JPP JOURNAL OF PROTEINS AND PROTEOMICS
4(3), December 2013, pp. 191-216

Research Article

A QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP AND MOLECULAR DOCKING STUDY ON A SERIES OF PYRIMIDINES ACTING AS ANTI-HEPATITIS C VIRUS AGENTS

Sakshi Gupta[1], Satya P. Gupta[1*] and Neeraj Agarwal[2]

[1]Department of Pharmaceutical Technology and [2]Department of Biotechnology, Meerut Institute of Engineering and Technology Meerut 250005, U.P, India

Abstract: A QSAR and molecular modeling study was performed on a series of pyrimidines acting as hepatitis C virus inhibitors. In this case, anti-HCV potency of the compounds was found to be significantly correlated with the hydrophobic property of the molecule, Kier's first-order valence molecular connectivity index for a particular substituent, total structure connectivity index of the molecule, and an indicator parameter used for the presence of benzothiazole ring. The validity of the correlation was judged by leave-one-out jackknife procedure and predicting the activity of some test compounds. Using the correlation obtained, some new compounds of high potency have been predicted in the series. A docking study using Molegro Virtual Docker was performed on these predicted compounds to decipher their interactions with the receptor. It was observed that all the predicted compounds had better interaction energy and docking score than the ligand complexed with the protein.

Keywords: Hepatitis C virus inhibitors; QSAR study; Pyrimidines; Docking

Introduction

Hepatitis C virus (HCV) is a single, enveloped positive sense RNA virus belonging to the family of *Flaviviridae* and having size between 55-65 nm. It is the major cause of blood borne diseases, such as non-A, non-B hepatitis and most commonly hepatitis C in humans affecting over 170 million people throughout the world. The length of HCV genome is approximately 9.6 kb with a polyprotein of about 3,000 amino acids encoded in it. The host and HCV-encoded proteases are responsible for its cleavage into 10 structural and non-structural components (Pfefferkorn *et al.*, 2005; Ishida *et al.*, 2006; Chinnaswamy *et al.*, 2010).

Corresponding Author: **Satya P. Gupta**
E-mail: spgbits@gmail.com

Received: October 19, 2013
Accepted: November 21, 2013
Published: December 31, 2013

HCV virus is classified into seven genotypes 1-7 based on genetic differences among different HCV species. Each genotype is further subdivided into various subtypes (denoted by lower case letters such as 1a, 1b, etc.). The genomic compositions of various subtypes of a particular genotype differ generally by 20-25%. 60% of all cases of HCV virus infection are caused by subtypes 1a and 1b which are found throughout the world. The genome of HCV virus is evolved rapidly due to the errors created during the replication of RNA-dependent RNA polymerase (Simmonds *et al.*, 1993; Labrador and Xavier, 2008).

Chronic HCV infections are also responsible for many liver diseases like hepatitis, liver cirrhosis, liver fibrosis, hepatocellular carcinoma, and other forms of liver malfunctioning which ultimately develop into liver cancer if not treated in time. The disease caused due to HCV has widespread impact on humans; therefore there is

an urgent need to develop novel and potent anti-HCV agents so as to compliment the already existing therapies. The current therapy that is available until now includes interferon-α (INF-α) and ribavirin, but the overall chances of curing the disease in this case is only up to 50% and the undesirable side effects are also associated with it (Zuoa *et al.*, 2007; Kwong *et al.*, 2012; Li *et al.*, 2010; Wei *et al.*, 2009).

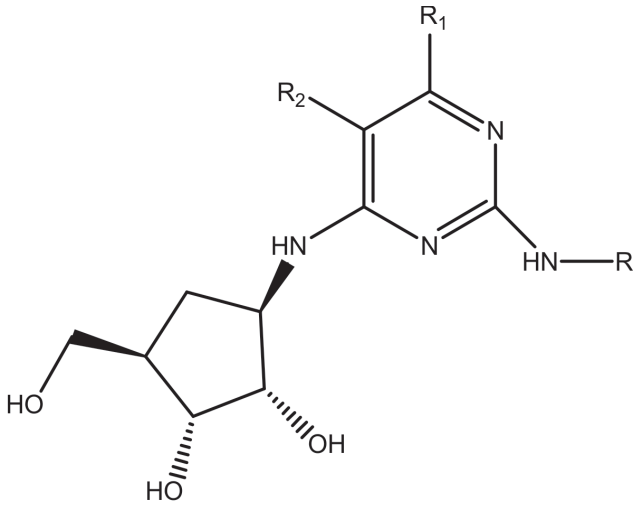
The main targets of HCV virus are liver cells where its replication takes place and the site for entering this virus is skin. The HCV goes on infecting other healthy cells after being released from the infected cell. The first phase of HCV infection is known as acute phase and it has flu like symptoms such as jaundice, abdominal pain, joint pain, nausea and vomiting, fatigue and appetite loss, dark urine, etc. (Labrador and Xavier, 2008).

The HCV antibody test is used to diagnose HCV infection. It is of two types: Enzyme Immunoassay (EIA) and Recombinant Immunoblot Assay (RIBA). A polymerase chain reaction RNA test is used to confirm the disease after the test for antibody comes positive (*http://www.medicinenet.com/hepatitis_c/patient-comments-552-page2.htm*).

The HCV inhibitors are mainly of two types: NS3 protease inhibitors and NS5B polymerase inhibitors. The NS3 protease inhibitors are further subdivided into two types, namely covalent and non-covalent inhibitors. Similarly, NS5B polymerase inhibitors are also of two types, namely nucleoside and non-nucleoside inhibitors (Paulson *et al.*, 2009). In the present communication, we have focused on developing the novel inhibitors of HCV NS5B RNA-dependent RNA polymerase with improved potency and lesser side effects with the help of quantitative structure-activity relationship (QSAR) and molecular modeling studies.

Materials and Methods

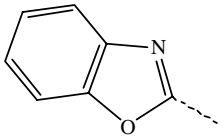
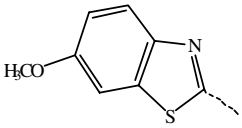
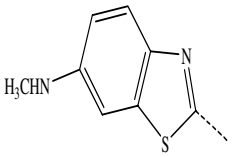
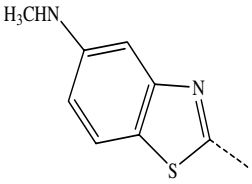
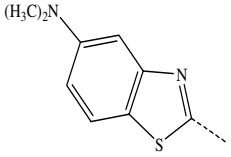
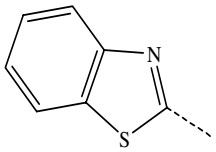
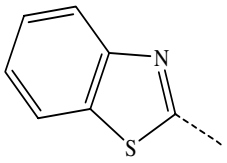
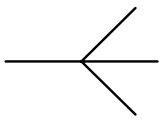
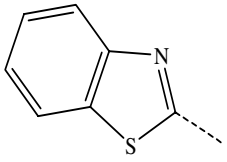
We have compiled a large series of pyrimidine derivatives (I) along with their anti-HCV activity from different communications (Arasappan *et al.*, 2012; Bennet *et al.*, 2012; Kwong *et al.*, 2012; Girijavallabhan *et al.*, 2012). The whole series of compounds was divided into two subsets, the training set comprising of 90 compounds and the test set comprising of 51 compounds. Compounds of training set are listed in Table 1 and those of test set in Table 2. In both the sets, compounds were selected keeping in view that there are wide variations in activities as well as in structures of the compounds. For making a QSAR study on the training set, a large number of descriptors were calculated using Chemdraw 2004, Chemsketch version 11.0 and e-Dragon software (*http://www.vcclab.org/lab/edragon/*). Both the tables list only those descriptors that were found to be significant in deriving QSAR model. Among these parameters, ClogP refers to hydrophobicity of the molecule, $^{1}\chi^{v}$ to Kier's first-order valence molecular connectivity index, and Xt to total structure connectivity index of the whole molecule. Several other parameters were also calculated but they were not found to be significant. An indicator parameter 'I' has also been used that refers to the presence of a benzothiazole ring at R_2-position. It has been given a value of 1 if Ar-substituent has a 2-methyl benzothiazole ring, otherwise its value is zero. In activity term log ($1/EC_{50}$), EC_{50} refers to molar concentration of the compound leading to 50% effect on inhibition of HCV RNA replication.

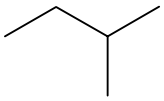
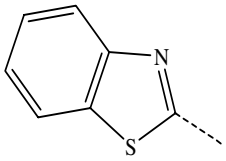
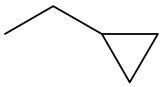
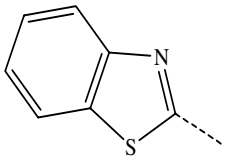
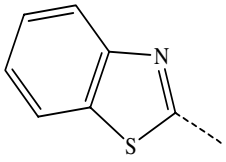
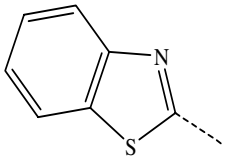
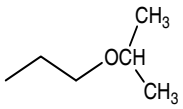
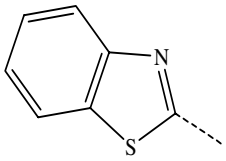
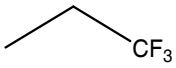
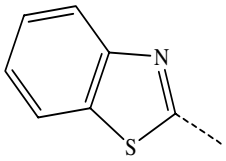
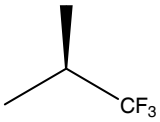
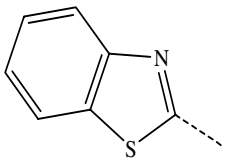
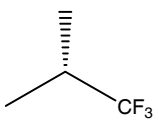
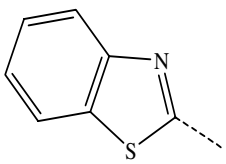


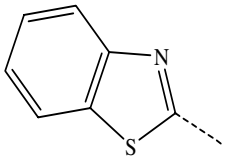
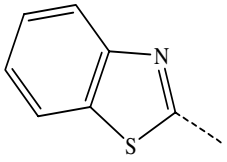
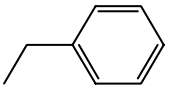
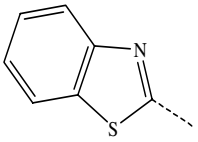
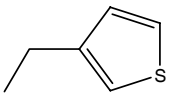
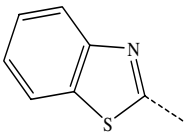
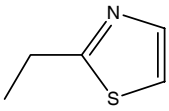
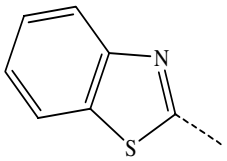
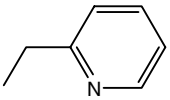
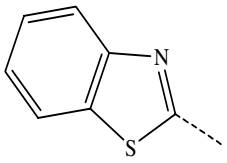
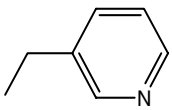
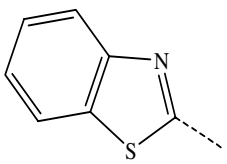
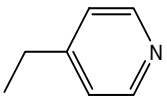
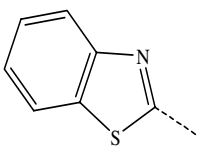
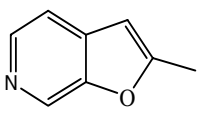
(I)

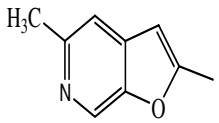
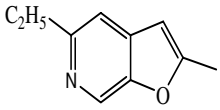
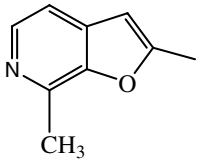
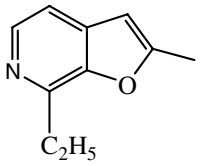
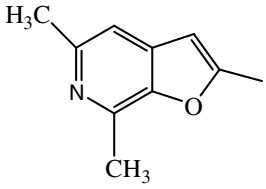
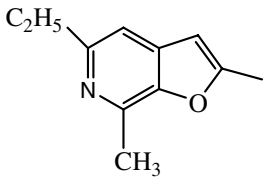
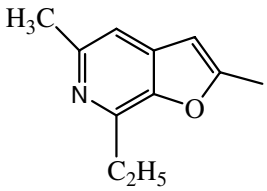
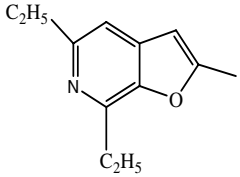
Docking. Molecular docking is a computational technique for exploration of the possible binding modes of a substrate or inhibitor in a given enzyme or receptor to give the optimal interactions (Gupta, 2011). To perform a docking,

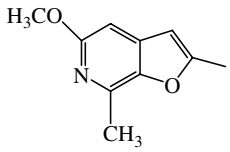
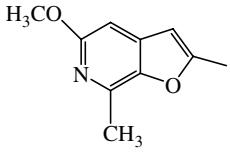
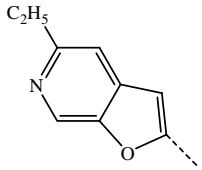
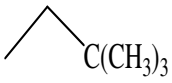
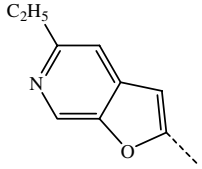
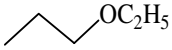
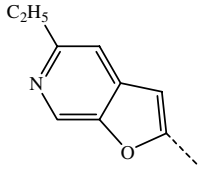
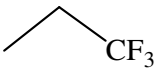
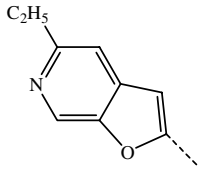
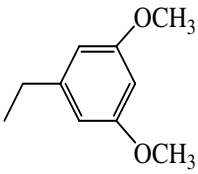
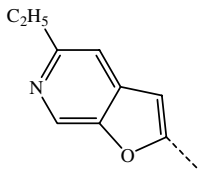
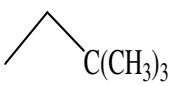
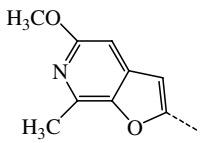
Table 1
Pyrimidine Derivatives (I as Shown in Materials and Methods) Belonging to Training Set and their anti-HCV Activity and Physicochemical Parameters

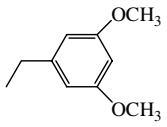
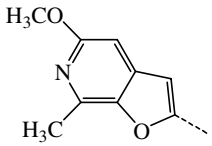
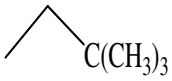
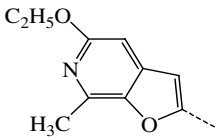
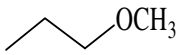
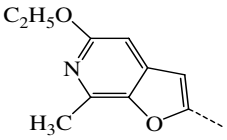
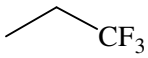
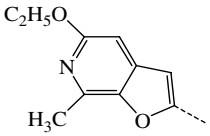
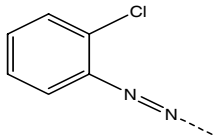
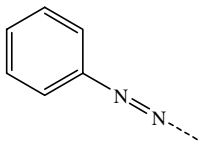
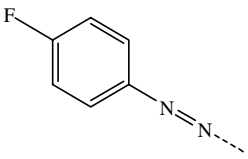
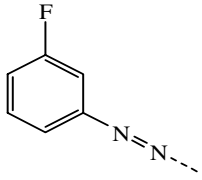
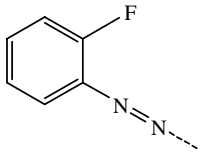
Compd	*R*	R_1	R_2	*ClogP*	$^1\chi^v_{R2}$	*Xt*	*I*	*log (1/EC$_{50}$)*		
								Obsd.	*Calcd Eq.1*	*Pred. LOO*
1	H	Cl		1.04	2.68	0.22	0.0	7.30	7.26	7.26
2	H	Cl		1.96	4.02	0.22	1.0	7.35	7.37	7.37
3	H	Cl		1.96	4.16	0.22	1.0	7.30	7.42	7.43
4	H	Cl		1.96	4.16	0.22	1.0	7.70	7.42	7.39
5	H	Cl		2.28	4.53	0.22	1.0	7.45	7.52	7.52
6	CH_3	CH_3		2.24	3.50	0.22	1.0	7.30	7.18	7.17
7	C_2H_5	CH_3		2.77	3.50	0.22	1.0	7.00	7.08	7.08
8		CH_3		3.48	3.50	0.21	1.0	7.04	7.14	7.15

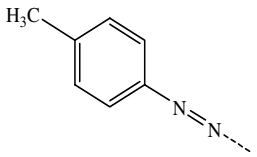
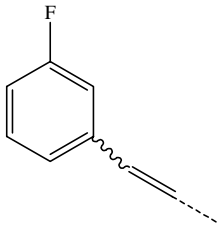
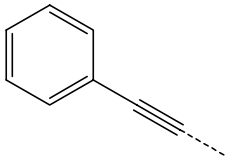
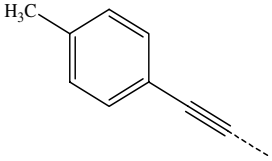
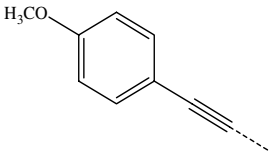
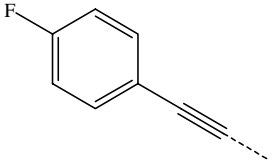
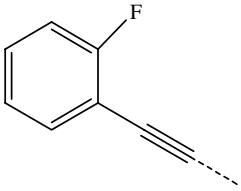
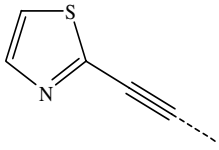
9		CH_3		3.70	3.50	0.21	1.0	7.22	7.03	7.02
10		CH_3		3.22	3.50	0.21	1.0	7.30	7.25	7.24
11		CH_3		2.24	3.50	0.21	1.0	7.45	7.49	7.49
12		CH_3		2.63	3.50	0.21	1.0	7.45	7.43	7.43
13		CH_3		2.94	3.50	0.21	1.0	7.70	7.35	7.33
14		CH_3		3.04	3.50	0.21	1.0	7.52	7.32	7.30
15		CH_3		3.34	3.50	0.21	1.0	6.85	7.20	7.22
16		CH_3		3.34	3.50	0.21	1.0	7.00	7.20	7.22

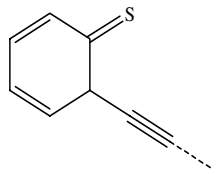
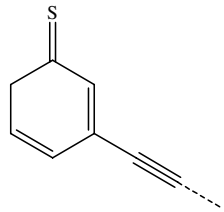
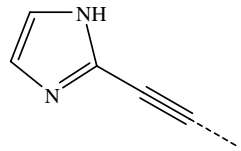
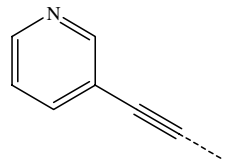
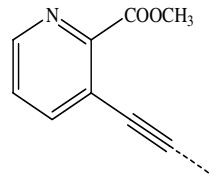
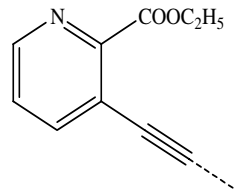
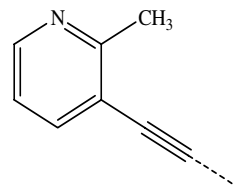
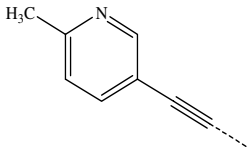
17		CH_3		3.01	3.50	0.21	1.0	7.52	7.33	7.32
18		CH_3		3.34	3.50	0.20	1.0	7.52	7.52	7.51
19		CH_3		3.69	3.50	0.20	1.0	7.52	7.35	7.34
20		CH_3		3.34	3.50	0.20	1.0	7.82	7.52	7.50
21		CH_3		2.03	3.50	0.20	1.0	7.52	7.83	7.85
22		CH_3		2.19	3.50	0.20	1.0	7.52	7.81	7.84
23		CH_3		2.19	3.50	0.20	1.0	7.82	7.81	7.81
24		CH_3		2.19	3.50	0.20	1.0	7.52	7.81	7.84
25	H	Cl		0.73	2.66	0.22	0.0	7.15	7.17	7.18

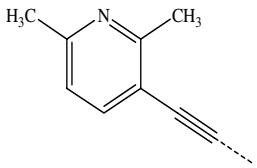
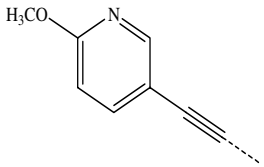
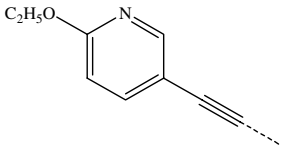
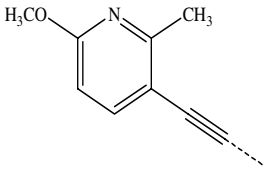
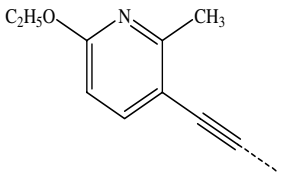
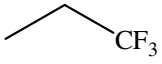
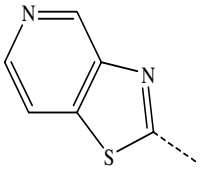
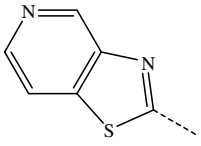
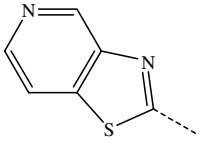
26	H	Cl		1.23	3.08	0.22	0.0	7.52	7.44	7.43
27	H	Cl		1.76	3.65	0.22	0.0	8.00	7.68	7.66
28	H	Cl		1.23	3.09	0.22	0.0	6.92	7.44	7.46
29	H	Cl		1.76	3.65	0.22	0.0	7.70	7.67	7.68
30	H	Cl		1.73	3.51	0.22	0.0	8.00	7.63	7.62
31	H	Cl		2.26	4.07	0.22	0.0	8.00	7.79	7.78
32	H	Cl		2.26	4.07	0.22	0.0	7.52	7.79	7.81
33	H	Cl		2.78	4.63	0.21	0.0	7.70	8.19	8.27

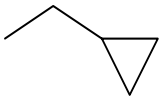
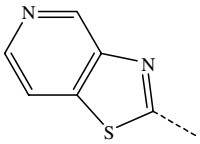
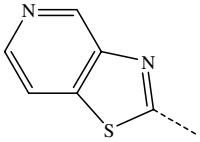
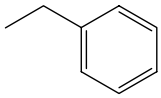
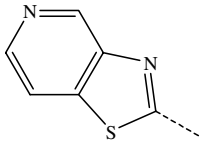
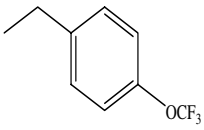
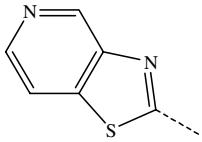
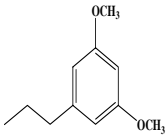
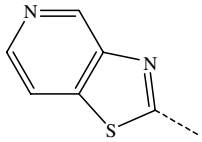
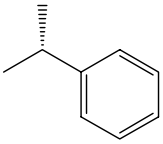
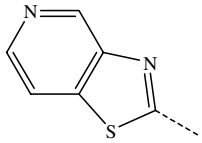
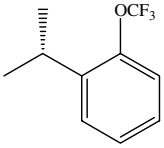
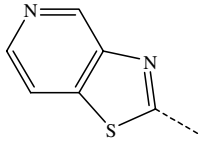
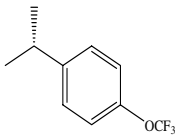
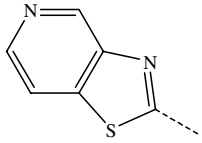
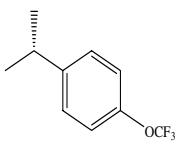
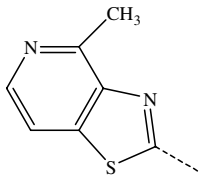
34	H	Cl		2.05	3.62	0.22	0.0	8.00	7.66	7.64
35	CH_3	Cl		2.87	3.62	0.21	0.0	7.70	7.83	7.84
36	CH_3	CH_3		2.26	3.65	0.22	0.0	7.52	7.65	7.66
37		CH_3		4.12	3.65	0.21	0.00	7.52	7.27	7.26
38		CH_3		2.65	3.65	0.20	0.00	8.40	8.22	8.20
39		CH_3		3.06	3.65	0.21	0.00	7.70	7.78	7.79
40		CH_3		3.72	3.65	0.19	0.0	8.52	8.13	8.11
41		CH_3		4.41	3.62	0.20	0.0	7.40	7.38	7.36

42		CH_3		4.01	3.62	0.18	0.0	8.00	8.27	8.30
43		CH_3		4.94	4.21	0.20	0.0	7.30	7.15	7.14
44		CH_3		3.08	4.21	0.20	0.0	8.04	8.28	8.30
45		CH_3		3.88	4.21	0.20	0.0	7.70	7.91	7.92
46	H	Cl		2.34	2.82	0.23	0.0	6.96	7.05	7.05
47	H	Cl		1.62	2.33	0.23	0.0	7.00	6.92	6.91
48	H	Cl		1.77	2.43	0.23	0.0	6.70	6.96	6.97
49	H	Cl		1.77	2.43	0.23	0.0	7.15	6.96	6.94
50	H	Cl		1.77	2.44	0.23	0.0	7.00	6.96	6.96

51	H	Cl		2.12	2.75	0.23	0.0	7.22	7.05	7.04
52	H	Cl		2.65	2.63	0.23	0.0	6.70	6.93	6.95
53	H	Cl		2.45	2.41	0.23	0.0	6.70	6.90	6.91
54	H	Cl		2.95	2.82	0.23	0.0	6.85	6.91	6.92
55	H	Cl		2.37	2.93	0.23	0.0	6.74	7.08	7.10
56	H	Cl		2.60	2.51	0.23	0.0	6.47	6.90	6.93
57	H	Cl		2.60	2.52	0.23	0.0	6.64	6.90	6.93
58	H	Cl		0.62	2.63	0.24	0.0	6.70	6.50	6.48

59	H	Cl		0.46	3.06	0.23	0.0	6.82	6.90	6.91
60	H	Cl		0.66	3.07	0.23	0.0	7.00	6.97	6.97
61	H	Cl		0.28	1.85	0.24	0.0	6.30	6.11	6.06
62	H	Cl		0.96	2.26	0.23	0.0	6.40	6.79	6.82
63	H	Cl		0.68	3.25	0.22	0.0	7.40	7.36	7.35
64	H	Cl		1.21	3.84	0.22	0.0	7.70	7.68	7.68
65	H	Cl		1.46	2.69	0.23	0.0	7.15	7.02	7.02
66	H	Cl		1.46	2.68	0.23	0.0	6.82	7.02	7.03

67	H	Cl		1.96	3.11	0.23	0.0	7.30	7.18	7.17
68	H	Cl		1.78	2.79	0.23	0.0	7.52	7.07	7.06
69	H	Cl		2.31	3.52	0.22	0.0	7.70	7.60	7.60
70	H	Cl		2.28	3.22	0.22	0.0	8.00	7.51	7.49
71	H	Cl		2.81	3.81	0.22	0.0	7.52	7.59	7.60
72		CH_3		1.93	3.12	0.21	0.0	7.79	7.82	7.81
73	H	CH_3		0.31	3.12	0.22	0.0	6.91	7.17	7.22
74	CH_3	CH_3		1.14	3.12	0.22	0.0	7.28	7.43	7.44

75		CH_3		2.11	3.12	0.21	0.0	8.10	7.81	7.80
76		CH_3		1.53	3.12	0.21	0.0	8.04	7.80	7.79
77		CH_3		2.59	3.12	0.20	0.0	8.52	8.05	8.02
78		CH_3		3.61	3.12	0.19	0.0	8.52	8.01	7.96
79		CH_3		3.24	3.12	0.19	0.0	7.92	8.17	8.19
80		CH_3		2.89	3.12	0.20	0.0	8.15	7.98	7.97
81		CH_3		3.92	3.12	0.19	0.0	7.77	7.84	7.84
82		CH_3		3.92	3.12	0.19	0.0	7.79	7.84	7.84
83		CH_3		4.42	3.77	0.19	0.0	7.79	7.73	7.72

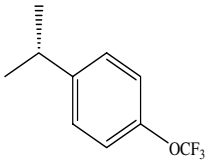
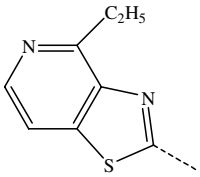
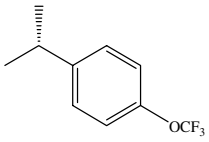
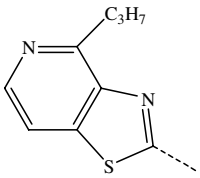
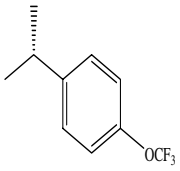
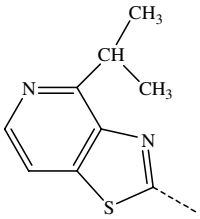
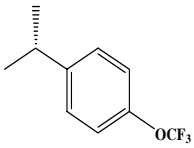
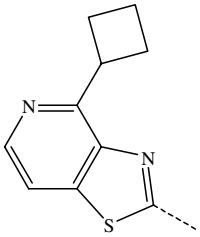
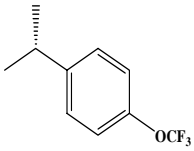
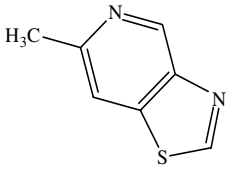
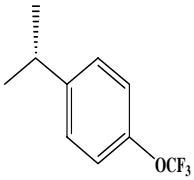
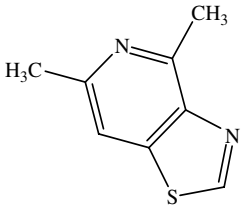
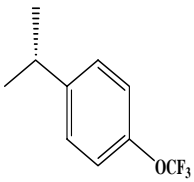
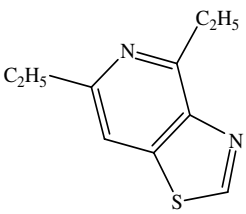
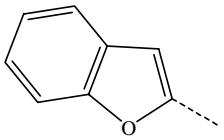
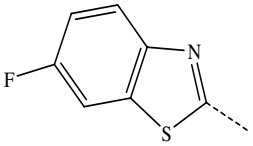
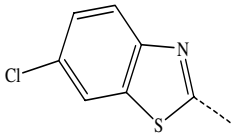
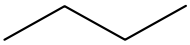
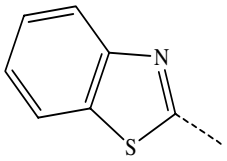
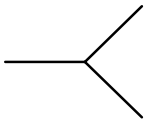
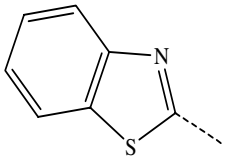
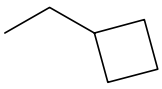
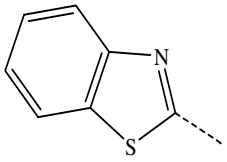
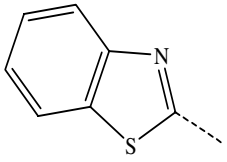
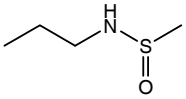
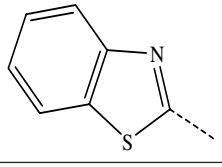
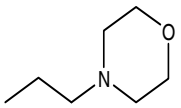
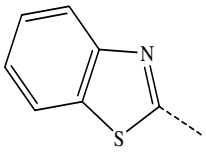
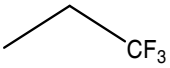
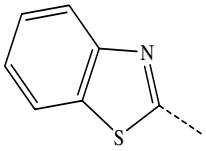
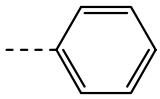
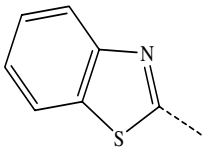
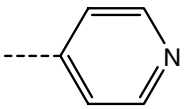
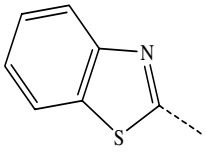
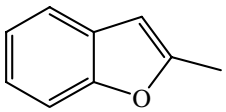
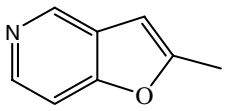
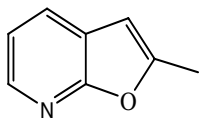
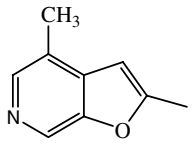
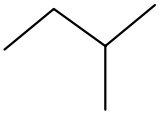
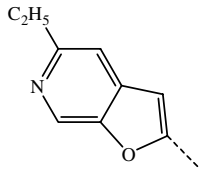
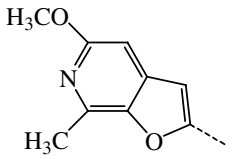
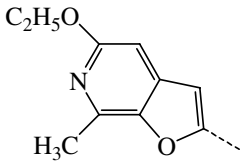
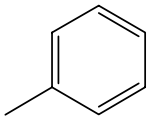
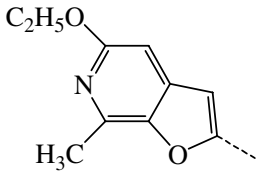
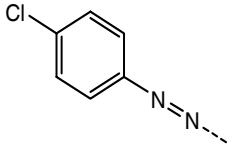
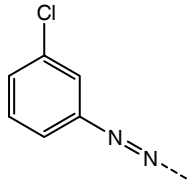
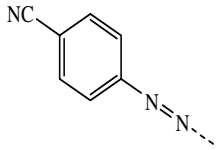
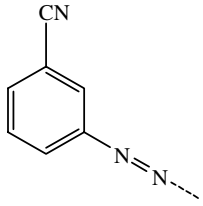
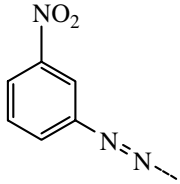
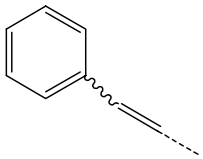
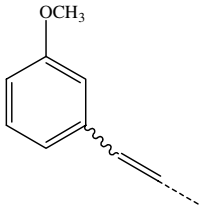
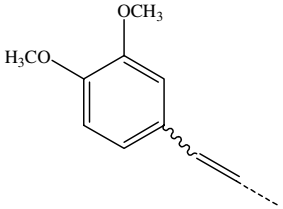
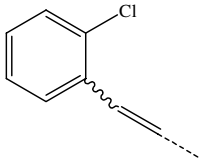
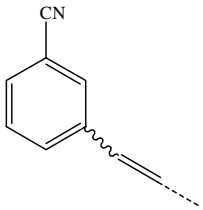
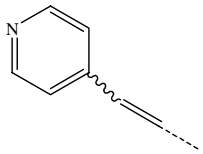
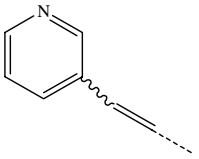
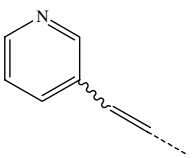
84		CH_3		4.95	4.33	0.18	0.0	7.70	7.82	7.82
85		CH_3		5.48	4.83	0.18	0.0	7.70	7.48	7.44
86		CH_3		5.35	4.72	0.18	0.0	7.58	7.57	7.56
87		CH_3		5.42	5.38	0.18	0.0	7.79	7.73	7.72
88		CH_3		4.42	3.77	0.19	0.0	7.42	7.73	7.76
89		CH_3		4.92	4.20	0.18	0.0	7.37	7.80	7.84
90		CH_3		5.98	5.32	0.18	0.0	7.16	7.10	7.09

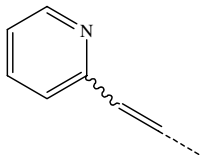
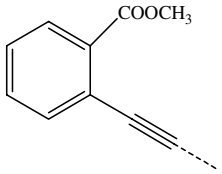
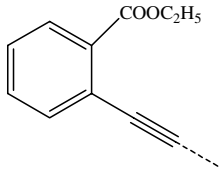
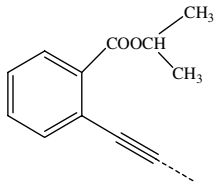
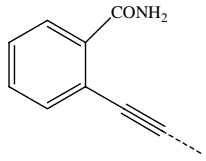
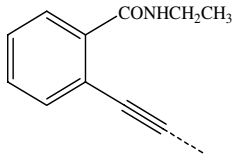
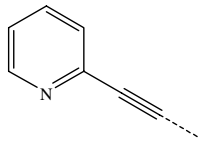
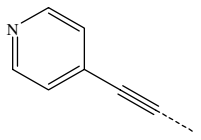
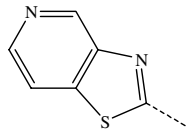
Table 2
Pyrimidine Derivatives (I as Shown in Materials and Methods) Belonging to Test Set and their Anti-HCV Activity and Physicochemical Parameters

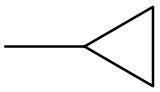
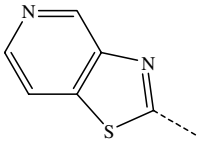
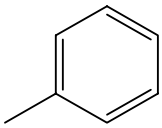
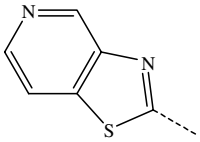
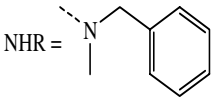
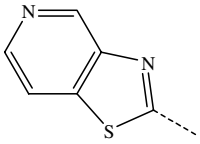
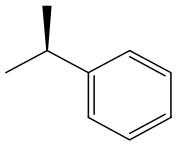
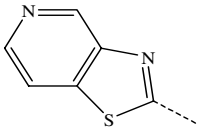
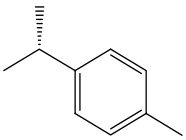
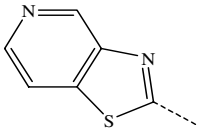
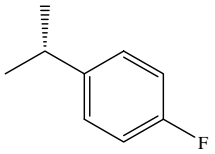
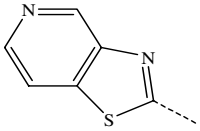
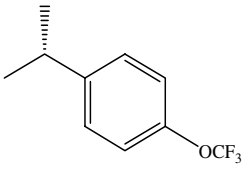
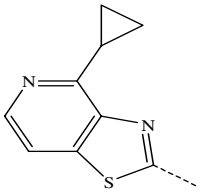
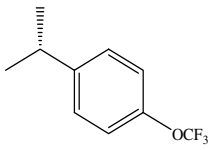
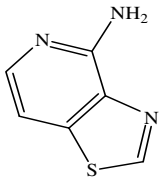
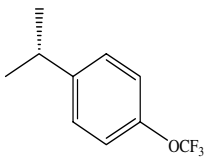
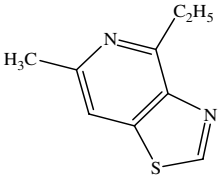
Compd	*R*	R_1	R_2	*ClogP*	$^1\chi^v_{R2}$	*Xt*	*I*	*log (1/EC$_{50}$)*	
								Obsd.	*Calc. (Eq.1)*
1	H	Cl		2.23	2.81	0.22	0.00	6.52	7.37
2	H	Cl		1.82	3.60	0.22	1.00	6.74	7.23
3	H	Cl		2.39	3.97	0.22	1.00	6.19	7.31
4		CH_3		3.30	3.50	0.21	1.00	6.51	7.21
5		CH_3		3.08	3.50	0.22	1.00	6.79	6.98
6		CH_3		3.77	3.50	0.20	1.00	6.60	7.31
7	NHR = $N(CH_3)_2$	CH_3		2.33	3.50	0.22	1.00	6.85	7.16
8		CH_3		1.60	3.50	0.20	1.00	6.58	7.82

9		CH_3		2.46	3.50	0.19	1.00	7.40	8.09
10		CH_3		3.04	3.50	0.21	1.00	6.10	7.31
11		CH_3		4.22	3.50	0.20	1.00	6.55	7.03
12		CH_3		2.83	3.50	0.20	1.00	6.48	7.69
13	H	Cl		2.23	2.81	0.22	0.00	6.52	7.37
14	H	Cl		0.73	2.66	0.22	0.00	5.77	7.17
15	H	Cl		0.73	2.67	0.22	0.00	6.40	7.18
16	H	Cl		1.23	3.08	0.22	0.00	5.85	7.43
17		CH_3		3.72	3.65	0.21	0.00	7.10	7.49

18	CH_3	CH_3		2.56	3.62	0.21	0.00	7.22	7.91
19	CH_3	CH_3		3.08	4.21	0.21	0.00	7.30	7.96
20		CH_3		5.06	4.21	0.19	0.00	6.82	7.35
21	H	Cl		2.34	2.81	0.23	0.00	6.52	7.04
22	H	Cl		2.34	2.81	0.23	0.00	6.49	7.04
23	H	Cl		1.06	2.72	0.23	0.00	5.92	6.97
24	H	Cl		1.06	2.72	0.23	0.00	6.22	6.97
25	H	Cl		1.37	2.83	0.22	0.00	6.04	7.37

26	H	Cl		2.51	2.53	0.23	0.00	6.04	6.92
27	H	Cl		2.43	3.06	0.23	0.00	6.22	7.11
28	H	Cl		2.17	3.58	0.22	0.00	6.92	7.63
29	H	Cl		3.22	3.01	0.23	0.00	6.52	6.88
30	H	Cl		1.94	2.92	0.23	0.00	6.22	7.11
31	H	Cl		1.01	2.38	0.23	0.00	5.47	6.84
32	H	Cl		1.01	2.38	0.23	0.00	6.00	6.84
33	H	Cl		1.01	2.38	0.23	0.00	5.40	6.84

34	H	Cl		1.01	2.39	0.23	0.00	5.64	6.84
35	H	Cl		2.42	3.39	0.22	0.00	6.89	7.54
36	H	Cl		2.95	3.98	0.22	0.00	7.00	7.61
37	H	Cl		3.26	4.38	0.22	0.00	6.52	7.63
38	H	Cl		0.97	3.07	0.22	0.00	6.70	7.38
39	H	Cl		1.70	4.09	0.22	0.00	6.70	7.82
40	H	Cl		0.96	2.27	0.23	0.00	6.10	6.79
41	H	Cl		0.96	2.26	0.23	0.00	6.04	6.79
42	$NHR = N(CH_3)_2$	CH_3		1.22	3.12	0.22	0.00	6.25	7.44

43		CH_3		1.49	3.12	0.21	0.00	6.90	7.80
44		CH_3		3.12	3.12	0.20	0.00	6.77	7.90
45		CH_3		2.99	3.12	0.20	0.00	6.52	7.94
46		CH_3		2.89	3.12	0.20	0.00	6.65	7.97
47		CH_3		3.39	3.12	0.19	0.00	6.94	8.11
48		CH_3		3.04	3.12	0.19	0.00	6.65	8.24
49		CH_3		4.87	4.88	0.18	0.00	7.22	8.05
50		CH_3		3.85	3.56	0.19	0.00	7.43	8.02
51		CH_3		5.45	4.76	0.18	0.00	6.58	7.48

the first requirement is to have 3D structure of the receptor or protein of interest which can be determined by X-ray crystallography or NMR spectroscopy. This protein structure and a 3D database of potential ligands serve as input to a docking program. The success of a docking program depends on two components, viz., the search algorithm and the scoring function.

Docking simulations. Molegro Virtual Docker (MVD) (*http:// www.molegro.com*; Thomsen and Christen, 2006) (trial version) was used for flexible ligand docking wherein the software makes use of differential evolution algorithm (Yang and Chen, 2004). Fast and accurate identification of potential binding modes during the search process is made by the use of predicted cavities. The scoring function makes use of piecewise linear potential (PLP) (Thomsen, 2003). It takes into account hydrogen bonding terms along with their directionality, ligand-protein interaction energy, and intramolecular interaction energy of the ligand. For enhanced docking accuracy, the highest ranked poses are yet again re-ranked (*www.rcsb.org*).

Results and Discussion

QSAR Results

When a multiple regression analysis was performed on the training set, it revealed the following correlation.

$$\log(1/EC_{50}) = 0.514(\pm0.165)ClogP - 0.140(\pm0.028)(ClogP)^2 + 0.332(\pm0.112)^1\chi^v_{R2} - 31.554(\pm5.835)Xt - 0.426(\pm0.132)I + 12.933(\pm1.487)$$

$n = 90$, $r = 0.871$, $r^2_{cv} = 0.768$, $r^2_{pred} = 0.648$, $s = 0.24$, $F_{5,84} = 52.63(3.24)$, $(ClogP)_o = 1.84$ (1)

In this equation, n is the number of data points, r is the correlation coefficient, r^2_{cv} is the square of cross-validated correlation coefficient obtained from leave-one-out (LOO) jackknife procedure, r^2_{pred} is the square of correlation coefficient showing the predictive ability of the correlation, s is the standard deviation, data within the parentheses with ± sign are 95% confidence intervals, and the F is F-ratio between the variances of calculated and observed activities. The figure within the parenthesis for F is the standard F-value at 99% level. Equation 1 represents a highly significant correlation between the HCV inhibition potency of the compounds and the physicochemical parameters. In this equation, the positive coefficients of hydrophobicity (ClogP) and valence connectivity index ($^1\chi^v_{R2}$) suggest that the activity of the compounds will increase as the values of these two parameters increase. However, since the correlation is parabolic in ClogP, CloP has an optimum value equal to 1.84, suggesting that the activity will decrease after this value of ClogP. It can be therefore assumed that the molecule may have some hydrophobic interaction with the receptor but very bulky molecule for which ClogP may be quite high may not be conducive to the activity. This is also corroborated by the negative coefficient of structural connectivity index Xt of the molecule which is purely indicative of the size of the molecule. Since, as shown in Table 3 there exists no mutual correlation between ClogP and Xt, both the variables play an independent role.

A negative coefficient of the indicator variable 'I' also suggested that an R_2-substituent being or containing a benzothiazole ring will also not be favorable to the activity. This negative role of such substituent may be due to any repulsive electronic role of this group with any charged electronic site of the receptor.

All the variables used in equation 1 were found to have no mutual correlation (Table 3) and each one of them was found to be statistically quite significant, as an appreciable drop in the overall significance in the successive equation was observed when they were dropped one by one (equations 2-5).

$$\log(1/EC_{50}) = 0.325(\pm0.188)ClogP - 0.101(\pm0.031)(ClogP)^2 + 0.238(\pm0.131)^1\chi^v_{R2} - 30.015(\pm7.066)Xt + 12.988(1.806)$$

$n = 90$, $r = 0.799$, $s = 0.296$, $F_{4,85} = 37.52(3.55)$ (2)

$$\log(1/EC_{50}) = 0.468(\pm0.249)ClogP - 0.082(\pm0.042)(ClogP)^2 + 0.397(\pm0.170)^1\chi^v_{R2} + 5.551(\pm0.598)$$

$n = 90$, $r = 0.579$, $s = 0.398$, $F_{3,86} = 14.44(4.02)$ (3)

$$\log(1/EC_{50}) = 0.486(\pm0.277)ClogP - 0.063(\pm0.046)(ClogP)^2 + 6.698(\pm0.382)$$

$n = 90$, $r = 0.409$, $s = 0.443$, $F_{2,87} = 8.73(4.85)$ (4)

$$\log(1/EC_{50})=0.119(\pm0.078)ClogP+7.130(\pm0.224)$$

$$n = 90,\ r = 0.310,\ s = 0.459,\ F_{1,88} = 9.34(6.93) \quad (5)$$

Thus the correlation expressed by equation 1 seems to be highly significant and its internal and external validations can be judged by r^2_{cv} and r^2_{pred} values, which are 0.768 and 0.648, respectively. The r^2_{cv} is calculated as follows.

$$r^2_{cv} = 1- [\Sigma_i(Y_{i,obsd} - Y_{i,pred})^2/ \Sigma_i(Y_{i,obsd} - Y_{av,obsd})^2] \quad (6)$$

where $Y_{i,obsd}$ and $Y_{i,pred}$ are the observed and predicted (from LOO) activity values of compound *i*, respectively, and $Y_{av,obsd}$ the average of the observed activities of all compounds used in the correlation. Similarly, the r^2_{pred} is calculated as

$$r^2_{pred} = 1- [\Sigma_i(Y_{i,obsd} - Y_{i,pred})^2/ \Sigma_i(Y_{i,obsd} - Y_{av,obsd})^2] \quad (7)$$

where $Y_{i,obsd}$ is the observed activity of compound *i* in the test set and $Y_{i,pred}$ is its activity predicted from equation 1. $Y_{av,obsd}$ is the same as in equation 6.

The activity values predicted from equation 1 for test set compounds are displayed in Table 2. A comparison shows that these predicted values are in fairly good agreement with the corresponding observed ones. In the training set also the predicted values were found to be in good agreement with the observed ones (Table 1). All these observations can be better visualized in the graph drawn between the predicted and observed activities for training as well as test sets (Figure 1). Using equation 1, we predicted some new compounds as shown in Table 4. The activities of these compounds are higher than any compound in the existing series (Tables 1 and 2).

Docking Results and Validation of Docking Method

For the present studies, we have selected the ligand 1BI_601 [B] (PDB code: 4GMC). The X-ray crystal structure of HCV NS5B polymerase in complex with a thumb inhibitor (4GMC) refined at 2.70 Å resolution was considered for the docking studies. The ligand was extracted from the complex (4GMC) and redocked using flexible docking simulations into the original structure of HCV NS5B polymerase. The protein complex of HCV NS5B polymerase with a thumb inhibitor (4GMC) was imported from the Protein Data Bank (*http:// www.molegro.com*). The scoring function MolDock Grid with 0.30 Å resolution along with an algorithm MolDock optimizer was used for docking. The essential parameters, i.e., number of runs = 10, population size = 50, maximum iterations = 2000, and termination scheme = variance based, were fixed. Docking results have been shown for all the predicted compounds in Table 5. The hydrogen-bond interactions, hydrophobic interactions, and electrostatic interactions have been shown in Figures 2-4, respectively, for one of the predicted compounds that has the highest number of H-bonds (Compound 5, Table 5). The docking study showed that all predicted compounds have better energies of interactions with the enzyme and docking scores than the original ligand complexed with the enzyme. In Table 5, it is observed that while ligand complexed with the enzyme forms no H-bond, all predicted compounds, except 6, form 1-4 H-bonds, in which mostly Glu (70), Ala (73), Lys (74), Val (186), Tyr (74) and Thr (77) residues of the enzymes are involved. Compound 5 with the highest number of H-bonds (4) forms the H-bonds with its NH that is present between pyrimidine and cyclopentyl rings and two of the hydroxyl groups present at the cyclopentyl ring.

In Figure 3 that shows the hydrophobic interaction of compound 5, it is indicated that except the pyrimidine ring, all other fragments of the molecule are present in strong hydrophobic region of the enzyme and thus have strong hydrophobic interaction. On the other hand, in Figure 4 that shows the electrostatic interaction of compound 5 with the enzyme, it is indicated that no portion of the molecule is embedded in strong electrostatic region of the enzyme and thus there is little electrostatic interaction between the molecule and the enzyme. These observations are in agreement to the findings of 2D QSAR which suggested that molecules involve dominantly only hydrophobic interactions with the receptor and that the bulky molecule may face steric hindrance.

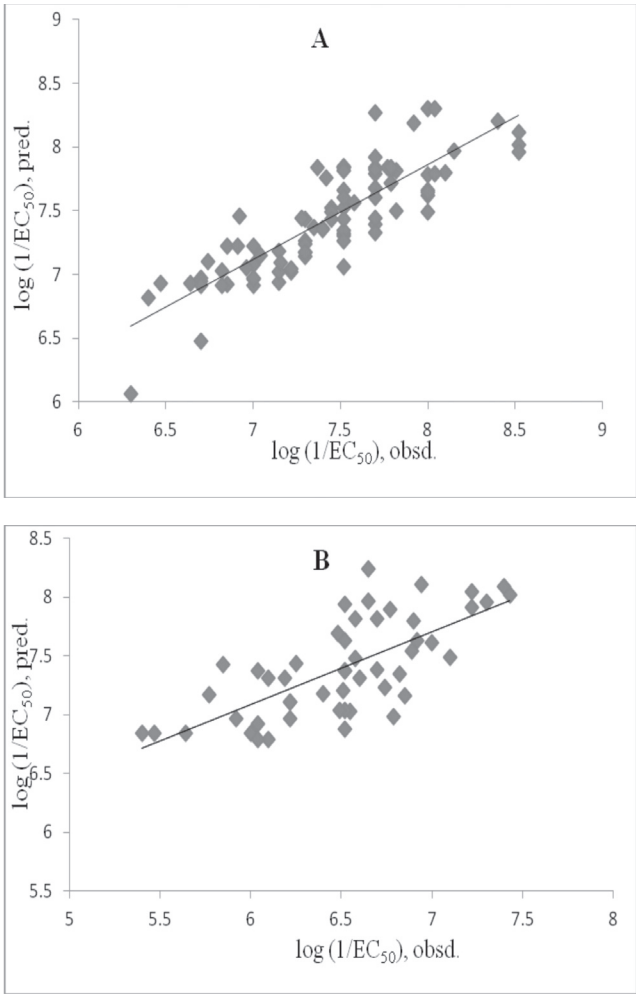


Figure 1: **Plots between observed and predicted activities for pyrimidine derivatives: A, for training set; B, for test set**

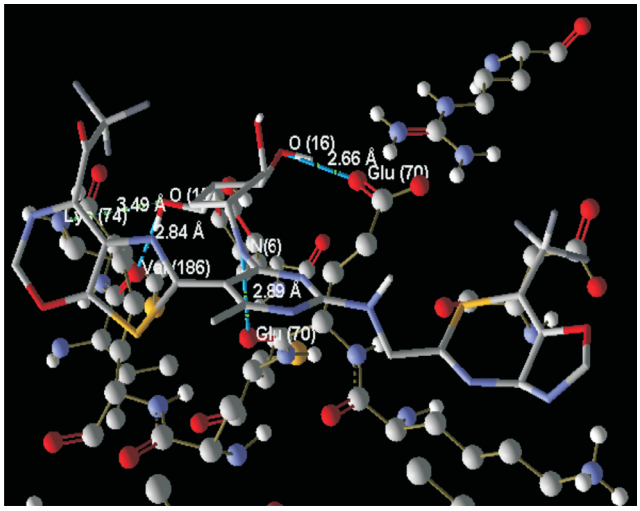


Figure 2: **A model showing hydrogen bond interactions of predicted compound 5 with the enzyme HCV Replicase. Compound 5 is one of the predicted compounds that has the highest number of H-bonds**

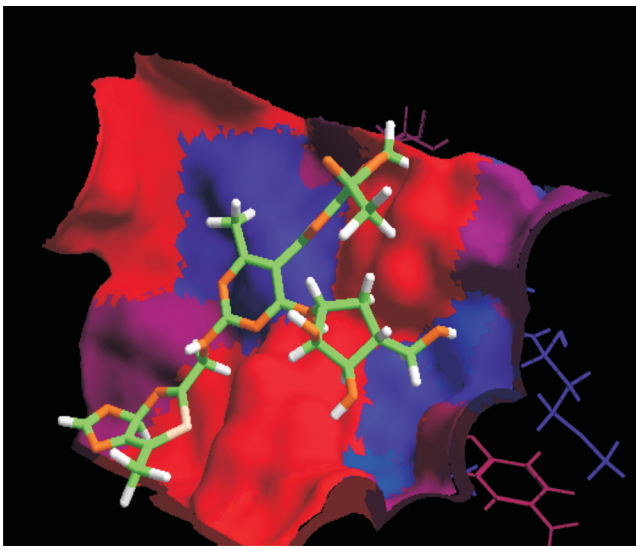


Figure 3: **The model showing hydrophobic interactions of predicted compound 5 with the enzyme HCV Replicase. The red surface shows strong hydrophobic zone and blue one the low hydrophobic zone**

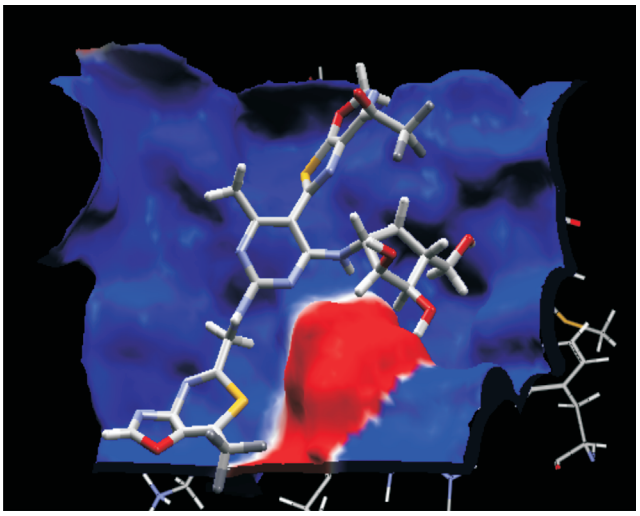
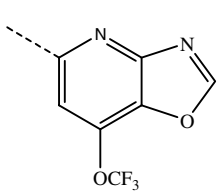
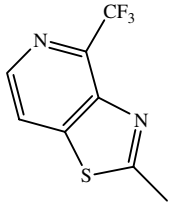
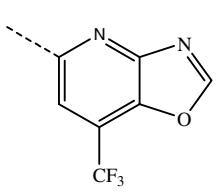
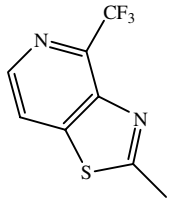
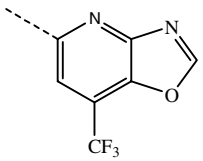
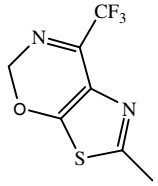
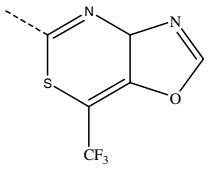
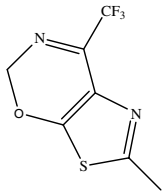
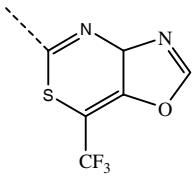
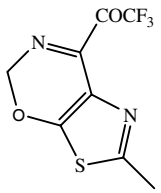
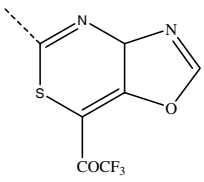
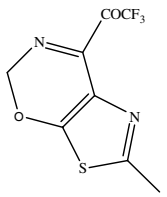
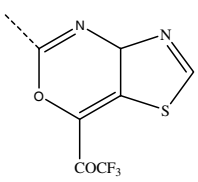
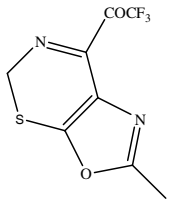


Figure 4: **The model showing electrostatic interactions of predicted compound 5 with the enzyme HCV Replicase. The blue surface shows strong electrostatic zone and red one the low electrostatic zone**

Table 3
Correlation Matrix Showing Mutual Correlations among the Variables used in Equation 1

	Clog P	*(Clog P)²*	$^1\chi^v$	*Xt*	*I*
ClogP	1.000	–0.902	0.113	0.139	– 0.352
$(ClogP)^2$		1.000	–0.242	0.158	0.420
$^1\chi^v_{R2}$			1.000	0.254	– 0.261
Xt				1.000	0.081
I					1.000

Table 4
Some Proposed Compounds Belonging to the Series of Pyrimidine Derivatives and their Predicted Activities

Compd	*R*	R_2	*ClogP*	$^1\chi^v_{R2}$	*Xt*	*log(1/EC$_{50}$) Cacld, Eq. 1*
1			3.35	4.095	0.175	9.25
2			2.63	4.095	0.176	9.13
3			1.59	4.024	0.176	9.18
4			2.82	4.024	0.176	9.06
5			3.08	4.48	0.173	9.22
6			1.15	4.48	0.17	9.43
7			1.04	4.65	0.17	9.46

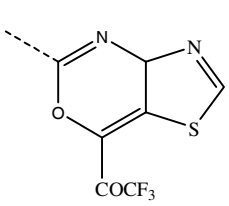
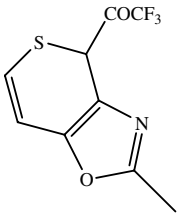
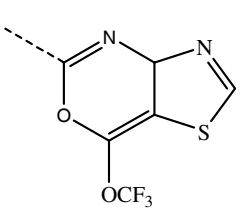
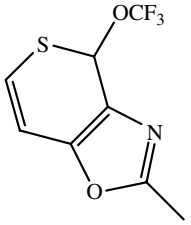
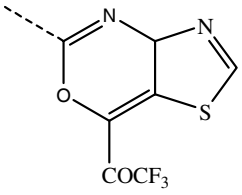
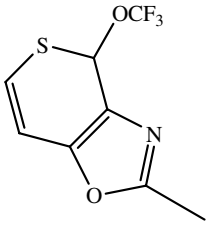
8			1.43	4.75	0.17	9.57
9			3.30	4.45	0.173	9.12
10			2.42	4.45	0.172	9.40

Table 5
Docking Results of Proposed Compounds with Reference to Compound Available in Complex with Protein

Predicted Compnd (Table 4)	*Overall Interac E (kJ/mol)*	*H -bond Energy (kJ/mol)*	*No of H-bonds*	*H-bonds Compd-Protein Atoms*	*H-bond Length(Å)*	*Mole Dock Score*	*Internal E of Pose (kJ/mol)*
Ligand	-119.444	0.000	0	-	-	-146.167	-26.723
1	-118.448	-5.00	2	O (31) – Glu (70)	3.52	-125.630	-7.181
				O (38) – Glu (70)	3.17		
2	-129.013	-2.012	13	N (12) – Glu (70)	3.06	-133.265	-4.252
3	-127.610	-5.307	3	N (12) – Glu (70)	3.17	-131.330	-3.721
				O (34) - Tyr (74)	2.72		
				O (16) - Thr (77)	3.30		
4	-120.240	-4.688	25	O (15) - Lys (69)	3.16	-124.706	-4.466
				N (24) - Lys (74)	3.05		
5	-133.002	-8.029	4	O (16) – Glu (70)	2.66	-133.477	-0.475
				N (6) – Glu (70)	2.89		
				O (15) - Val (186)	2.84		
				O (15) – Lys (74)	3.49		
6	-104.970	0.000	0	-	-	-130.950	-25.980
7	-114.390	-0.222	1	N (6) - Ala (73)	3.46	-132.085	-17.696
8	-110.691	-5.482	3	N (12) - Ala (73)	2.79	-147.859	-37.168
				O (17) - Ala (73)	2.41		
				O (15) - Lys (74)	2.68		
9	-147.938	-5.00	2	O (38) - Lys (74)	2.93	-147.769	0.169
				O (16) - Glu (70)	3.03		
10	-127.709	-2.5	1	O (17) - Lys (74)	2.94	-151.781	-24.072

Conclusion

The HCV inhibition potency of pyrimidines is found to be controlled by hydrophobicity, Kier's first-order valence connectivity index of a substituent, total structure connectivity index of the molecule and one indicator parameter used for the presence of benzothiazole ring. Of all these parameters, the positive effect has been found to be produced by only two parameters, hydrophobicity and valence connectivity index. Based on QSAR equation, some new compounds with higher activity have been predicted. The docking study showed that these predicted compounds have better energies of interactions with the enzyme and docking scores than the original ligand complexed with the enzyme.

Abbreviations

HCV, Hepatitis C Virus; QSAR, Quantitative Structure-Activity Relationship; RNA, Ribonucleic Acid; INF-α, Interferon-α; EIA, Enzyme Immunoassay; RIBA, Recombinant Immunoblot Assay; ClogP, Calculated hydrophobic paramer; $^{1}\chi^{v}$, Kier's first-order valence molecular connectivity index; Xt, Total structure connectivity index; LOO, Leave-one-out; F, Fischer ratio; MVD, Molegro Virtual Docker; PLP, Piecewise Linear Potential; Compd, Compound

References

Arasappan, A., Bennett, F., Girijavallabhan, V., Huang, Y., Huelgas, R., Alvarez, C., Chen, L., Gavalas, S., Kim, S.-H., Kosinski, A., Pinto, P., Rizvi, R., Rossman, R., Shankar, B., Tong, L., Velazquez, F., Venkatraman, S., Verma, V. A., Kozlowski, J., Shih, N. Y., *et al.* (2012). 5-Benzothiazole substituted pyrimidine derivatives as HCV replication (replicase) inhibitors. Bioorg. Med. Chem. Lett. *22*, 3229-3234.

Belda, O. and Targett-Adams, P. (2012). Small molecule inhibitors of the hepatitis C virus-encoded NS5A protein. Virus Res. *170*, 1–14.

Bennett, F., Kezar III, H. S., Girijavallabhan, V., Huang, Y., Huelgas, R., Rossman, R., Shih, N.-Y., Piwinski, J. J., MacCoss, M., Kwong, C. D., Clark, J. L., Fowler, A. T., Geng, F., Roychowdhury, A., Reynolds, R. C., Maddry, J. A., Ananthan, S., Secrist III, J. A., Li, C., Chase, R., *et al.* (2012). Pyridofuran substituted pyrimidine derivatives as HCV replication (replicase) inhibitors. Bioorg. Med. Chem. Lett. *22*, 5144-5149.

Chinnaswamy, S., Cai, H. and Kao, C. (2010). An update on small molecule inhibitors of the HCV NS5B polymerase: effects on RNA synthesis *in vitro* and in cultured cells, and potential resistance in viral quasispecies. Virus Adapt. Treat. *2*, 73–89.

Girijavallabhan, V. M., Alvarez, C., Bennett, F., Chen, L., Gavalas, S., Huang, Y., Kim, S.-H., Kosinski, A., Pinto, P., Rizvi, R., Rossman, R., Shankar, B., Tong, L., Velazquez, F., Venkatraman, S., Verma, V. A., Kozlowski, J., Shih, N.-Y., Piwinski, J. J., MacCoss, M., *et al.* (2012). Synthesis and SAR of pyridothiazole substituted pyrimidine derived HCV replication inhibitors. Bioorg. Med. Chem. Lett. *22*, 5652-5657.

Gupta, S.P. (2011). QSAR and Molecular Modeling. Anamaya Publishers, New Delhi.

Ishida, T., Suzuki, T., Hirashima, S., Mizutani, K., Yoshida, A., Ando, I., Ikeda, S., Adachi, T. and Hashimoto, H. (2006). Benzimidazole inhibitors of hepatitis C virus NS5B polymerase: identification of 2-[(4-diarylmethoxy) phenyl]-benzimidazole. Bioorg. Med. Chem. Lett. *16*, 1859–1863.

Kwong, C. D., Clark, J. L., Fowler, A. T., Geng, F., Kezar III, H. S., Roychowdhury, A., Reynolds, R. C., Maddry, J. A., Ananthan, S., Secrist III, J. A., Shih, N. Y., Piwinski , J. J., Li, C., Feld, B., Huang, H. C., Tong, X., Njoroge, F. G. and Arasappan. A. (2012). Novel substituted pyrimidines as HCV replication (replicase) inhibitors. Bioorg. Med. Chem. Lett. *22*, 1160-1164.

Labrador L. and Xavier F. (2008). Hepatitis C virus NS3/4A protease inhibitors. Recent Patents on Anti-Infective Drug Discovery *3*, 157-167.

Li, X., Zhang, Y.-K., Liu, Y., Ding, C. Z., Li, Q., Zhou, Y., Plattner, J. J., Baker, S. J., Qian, X., Fan, D., Liao, L., Ni, Z.-J., White, G. V., Mordaunt, J. E., Lazarides, L. X., Slater, M. J., Jarvest, R. L., Thommes, P., Ellis, M., Edge, C. M., et al. (2010). Synthesis and evaluation of novel a-amino cyclic boronates as inhibitors of HCV NS3 protease. Bioorg. Med. Chem. Lett. *20*, 3550–3556.

Paulson, M. S., Yang, H., Shih, I.-H., Feng, J. Y., Mabery, E. M., Robinson, M. F., Zhong, W. and Delaney IV, W. E. (2009). Comparison of HCV NS3 protease and NS5B polymerase inhibitor activity in 1a, 1b and 2a replicons and 2a infectious virus. Antivir. Res. *83*, 135-142.

Pfefferkorn, J. A., Nugent, R., Gross, R. J., Greene, M., Mitchell, M. A., Reding, M. T., Funk, L. A., Anderson, R., Wells, P. A., Shelly, J. A., Anstadt, R., Finzel, B. C., Harris, M. S., Kilkuskie, R. E., Kopta, L. A. and Schwende, F. J. (2005). Inhibitors of HCV NS5B polymerase. Part 2: Evaluation of the northern region of (2Z)-2-benzoylamino-3-(4-phenoxy-phenyl)-acrylic acid. Bioorg. Med. Chem. Lett. *15*, 2812–2818.

Simmonds, P., Holmes, E. C., Cha, T.-A., Chan, S.-W., McOmish, F., Irvine, B., Beall, E., Yap, P. L., Kolberg, J. and Urdea, M. S. (1993) Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. J. Gen. Virol. *74*, 2391-2399.

Thomsen, R. (2003). Flexible ligand docking using evolutionary algorithms: investigating these effects of variation operators and local search hybrids. Biosystems *72*, 57-73.

Thomsen, R. and Christen, M. H. (2006). MolDock. A new technique for high accuracy molecuar docking. J. Med. Chem. *49*, 3315-3321.

Wei , W., Cai, C., Kota, S., Takahashi, V., Ni, F., Strosberg , A. D. and Snyder, J. K. (2009). New small molecule inhibitors of hepatitis C virus. Bioorg. Med. Chem. Lett. *19*, 6926–6930.

Yang, J. M. and Chen, C. C. (2004). GEMDOCK: A generic evolutionary method for molecular docking. Proteins *55*, 288-304.

Zuo, G., Li, Z., Chen, L. and Xu, X. (2007). Activity of compounds from Chinese herbal medicine *Rhodiola kirilowii* (Regel) maxim against HCV NS3 serine protease. Antivir. Res. *76*, 86–92.