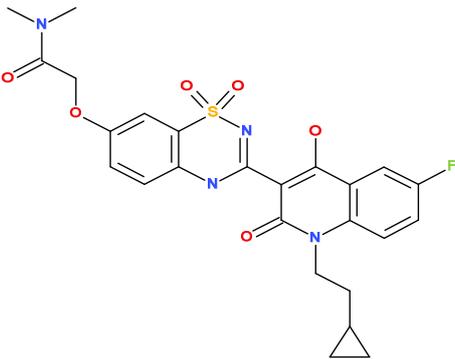
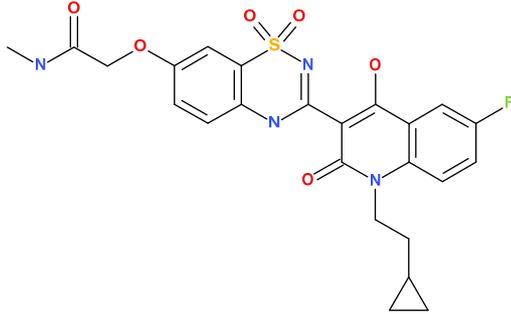
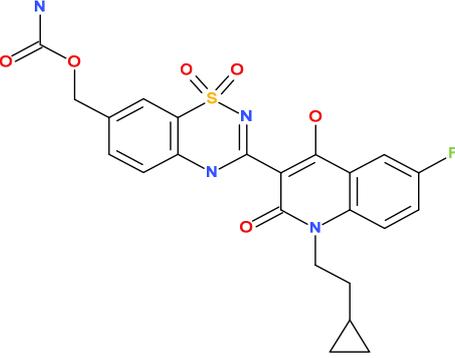
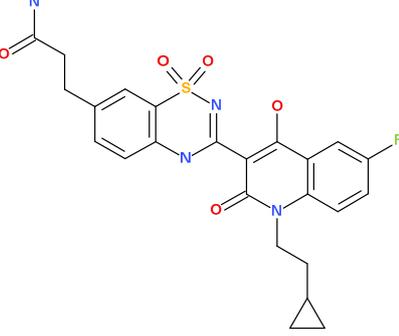
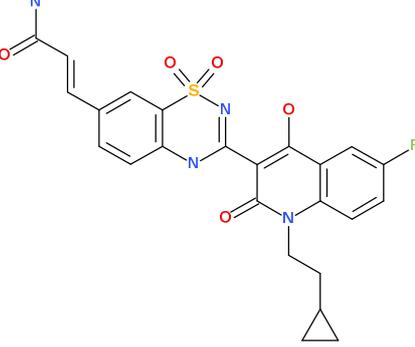
34	CMPD-34		679	6.168
35	CMPD-35		975	6.010
36	CMPD-36		774	6.111
37	CMPD-37		288	6.540
38	CMPD-38		728	6.137

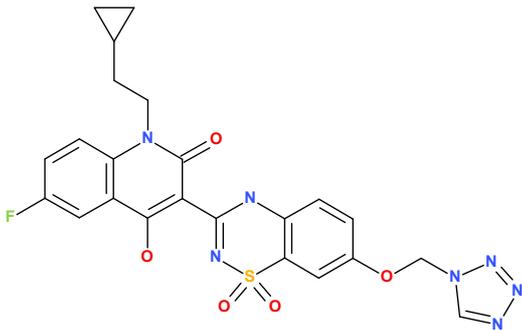
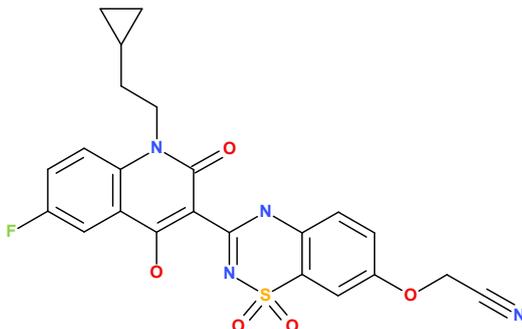
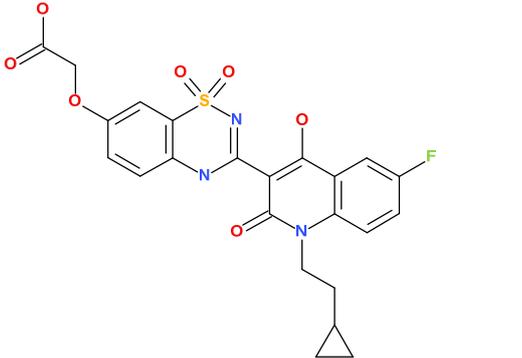
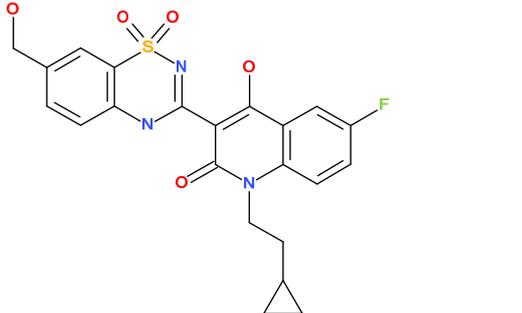
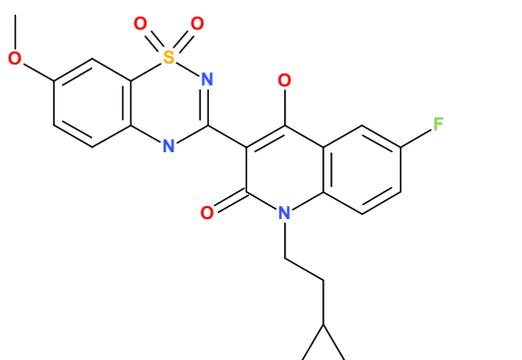
39	CMPD-39		276	6.559
40	CMPD-40		132	6.879
41	CMPD-41		879	6.056
42	CMPD-42		1527	5.816
43	CMPD-43		604	6.218

44	CMPD-44		200	6.698
45	CMPD-45		3016	5.520
46	CMPD-46		57	7.244
47	CMPD-47		20	7.698
48	CMPD-48		16	7.795

49	CMPD-49		51	7.292
50	CMPD-50		4936	5.306
51	CMPD-51		3337	5.476
52	CMPD-52		1168	5.932
53	CMPD-53		132	6.879

54	CMPD-54		48	7.318
55	CMPD-55		9	8.045
56	CMPD-56		22	7.657
57	CMPD-57		12	7.920
58	CMPD-58		< 5	8.301

59	CMPD-59		12	7.920
60	CMPD-60		< 5	8.301
61	CMPD-61		36	7.443
62	CMPD-62		5	8.301
63	CMPD-63		202	6.694

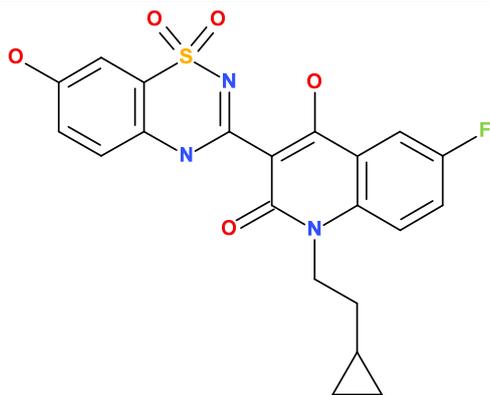
64	CMPD-64		6	8.221
65	CMPD-65		< 5	8.301
66	CMPD-66		< 5	8.301
67	CMPD-67		17	7.769
68	CMPD-68		24	7.619

69

CMPD-69

< 5

8.301

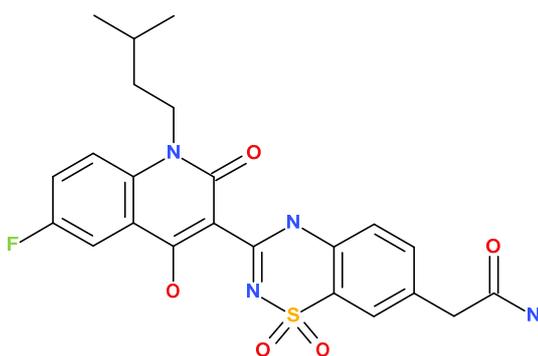


70

CMPD-70

18

7.744

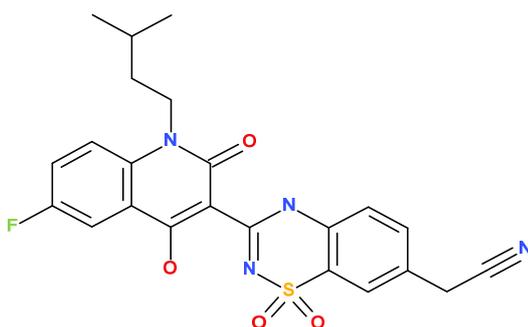


71

CMPD-71

5

8.301

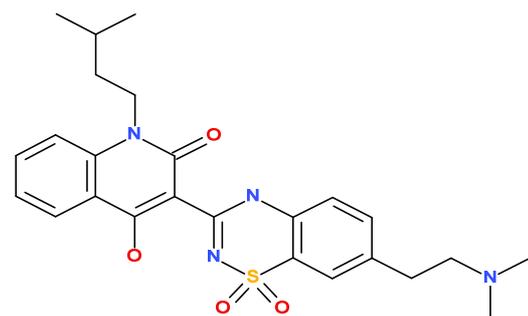


72

CMPD-72

685

6.164

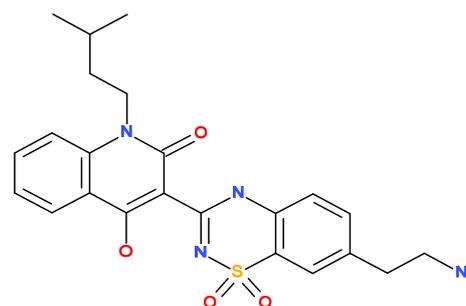


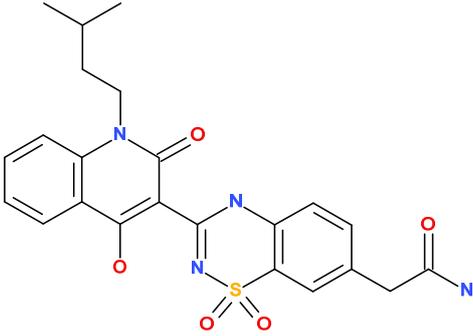
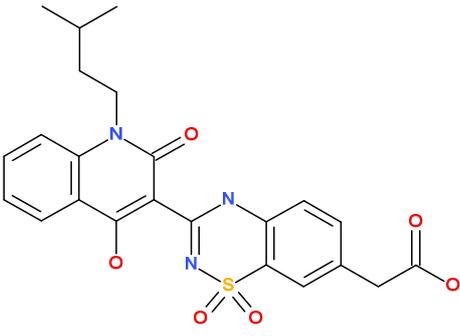
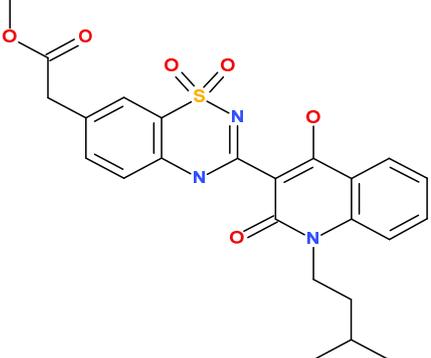
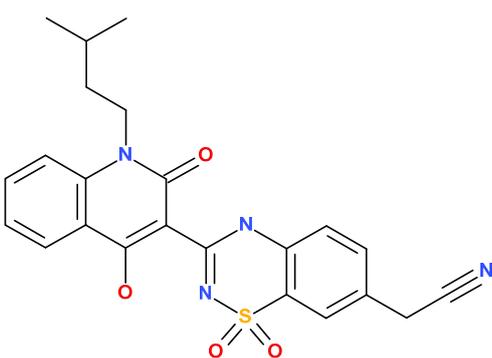
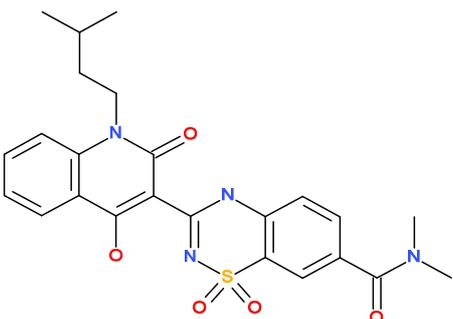
73

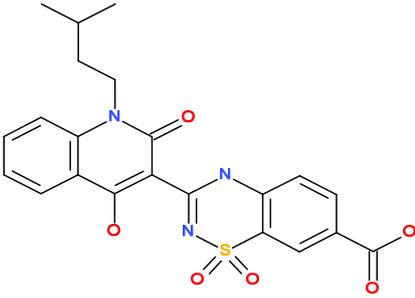
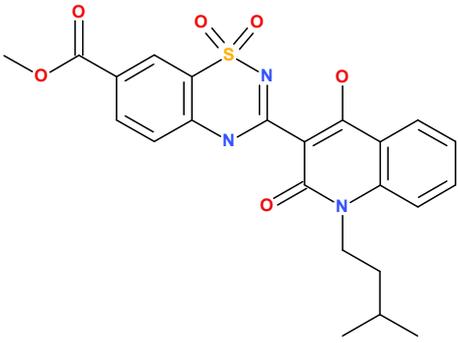
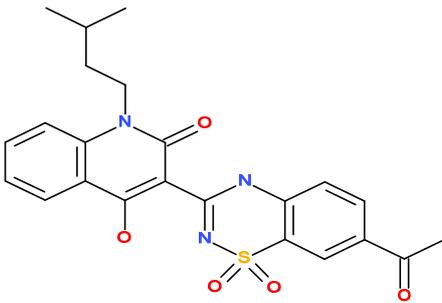
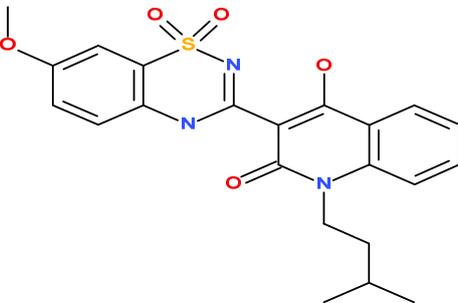
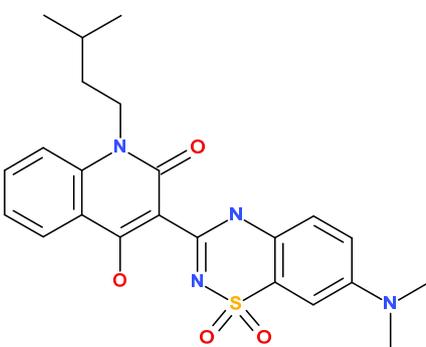
CMPD-73

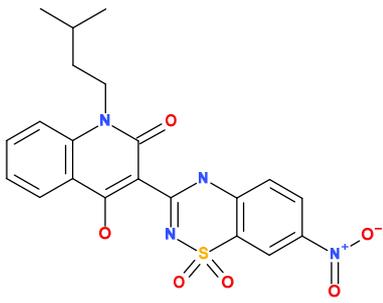
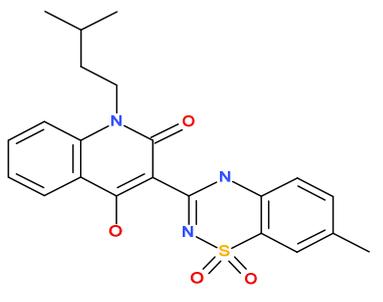
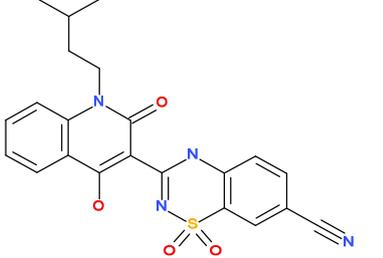
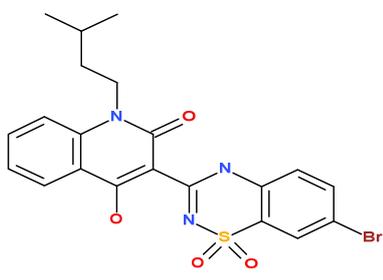
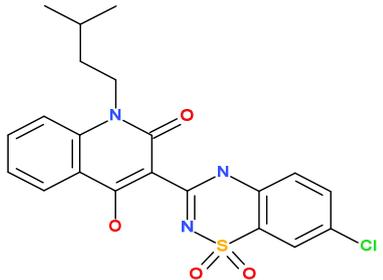
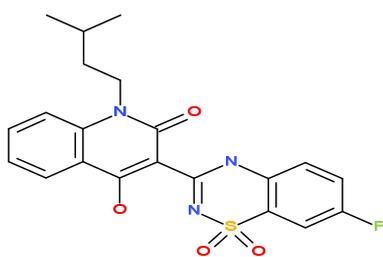
142

6.847



74	CMPD-74		6	8.221
75	CMPD-75		8	8.096
76	CMPD-76		23	7.638
77	CMPD-77		7	8.154
78	CMPD-78		102	6.991

79	CMPD-79		45	7.346
80	CMPD-80		65	7.187
81	CMPD-81		127	6.896
82	CMPD-82		55	7.259
83	CMPD-83		142	6.847

84	CMPD-84		238	6.623
85	CMPD-85		50	7.301
86	CMPD-86		57	7.244
87	CMPD-87		173	6.761
88	CMPD-88		76	7.119
89	CMPD-89		26	7.585

Creation of pharmacophore sites and common pharmacophore hypothesis

As the 3D-QSAR relies on the relationship between free energy change and equilibrium constant ($\Delta G^\circ = -2.3RT \log K_c$), the data for QSAR model generation should be expressed in terms of free energy changes occurring during biological response [16].

$$\Delta G^\circ = -2.3RT \log_{10} K_c$$

ΔG is Gibb's free energy, R is gas constant, T is temperature, K_c is equilibrium constant which is the ratio of the concentration of products and concentration of reactants. When measuring concentration values under equilibrium conditions, the concentration of the compound should be expressed as the inverse logarithm of compound concentration. The change in Gibb's free energy is directly proportional to the inverse logarithm of compound concentration [16].

$$\Delta G^\circ = -2.3RT \log_{10} K_c = \log 1/[S]$$

Where [S] is compound concentration

Hence, the *in vitro* inhibitory concentration (IC_{50}) values of 89 molecules were converted to corresponding pIC_{50} ($\log 1/IC_{50}$) values and used as dependent variable. The entire dataset is divided into actives and inactives based on activity threshold value (7.8) which was derived from dataset activity distribution (ranging from 6.0 to 8.3). All the ligands with activity value greater than threshold value were categorized as active (31 molecules) and those with activity value less than threshold value were considered inactive (58 molecules). Active ligands were chosen for identification of pharmacophore features, their frequency of occurrence and creation of pharmacophore sites points. The common pharmacophore hypothesis (CPH) was identified based on these site points using tree-based partitioning technique [17].

First, variant list of pharmacophore hypotheses was generated using the features and site points where the criteria were, a hypothesis should have minimum and maximum of 5 site points and must match at least 25/31 actives. Variant hypothesis not present in any of the ligands was filtered out and those common among maximum ligand molecules were picked for scoring and ranking. The best candidate CPH, which conveys 3D chemical characteristics that are critical for binding was selected for building the 3D-QSAR model.

Building 3D-QSAR model

3D-QSAR model was generated using the best candidate hypothesis selected based on scoring parameters. For a model generation, the entire data set was divided into a training set (70%) and test set (30%) in a random manner using "Automated Random Selection" option available in PHASE module [18]. PHASE has two approaches for building 3D-QSAR model—pharmacophore based and atom based

approach. In pharmacophore based QSAR, only pharmacophore features are considered and not whole ligand features, hence, in this study, an atom based QSAR model was built. In atom based QSAR modelling, each molecule is considered as a set of Van Der Waal's spheres (each atom corresponds to a sphere). Each Sphere is placed into one of the six categories—Hydrogen bond donor (Hydrogens attached to polar atoms), hydrophobic atom (carbons, halogens and hydrogens attached to them), Negative ionic atom (atoms with negative charge), Positive ionic atom (atoms with positive charge), electron withdrawing atoms (Polar atoms) and miscellaneous atoms (Other types of atoms). Van der Waal's spheres of training set molecules are placed into a grid of cubes and each cube is allocated zero or more bits depending on different types of atoms occupying the cube. This occupation pattern is used to create partial least squares (PLS) QSAR models which can be used as an independent variable for QSAR model generation. The number of PLS factors in each model can be 1/5 of the training set molecules.

RESULTS AND DISCUSSION

NS5B inhibitors offer therapeutic potential for treatment of hepatitis C. Several compounds binding to allosteric sites of the enzyme were discovered, however, none of them was launched yet. In an attempt to identify more potent inhibitors binding to Palm I allosteric site of the enzyme, pharmacophoric characteristic features that are crucial for binding to NS5B allosteric site were identified using already reported NS5B inhibitors. From a set of 31 actives among 89 benzothiadiazine analogues, a total of 102 variant hypotheses were generated. From the variant list, CPHs were identified and include AANRR which comprise pharmacophores of 22 ligands—AANRR (22), AAARR (15), AAANR (11), AARRR (10), AHNRR (10), and AAPRR (10).

The CPHs were scored and ranked to identify the best candidate hypothesis. The scoring algorithm includes alignment with actives, the site points and vectors, volume overlap, selectivity, number of active ligands matched, relative conformational energy and activity [17]. Also, these pharmacophore hypotheses should discriminate between actives and inactives. A pharmacophore hypothesis comprises information on critical sites required for binding of actives and also sites which are preventing inactive from binding.

To identify the pharmacophore models with more active and less inactive features, they were aligned with inactives and scored (In active). The survival score was adjusted by subtracting survival score of active with survival score of inactive (Survival (inactive)).

The hypothesis with maximum adjusted survival score is more preferred. After analysis of scores, the CPHs were ranked and best candidate hypothesis was selected which was used for building the 3D-QSAR model. The best candidate hypothesis along with its corresponding scores is summarized in table 2.

Table 2: Best candidate hypothesis along with scoring parameters

ID	Survival	Survival (Inactive)	Site	vector	Volume	Selectivity	#Matches	Energy	Activity	Inactive
AANRR.6	3.668	1.401	0.95	0.996	0.726	2.249	25	0	7.886	2.267

Survival—A weighed combination of Site, vector and volume, larger the survival value better is the alignment of ligand over common pharmacophore, Survival (inactive)—Survival score of active with survival score of inactive subtracted, Site—this score determines how closely the site points are superimposed on the pharmacophore, Vector—Alignment score, Volume—Measurement of volume overlap in ligand and pharmacophore alignment, Selectivity—uniqueness of the pharmacophore to a particular ligand, #Match—number of actives fitting to the hypothesis, Energy—Relative energy of the reference ligand in kcal/mol, Activity—Relative activity of the reference ligand, Inactive—Survival score of inactive.

The best candidate hypothesis AANRR.6 was associated with five-point hypotheses and comprise of two hydrogen bond acceptors (pink sphere with arrows), negatively charged group (pink sphere), two aromatic rings (orange rings) (fig. 1A). When aligned with the molecule with maximum fitness score, the two oxygens of benzothiadiazinyl ring occupied the hydrogen bond acceptor spheres;

oxygen at C-2 position of quinolinone is engaged with a negatively charged sphere and aromatic rings were mapped to quinolinone and benzothiadiazinyl rings (fig. 1B). Alignments of actives and inactives are shown in fig. 1C and D respectively. From fig. 1 we can easily identify that active ligand is having good alignment than an inactive one.

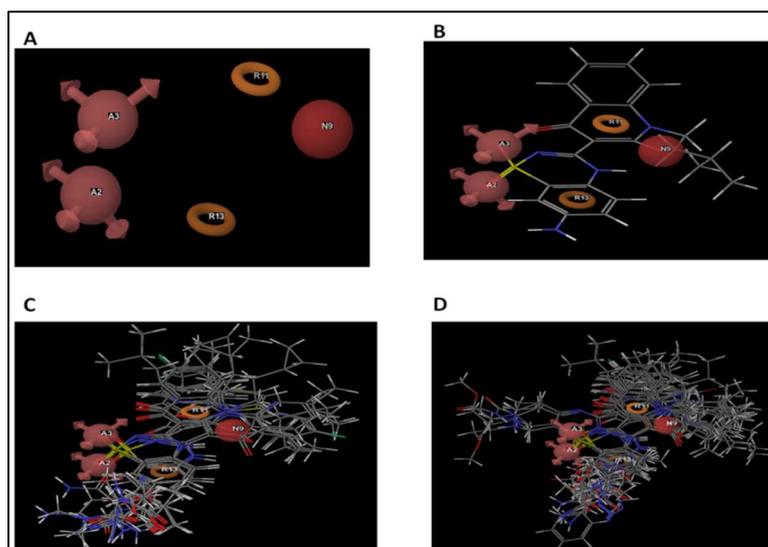


Fig. 1: Best common pharmacophore hypothesis: (A) Pharmacophore site points (B) Alignment of best fit molecule (CMPD-20) and site points (C) alignment of active ligands with CPH (D) alignment of inactive ligands with CPH

The spatial arrangement of features along with distances and angles were depicted in fig. 2 A and B. AANRR.6 was selected for building an atom based 3D-QSAR model. For generating the 3D-QSAR model, the dataset was randomly divided into training set and test set. A five-component PLS model with good statistics was generated using 62

training set compounds and validated using 24 test set compounds. A PLS factor of five is used to generate the model as it was observed that there was a significant increment in statistics and predictivity with an increase in PLS factors. The pharmacophore hypothesis AANRR.6 yielded a 3D QSAR model with good statistical scores.

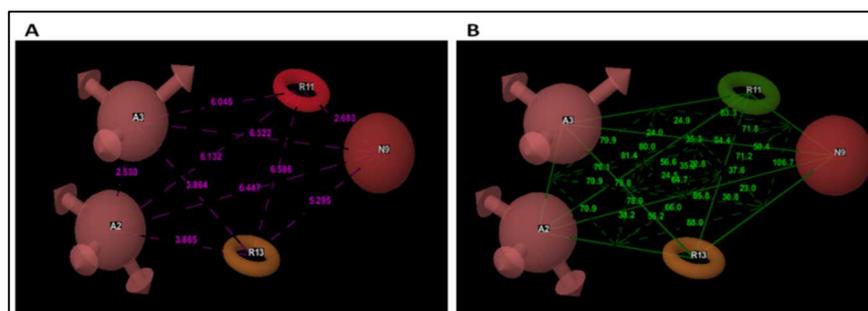


Fig. 2: Pharmacophore site distances (A) Pharmacophore site of angles (B)

As suggested by many QSAR modellers the model should be verified by goodness of Fit (R^2), robustness or internal predictivity (Q^2) and external predictivity (validation using external chemical compounds that were never included in the training set even at least once during the model building) [19]. The aim of internal predictivity is preliminary model validation and to select more robust and predictive model whereas, the goal of external predictivity is to evaluate the predictive power of the developed model by using it against a set of completely unseen chemical structures.

The model showed a good R^2 value of 0.924, Fit value (F) =138, standard deviation, (SD) =0.24 for the training set. Lower value of standard deviation (SD) reflects a lesser deviation from biological data. The model was validated using the internal test set molecules

and the validity of the model is expressed as cross-validated R^2 (Q^2) which is the characteristic of the predictive ability of the model. The Q^2 value less than R^2 and close to R^2 indicate the good predictive ability of the model. Cross-validation was done using leave one out (LOO) method, where one molecule from the training set is left out (only once) and R is computed for the left out molecule using model generated with remaining training set molecules. Sometimes more than one molecule is left out at one time (leave many out (LMO)). The Q^2 value (0.77) so obtained emphasizes the model generated is best with RMSE value of 0.299. The fit of the model can be determined by root-mean-squared error (RMSE), an error between the mean of experimental activities and predicted activities. Lower RMSE reflects good predictive quality of the model. The statistical results of atom based QSAR model built are summarized in table 3.

Table 3: Statistical parameters of QSAR model built based on AANRR.6

ID	PLS Factors	SD	R^2	F	P	RMSE	Q^2	Pearson-R
AANRR.6	1	0.624	0.4805	55.5	4.278e-010	0.4617	0.4639	0.724
	2	0.4927	0.6821	63.3	2.08e-015	0.3975	0.6027	0.83
	3	0.3882	0.806	80.3	1.236e-020	0.3393	0.7105	0.857
	4	0.3503	0.8447	77.5	2.213e-022	0.3605	0.6733	0.839
	5	0.2475	0.9249	138	3.454e-030	0.2993	0.7748	0.908

The interaction between ligand and receptor active site is analyzed using QSAR visualization option in PHASE and the interaction is represented by clouded cubes where blue and red cubes represent favourable and unfavourable regions respectively. The pharma-

cophore hypothesis that yielded 3D-QSAR model is aligned with ligands and features essential for NS5B inhibitory activity were analyzed. The 3D-QSAR visualization for most active compound and least active compound were given in fig. 3 and fig. 4 respectively.

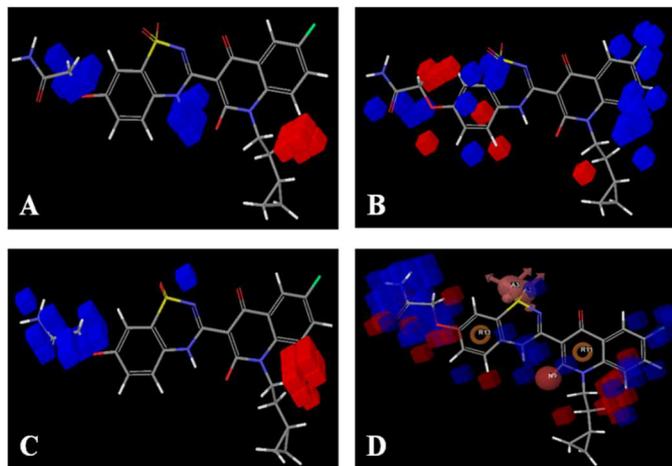


Fig. 3: Visualization of QSAR Model in the context of the active compound (Compound 22) in the training set. A) Hydrogen bond feature B) Hydrophobic feature C) Electron withdrawing feature D) Combination of all the pharmacophoric features

In the most active compound, positive correlation of H-Bond donor feature was seen at the nitrogen atom of benzothiadiazine sulfamide ring and the oxygen atom of oxyacetamide group attached to C7 carbon. Benzothiadiazole analogues with oxyacetamide substitution at C7 position showed low micromolar inhibitory activity in the 1b replicon assay also it makes favourable hydrogen bonding with Asn 291 [15]. The electron withdrawing atoms which are part of oxyacetamide group were also positively correlated with NS5B inhibitory activity. Among the hydrophobic groups present in the

molecule, Quinoline ring is positively correlated with activity. Substitution of fluorine at C6 position of the quinolone ring has been shown to improve the cellular activity [20]. In the less active compound, nitrogen atoms of 5-hydroxy pyridazine ring (hydrogen bond donor) and benzyl moiety of benzothiadiazine ring (hydrophobic group) are negatively correlated with activity (fig. 4). This molecule was also lacking a substituent at the C7 position, where there more probability of H-bond formation and filling the binding pocket.

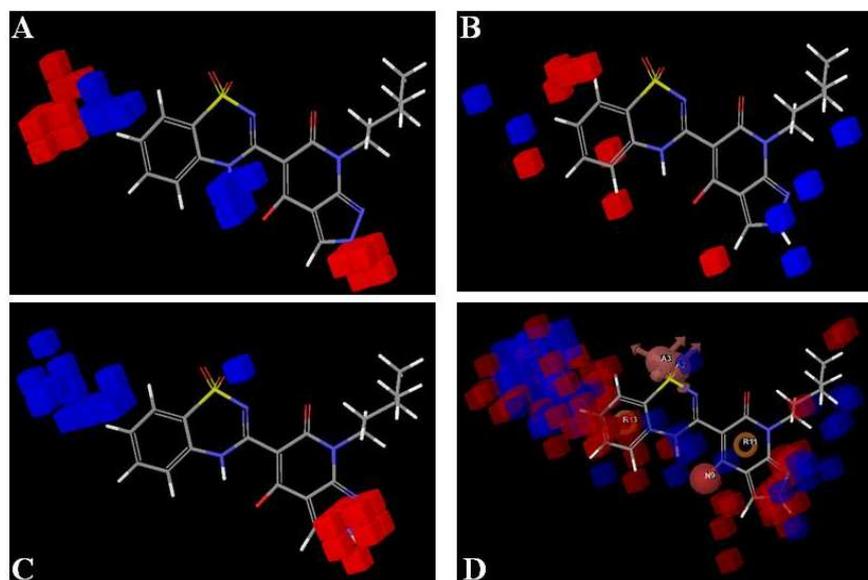


Fig. 4: Visualization of QSAR model in the context of inactive compound (Compound 50) in the training set. A) Hydrogen bond feature B) hydrophobic feature C) electron withdrawing feature D) combination of all the pharmacophoric features

Fitness graph of observed values Vs PHASE predicted values with 45° regression line are depicted in fig. 5. The Fitness graph emphasizes reasonably good alignment between experimental values and

predicted values for more than 70% of the dataset molecules. Predicted activity values of training and test set molecules are furnished in table 4.

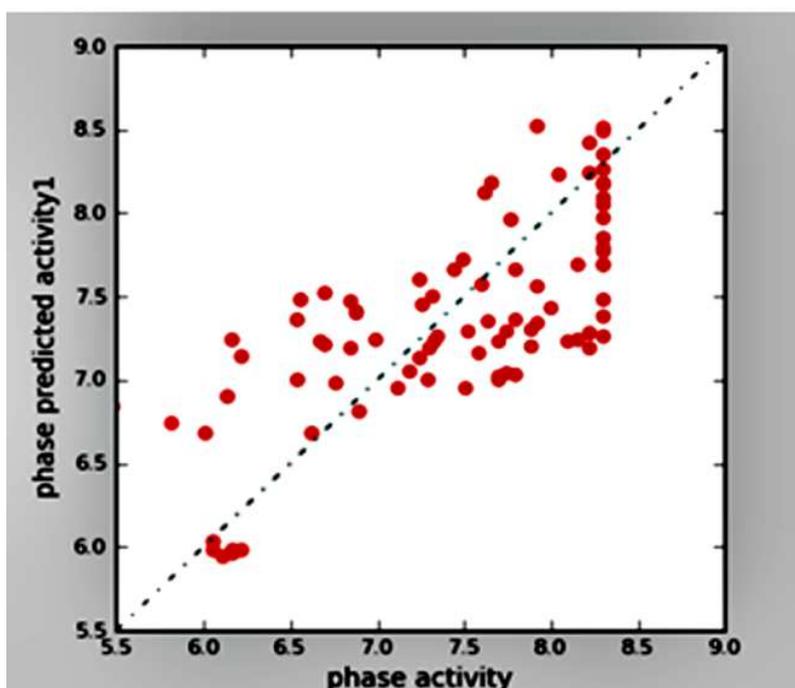


Fig. 5: Fitness graph between observed and predicted values

Table 4: Predicted activities of training and test set molecules

S. No.	Compound No	QSAR set	Observed activity	Predicted activity	Pharm set	Fitness
1	CMPD-1	Test	6.056	5.73	Inactive	1.54
2	CMPD-2	Test	6.168	5.95	Inactive	1.52
3	CMPD-3	Training	6.218	6.30	Inactive	2.12
4	CMPD-4	Test	7.745	7.32	Inactive	2.99
5	CMPD-5	Test	7.509	7.50	Inactive	2.95
6	CMPD-6	Training	7.698	7.68	Inactive	2.29
7	CMPD-7	Test	8.301	8.08	Active	2.60
8	CMPD-8	Training	8.301	8.33	Active	2.61
9	CMPD-9	Training	8.301	7.80	Active	2.62
10	CMPD-10	Training	8.221	8.21	Active	2.54
11	CMPD-11	Training	8.301	8.24	Active	1.05
12	CMPD-12	Training	8.301	8.25	Active	1.04
13	CMPD-13	Training	8.301	8.28	Active	1.00
14	CMPD-14	Training	8.301	8.35	Active	1.01
15	CMPD-15	Test	7.920	7.78	Active	1.04
16	CMPD-16	Test	7.795	7.55	Inactive	1.10
17	CMPD-17	Training	6.673	7.08	Inactive	1.10
18	CMPD-18	Training	8.221	7.91	Active	2.30
19	CMPD-19	Test	7.522	7.97	Inactive	3.00
20	CMPD-20	Training	7.886	7.99	Active	3.00
21	CMPD-21	Training	8.301	8.52	Active	2.69
22	CMPD-22	Training	8.301	8.22	Active	2.52
23	CMPD-23	Training	8.154	8.14	Active	1.03
24	CMPD-24	Training	7.327	7.23	Inactive	1.06
25	CMPD-25	Training	8.000	7.92	Active	2.74
26	CMPD-26	Test	7.698	7.27	Inactive	1.06
27	CMPD-27	Training	7.886	7.76	Active	2.95
28	CMPD-28	Training	7.494	7.45	Inactive	2.58
29	CMPD-29	Training	6.698	7.03	Inactive	1.06
30	CMPD-30	Test	7.602	7.48	Inactive	2.67
31	CMPD-31	Training	8.301	8.44	Active	2.60
32	CMPD-32	Training	7.795	7.99	Inactive	2.67
33	CMPD-33	Training	6.539	6.61	Inactive	2.22
34	CMPD-34	Training	6.168	6.13	Inactive	1.52
35	CMPD-35	Training	6.010	6.89	Inactive	2.31
36	CMPD-36	Training	6.111	5.83	Inactive	1.55
37	CMPD-37	Test	6.540	5.97	Inactive	2.41
38	CMPD-38	Training	6.137	5.71	Inactive	2.45

39	CMPD-39	Test	6.559	6.57	Inactive	2.13
40	CMPD-40	Training	6.879	6.93	Inactive	2.29
41	CMPD-41	Test	6.056	6.04	Inactive	1.55
42	CMPD-42	Training	5.816	5.82	Inactive	2.46
43	CMPD-43	Training	6.218	6.15	Inactive	1.53
44	CMPD-44	Training	6.698	6.83	Inactive	2.50
45	CMPD-45		5.520		Inactive	
46	CMPD-46	Test	7.244	7.31	Inactive	2.53
47	CMPD-47	Test	7.698	7.35	Inactive	2.87
48	CMPD-48	Training	7.795	7.64	Inactive	2.95
49	CMPD-49	Training	7.292	7.24	Inactive	2.31
50	CMPD-50	Training	5.306	5.88	Inactive	2.66
51	CMPD-51	Training	5.476	5.05	Inactive	1.03
52	CMPD-52				Inactive	
53	CMPD-53	Training	6.879	7.06	Inactive	2.66
54	CMPD-54	Test	7.318	7.56	Inactive	2.66
55	CMPD-55	Training	8.045	7.82	Active	2.51
56	CMPD-56	Test	7.657	7.69	Inactive	2.52
57	CMPD-57	Test	7.920	8.30	Active	2.85
58	CMPD-58	Training	8.301	8.35	Active	2.51
59	CMPD-59	Training	7.920	7.92	Active	2.50
60	CMPD-60	Training	8.301	8.05	Active	2.51
61	CMPD-61	Training	7.443	7.98	Inactive	2.86
62	CMPD-62	Training	8.301	8.07	Active	2.87
63	CMPD-63		6.694		Inactive	
64	CMPD-64	Training	8.221	8.41	Active	2.44
65	CMPD-65	Training	8.301	8.21	Active	2.49
66	CMPD-66	Training	8.301	7.98	Active	2.52
67	CMPD-67	Training	7.769	7.65	Inactive	2.57
68	CMPD-68	Training	7.619	7.57	Inactive	2.57
69	CMPD-69	Training	8.301	8.28	Active	2.96
70	CMPD-70	Training	7.744	7.76	Inactive	2.90
71	CMPD-71	Training	8.301	7.88	Active	2.94
72	CMPD-72	Training	6.164	6.34	Inactive	2.90
73	CMPD-73	Test	6.847	7.08	Inactive	2.95
74	CMPD-74	Test	8.221	7.70	Active	2.91
75	CMPD-75	Training	8.096	7.76	Active	2.90
76	CMPD-76	Test	7.638	7.24	Inactive	0.99
77	CMPD-77	Training	8.154	7.82	Active	2.95
78	CMPD-78	Training	6.991	7.10	Inactive	2.19
79	CMPD-79	Training	7.346	7.60	Inactive	2.96
80	CMPD-80	Training	7.187	7.34	Inactive	2.24
81	CMPD-81	Training	6.896	6.92	Inactive	2.35
82	CMPD-82	Test	7.259	7.61	Inactive	1.08
83	CMPD-83	Training	6.847	7.46	Inactive	2.96
84	CMPD-84	Training	6.623	6.81	Inactive	2.35
85	CMPD-85	Test	7.301	7.46	Inactive	3.00
86	CMPD-86	Training	7.244	7.50	Inactive	2.31
87	CMPD-87	Training	6.761	7.08	Inactive	2.23
88	CMPD-88	Test	7.119	7.04	Inactive	2.23
89	CMPD-89	Training	7.585	7.50	Inactive	3.00

There has been a debate on QSAR model validation topic over the past decade and was the subject of discussion in scientific communities [21, 22]. The consensus was that both cross-validation (internal validation) and external validation are philosophically different approaches and both are required to characterize the model [19]. As per organisation for economic co-operation and development (OECD) Principles, the QSAR model generated should be verified for the goodness of fit, applicability domain, robustness and predictive capability [23].

In internal validation, the model will be validated using test set molecules that were also part of training set during multiple iterations of model generation. Internal Validation determines the statistical robustness of the model and this preliminary validation is required to proceed for external validation. In external validation, a new set of molecules (prediction set) which were never part of

training set during model development will be used for verifying the model. The prediction set can be either separated from the dataset selected before the model generation or collated from the literature after model generation. The number of molecules in the prediction set should be 10–20% of the dataset (both training and test set molecules) used for model generation.

External validation demonstrates the predictive power of the model. In the current study, though the QSAR model was validated using cross-validation, this validation is not sufficient to demonstrate the predictive capability of the model (24).

Hence, the generated model was also validated using the prediction set comprising 10 benzothiadiazine analogues [25]. The activity values predicted with 3D QSAR model built using AANRR.6 were given the table 5. For most of the compounds, the observed values were aligned with predicted values.

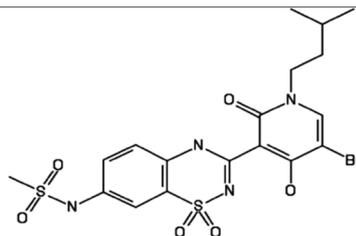
Table 5: Predicted and experimental values of external dataset comprising benzothiadiazine analogues

S. No.	Compound Name	Structure	Observed activity	Predicted activity
1	EX_VT_1		6.09	6.42
2	EX_VT_2		6.32	6.75
3	EX_VT_3		7.17	7.23
4	EX_VT_4		6.39	6.75
5	EX_VT_5		7.46	6.98
6	EX_VT_6		7.62	6.89
7	EX_VT_7		7.96	6.93
8	EX_VT_8		6.48	6.99

9 EX_VT_9

7.26

6.93



The robustness and predictive power of the model were proved through internal and external validation. This model was used to predict the NS5B inhibitory potential of the drugs (Aprepitant, Doripenem and Argathroban) that showed desired binding interactions with residues of the binding pocket [26]. It was found that the predicted activities of only Aprepitant are in line the

activities of the existing active palm I inhibitors. The details of phase predicted activities were furnished in table 6. Structural features of Aprepitant fit with 4 (A2, A3, R14 and N11) out of 5 pharmacophoric site points. The genomics analysis performed earlier also emphasized desired overlap between HCV pathophysiology and Aprepitant effects [26].

Table 6: Phase predicted activities of aprepitant

Compound	No of sites matched	Phase predicted activity_1	Phase predicted activity_2	Phase predicted activity_3	Phase predicted activity_4	Phase predicted activity_5
Aprepitant	4	6.91	6.84	6.95	6.90	6.93

HCV research has gained momentum in the recent past. Many hepatitis C drugs were launched in the market since 2011 and sofosbuvir is considered as a breakthrough drug. However, in spite of tremendous progress in hepatitis C drug discovery still there exist an unmet medical need in this therapeutic area. Also, the marketed NS5B inhibitors act by binding to the active site and none of the NS5B allosteric inhibitors was launched yet. In order to strengthen the pipeline of NS5B inhibitors targeting palm I site (present close to active site), a 3D QSAR model was generated in the present study that can be used to optimize existing NS5B inhibitors binding to palm I, to predict NS5B inhibitory activity of new molecular entities and also to identify new NS5B inhibitors through virtual screening. An Atom-based QSAR model was built using pharmacophoric features common among benzothiadiazine derivatives.

The common pharma-cophore hypothesis AANRR.6 is the best hypothesis with survival score of 3.668. The best candidate hypothesis AANRR.6 has two hydrogen bond acceptors, one negative ionization potential and two aromatic rings. The 3D QSAR model built using AANRR.6 showed good statistical scores ($R^2 = 0.924$, Q^2 -value 0.774) and prediction power was demonstrated using external dataset. Prediction of NS5B inhibitory activity of drugs identified in the earlier work emphasized the fact that Aprepitant could be a potential inhibitor.

CONCLUSION

The 3D QSAR model emphasized the positive correlation of nitrogen atom of benzothiadiazine sulfamide ring, oxyacetamide group attached to C7 carbon of benzothiadiazine and sulfonamide oxygens with NS5B inhibitory activity. NS5B inhibitory potential predicted by 3D-QSAR model emphasizes the Aprepitant as a potential inhibitor but needs to be further evaluated under *in vitro* conditions.

CONFLICTS OF INTERESTS

All authors have none to declare

Prasanthi Polamreddy: Carried out QSAR studies and drafted the work

Vinita Vishwakarma: Provided guidance and suggestions during the work and reviewed the article draft.

Manoj Kumar Mahto: Provided guidance and suggestions during the work and reviewed the article draft

REFERENCES

1. Waheed Y, Bhatti A, Ashraf M. RNA dependent RNA polymerase of HCV: a potential target for the development of antiviral drugs. *Infect Genet Evol* 2015;14:247–57.
2. Deore RR, Chern JW. NS5B RNA dependent RNA polymerase inhibitors: the promising approach to treat hepatitis C virus infections. *Curr Med Chem* 2010;17:3806–26.
3. Lee H, Liu Y, Mejia E, Paul AV, Wimmer E. The C-terminal hydrophobic domain of hepatitis C virus RNA polymerase NS5B can be replaced with a heterologous domain of poliovirus protein 3A. *J Virol* 2006;80:11343–54.
4. Ferrari E, Wright-Minogue J, Fang JWS, Baroudy BM, Lau JYN, Hong Z. Characterization of soluble hepatitis C virus RNA-dependent RNA polymerase expressed in escherichia coli. *J Virol* 1999;73:1649–54.
5. Ferrari E, He Z, Palermo RE, Huang HC. Hepatitis C virus NS5B polymerase exhibits distinct nucleotide requirements for initiation and elongation. *J Biol Chem* 2008;283:33893–901.
6. Kim Y, Russell WK, Thomson M, David H, Kao CC, Kim Y, *et al.* Functional analysis of RNA binding by the hepatitis C virus RNA-dependent RNA polymerase. *J Biol Chem* 2005;280:38011-9.
7. Choi KH, Rossmann MG. RNA-dependent RNA polymerases from flaviviridae. *Curr Opin Struct Biol* 2009;19:746–51.
8. Pauwels F, Mostmans W, Quiryren LMM, van der Helm L, Boutton CW, Rueff AS, *et al.* Binding-site identification and genotypic profiling of hepatitis C virus polymerase inhibitors. *J Virol* 2007;81:6909–19.
9. Eltahla AA, Luciani F, White PA, Lloyd AR, Bull RA. Inhibitors of the hepatitis C virus polymerase; mode of action and resistance. *Viruses* 2015;7:5206–24.
10. Kayali Z, Schmidt WN. Finally sofosbuvir: an oral anti-HCV drug with wide performance capability. *Pharmgenomics Pers Med* 2014;7:387–97.
11. Feld JJ, Foster GR. Review second generation direct-acting antivirals—Do we expect major improvements? *J Hepatol* 2016;65:S130–42.
12. Guide QS. Think and Do Quick Start Guide. Phase 3.3; 2008.
13. Das D, Hong J, Chen SH, Wang G, Beigelman L, Seiwert SD, *et al.* Recent advances in drug discovery of benzothiadiazine and related analogs as HCV NS5B polymerase inhibitors. *Bioorg Med Chem* 2011;19:4690–703.
14. Tedesco R, Chai D, Darcy MG, Dhanak D, Fitch DM, Gates A, *et al.* Synthesis and biological activity of heteroaryl 3-(1,1-dioxo-2H-(1,2,4)-benzothiadiazin-3-yl)-4-hydroxy-2(1H)-quinolinone

- derivatives as hepatitis C virus NS5B polymerase inhibitors. *Bioorg Med Chem Lett* 2009;19:4354–8.
15. Shaw AN, Tedesco R, Bambal R, Chai D, Concha NO, Darcy MG, *et al.* Substituted benzothiazine inhibitors of Hepatitis C virus polymerase. *Bioorg Med Chem Lett* 2009;19:4350–3.
 16. Moro S. Statistical concepts in QSAR 2004;50:1–10.
 17. Dixon SL, Smondyrev AM, Rao SN. PHASE: a novel approach to pharmacophore modelling and 3D database searching. *Chem Biol Drug Des* 2006;67:370–2.
 18. Shah Ua, Deokar HS, Kadam SS, Kulkarni VM. Pharmacophore generation and Atom-based 3D-QSAR of novel 2-(4-methylsulfonylphenyl)pyrimidines as COX-2 inhibitors. *Mol Divers* 2010;14:559–68.
 19. Gramatica P. External evaluation of QSAR models, in addition to cross-validation: verification of predictive capability on totally new chemicals. *Mol Inform* 2014;33:311–4.
 20. Tedesco R, Shaw AN, Bambal R, Chai D, Concha NO, Darcy MG, *et al.* 3-(1,1-dioxo-2H-(1,2,4)-benzothiadiazin-3-yl)-4-hydroxy-2(1H)-quinolinones, potent inhibitors of hepatitis C virus RNA-dependent RNA polymerase. *J Med Chem* 2006;49:971–83.
 21. Gramatica P. Principles of QSAR models validation: internal and external. *QSAR Comb Sci* 2007;26:694–701.
 22. Hawkins DM, Basak SC, Mills D. Assessing model fit by cross-validation. *J Chem Inf Comput Sci* 2003;43:579–86.
 23. Tichy M, Rucki M. Validation of QSAR models for legislative purposes. *Interdiscip Toxicol* 2009;2:184–6.
 24. Thangaraj Sindhu, Sundaraj Rajamanikandan, Dhanapal Durgapriya, Jebamalai Raj Anitha, Selvaraj Akila, Velliyur Kanniyapan Gopalakrishnan. Molecular docking and qsar studies on plant derived bioactive compounds as potent inhibitors of dekoprotein. *Asian J Pharm Clin Res* 2011;4:67–71.
 25. Donner PL, Xie Q, Pratt JK, Maring CJ, Kati W, Jiang W, *et al.* Des-A-ring benzothiadiazines: inhibitors of HCV genotype 1 NS5B RNA-dependent RNA polymerase. *Bioorg Med Chem Lett* 2008;18:2735–8.
 26. Polamreddy P, Vishwakarma V, Gattu N. Discovery of HCV NS5B Palm I allosteric inhibitors using computational techniques-drug repurposing; 2016.