Available online at www.derpharmachemica.com

ISSN 0975-413X
CODEN (USA): PCHHAX

Der Pharma Chemica, 2016, 8(11):45-61
(http://derpharmachemica.com/archive.html)

A Quantitative Structure-Activity Relationship Study on a Indo-2-yl Derivatives as Anti-HCV agents

Anjana Sharma[1], Satya Prakash Gupta[2], Anees Ahmed Siddiqui[3] and Nitin Sharma[1]

[1]*Department of Pharmaceutical Technology, Meerut Institute of Engineering and Technology, Meerut 250005, India*
[2]*Department of Applied Sciences, National Institute of Technical Teacher's Training and Research (NITTTR), Bhopal 462002, India*
[3]*Faculty of Pharmacy, Jamia Hamdard, New Delhi 110062, India*

ABSTRACT

A quantitative structure-activity relationship (QSAR) study has been performed on some large series of (Indol-2-yl)phenyl/pyridine sulfonamides derivatives reported by Chen et al. [Bioorg Med Chem Lett(2013)] ***[9,10]****, acting as anti-hepatitis C virus (anti-HCV) agents. The activity of the compounds was found to be correlated significantly with density, surface tension and some indicator variables. The whole series containing 73 compounds was divided into two subsets: training set and test set containing 53 and 20 compounds, respectively. The validity of correlation has been judged by LOO (leave one out) method and predicting the activity of some test compounds. Using the correlation obtained, some new compounds having high potency have been predicted.*

Keywords: QSAR study, anti-HCV agents, Hepatitis C virus inhibitors, (Indo-2-yl) phenyl /sulphonamide.

INTRODUCTION

Hepatitis C virus (HCV) infection is widespread, underdiagnosed, and drastically undertreated. Hepatitis can be caused by any of the six hepatotropic viruses viz hepatitis A virus, hepatitis B virus, hepatitis C virus, hepatitis D virus, hepatitis E virus, hepatitis G virus, representating a major health problem worldwide. Globally, HCV infection is a major cause of acute hepatitis and chronic liver disease **[1]**. Hepatitis C virus is a single stranded RNA virus belonging to *Flaviviridae* family. Its genome consists of structural proteins, envelope proteins, and non-structural (NS) proteins **[2]**. About 3% of the world's population is chronically infected by HCV and the most common cause for the need of liver transplantation **[3]**. About 80% of the acute infection become chronic that finally leads to liver cirrhosis and hepatocellular carcinoma **[4]**.

Existing therapies are based on combinations of pegylated interferon's and ribavirin which provide a sustained response only in fraction of patients and side effects are quiet serious **[5]**. Ribavirin (purine analog broad spectrum antiviral) alone is not effective against HCV, but when combined with peginterferon(affecting the immune response) the rate of viral eradication gets increased two to three times more than interferon's **[6]**. As, there is still not effective, well tolerated medication available for HCV infection so, there is an urgent need to design and discover specific antiviral agents that lead to improved efficacy, tolerability, and compliance to meet the medical need **[7,8]**. Quantitative structure activity relationship (QSAR) studies have been of great importance for design and development of potent drugs. These studies not only provide aid in the rationalization of drug development but also guidelines for drug design. We, therefore present here a QSAR studies on some Indolyl derivatives.

MATERIALS AND METHODS

Methodology

For our study we have taken a series of compounds from the two consecutive papers of Chen *et al* [9, 10] as shown in Table **1**. The activity of each compound was given in terms EC_{50} that refers to molar concentration of the compound leading inhibition to 50% inhibition of HCV replication. Table **1** contains 73 compounds in total, along with their biological activity, relevant physicochemical parameters, and some indicator variables that were used to describe the specific role of some substituents in the molecules which are defined later.

Cross-validation

The cross-validation evaluates the validity of a model by how well it predicts data rather than how well it fits the activity data. The analysis uses a Leave-one-out (LOO), Jackknife scheme to calculate r^2_{cv} using the equation:

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

where, $Y_{i,obsd}$ and $Y_{i,pred}$ are the observed and predicted activity values, respectively of compound *i* and $Y_{av,\ obsd}$ is the average of the observe activity of all compound used in the correlation. The model is build with the n-1 compound by removing the compound from the data set each time. By using this model the activity of the compound removed is predicted. This procedure is repeated till all the compounds are removed one by one and their activities are predicted from the new model formed each time. The correlation is supposed to be valid, if r^2_{cv} is greater than 0.60 that establishes the internal validation of the model. However, the external validation (predictive ability) of any model is measure by r^2_{pred}, also has a good value greater than 0.50. The predictive ability of any correlation equation is measured by using it to predict the activity of the compound and the test set and calculating the value of r^2_{pred} by the using the equation.

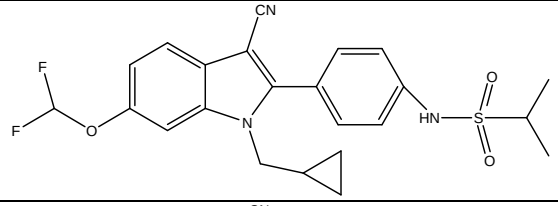
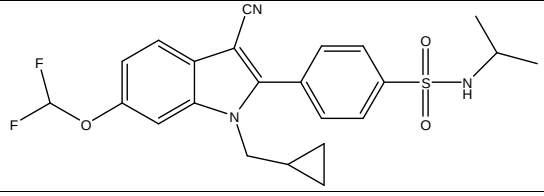
$$r^2_{pred} = 1 - [\ \Sigma_i(Y_{i,obsd}-Y_{i,pred})^2 / \ \Sigma_i(Y_{i,obsd}-Y_{av,obsd})^2] \quad (2)$$

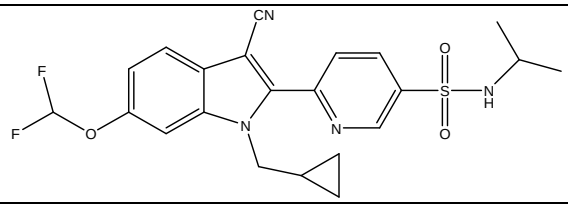
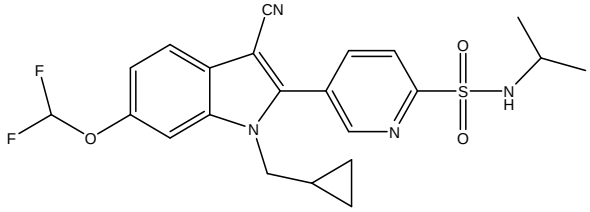
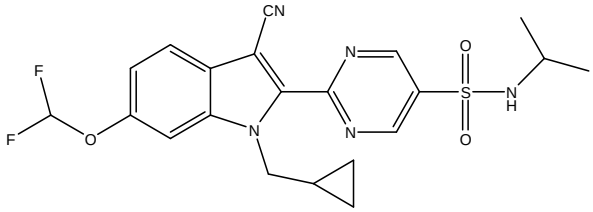
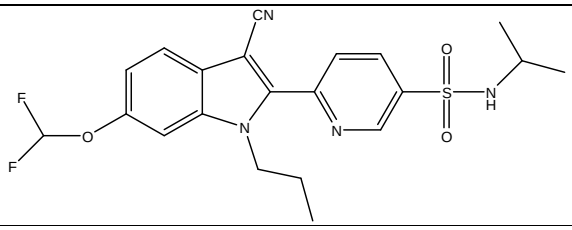
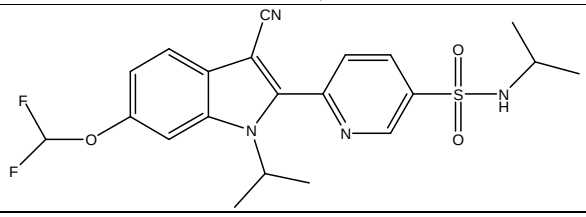
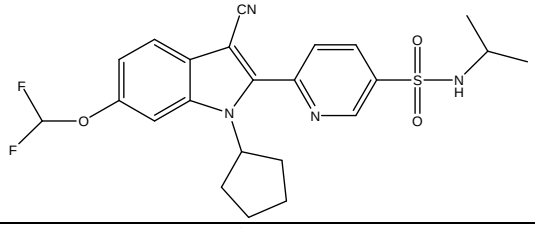
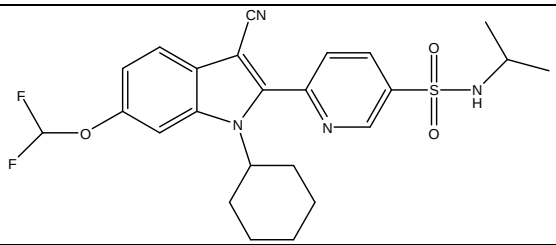
where $Y_{i,obsd}$ is the observed activity of compound *i* in the test set and $Y_{i,pred}$ is its activity predicted from Eq. (3). $Y_{av,obsd}$ is same as in Eq. (1).

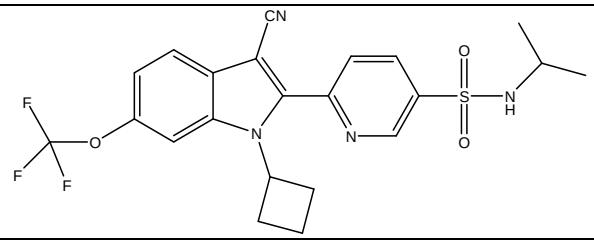
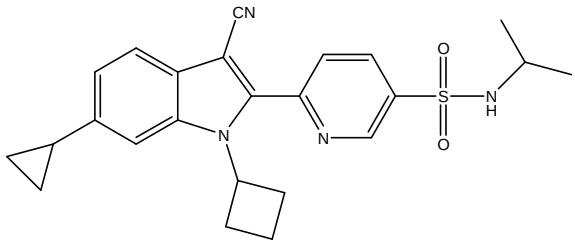
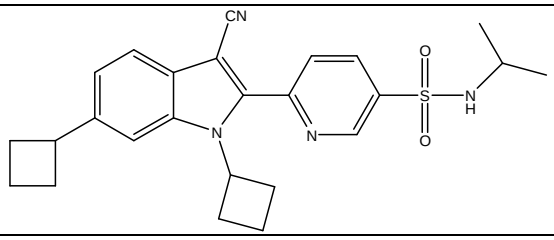
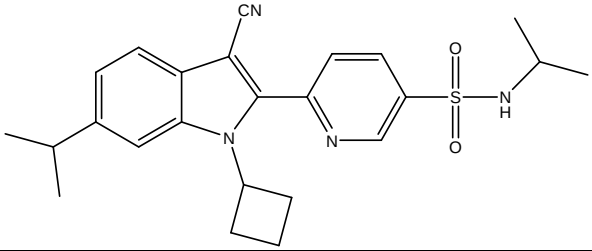
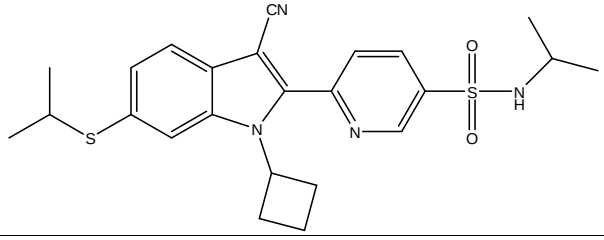
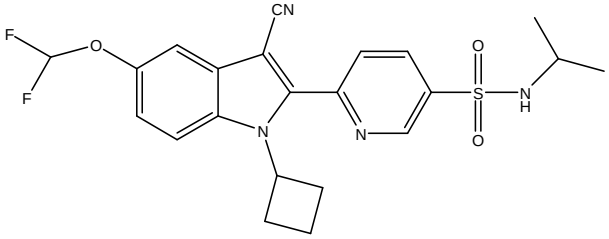
RESULTS AND DISCUSSION

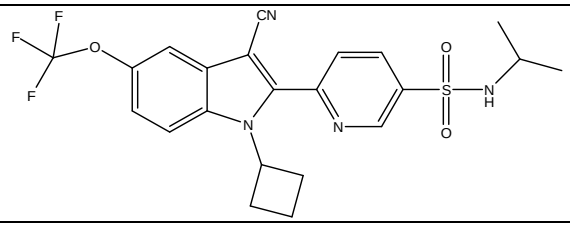
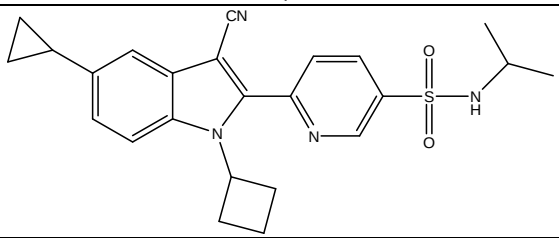
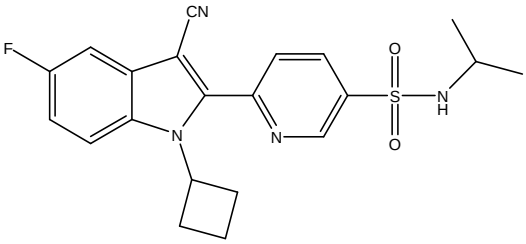
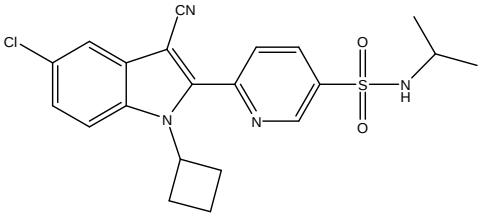
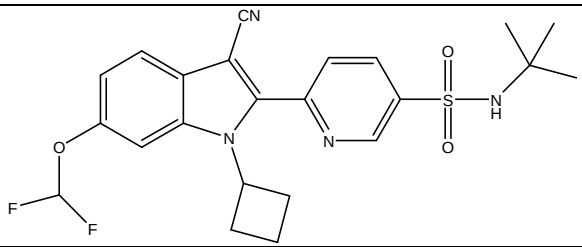
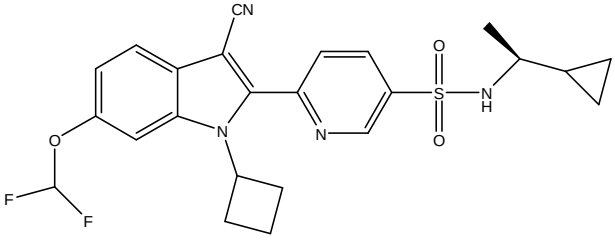
All the compounds of Table **1** were divided into two subsets: the training set and test set. The training and test sets consisted of 53 and 20 respectively. Compounds for test were selected arbitrarily by keeping in mind the wide variation in their structures as well in their activity. All the test compounds are given with superscript 'b' in Table **1**. The remaining compounds were taken for training set.

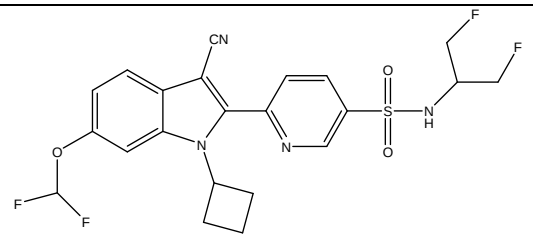
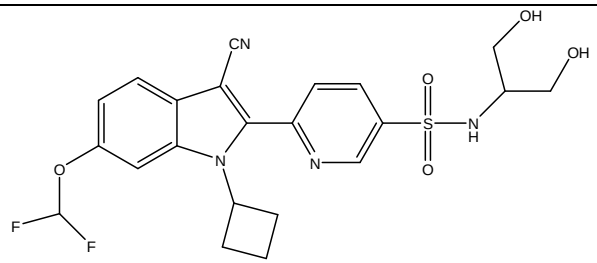
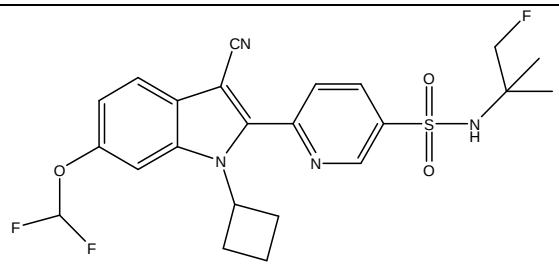
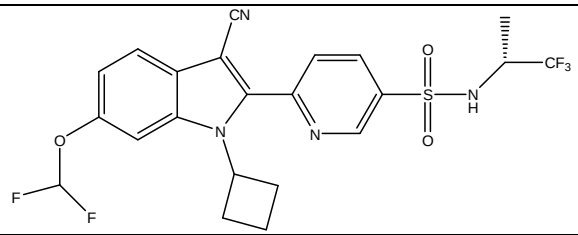
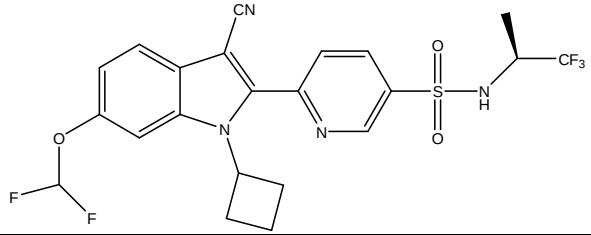
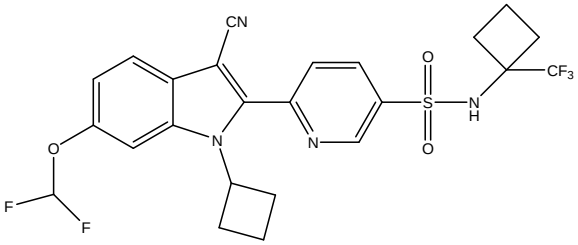
Table 1. *N*-(4'-(indol-2-yl)phenyl/pyridine sulphonamide derivatives and Their Physicochemical Parameters

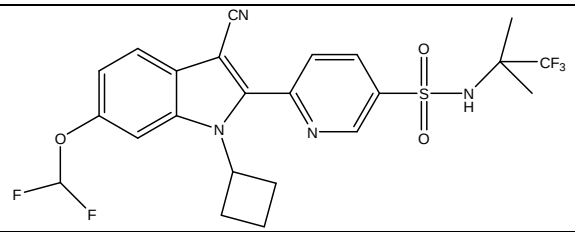
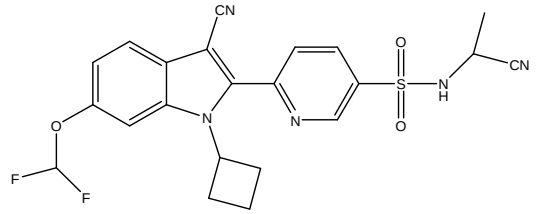
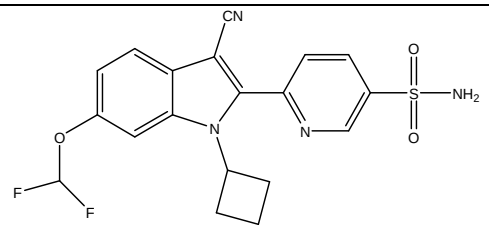
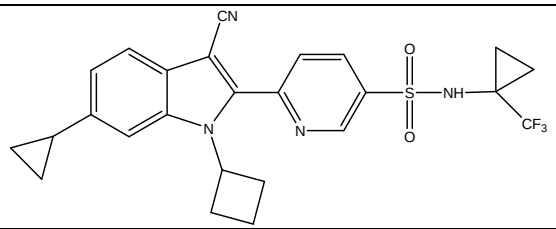
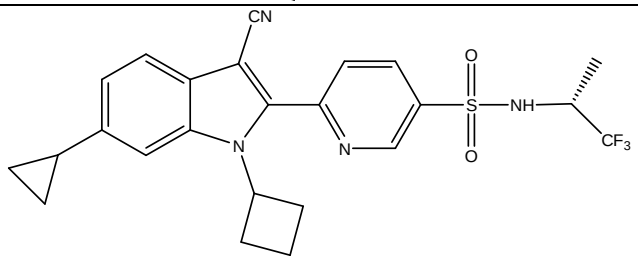
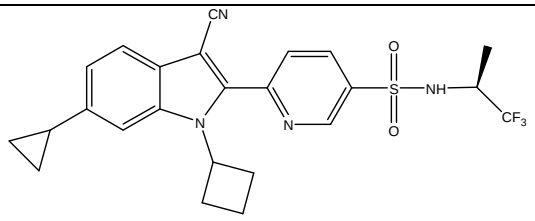
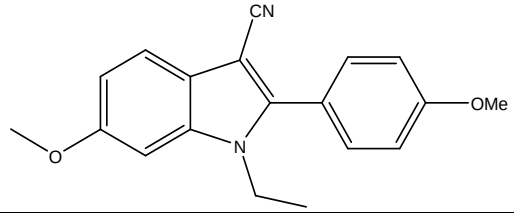
S.No	Compound	D	ST	I_4	I_1
1[b]		**1.38**	**47.30**	**0.00**	**1.00**
2		1.38	47.30	1.00	1.00

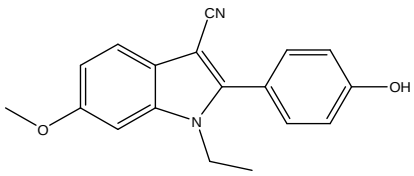
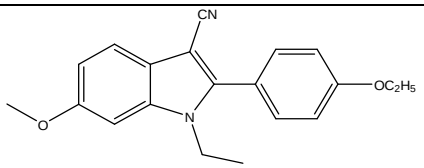
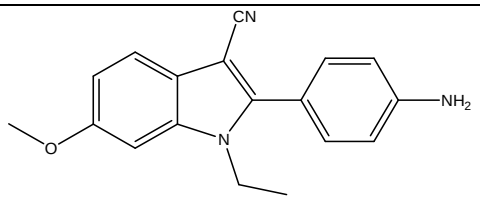
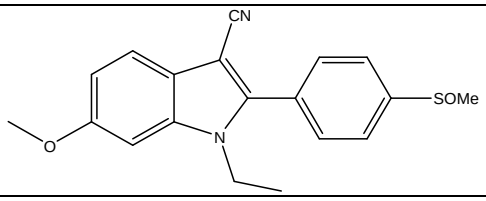
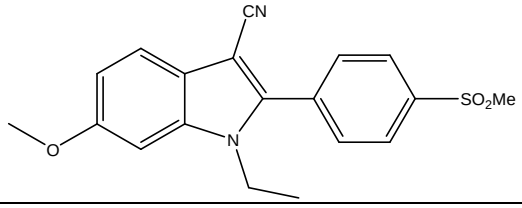
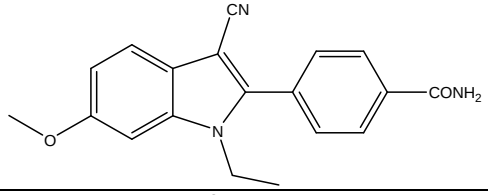
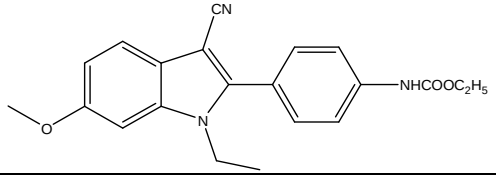
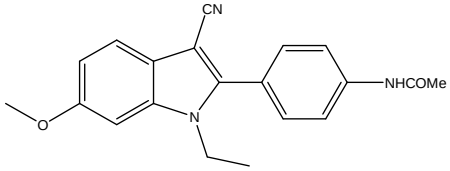
3[b]		**1.43**	**49.80**	**1.00**	**1.00**
4		1.43	49.80	1.00	1.00
5		1.49	52.70	1.00	1.00
6		1.35	43.90	1.00	1.00
7		1.35	43.90	1.00	1.00
8		1.41	48.9	1.00	1.00
9		1.38	48.10	1.00	1.00

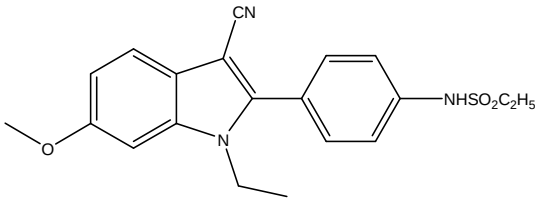
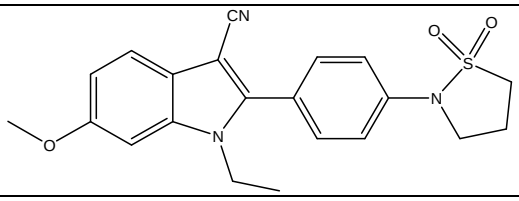
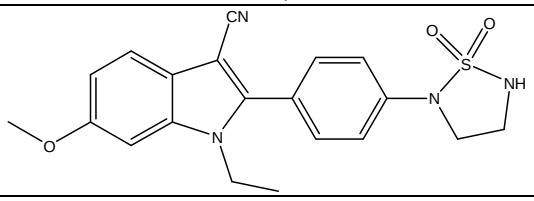
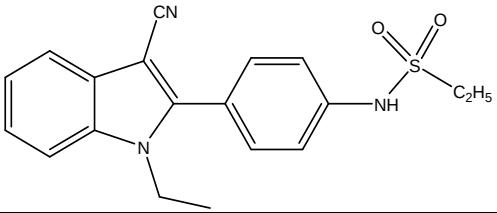
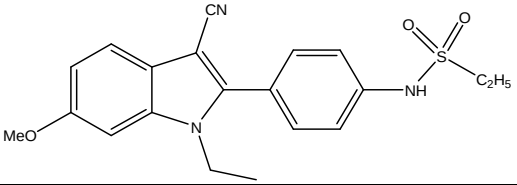
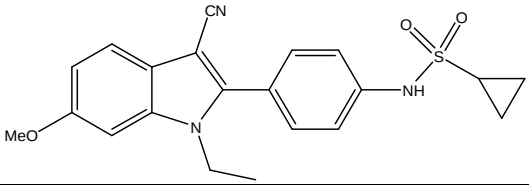
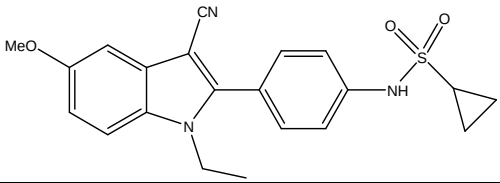
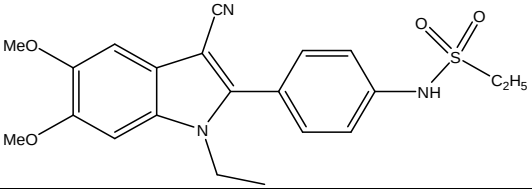
10		1.45	48.30	1.00	0.00
11[b]		**1.39**	**56.30**	**1.00**	**0.00**
12		1.36	55.00	1.00	0.00
13		1.41	49.40	1.00	0.00
14		1.32	51.00	1.00	0.00
15[b]		**1.31**	**49.00**	**1.00**	**0.00**

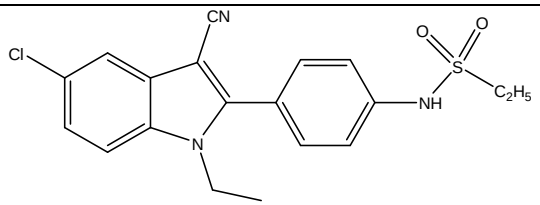
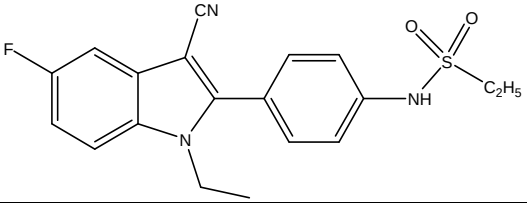
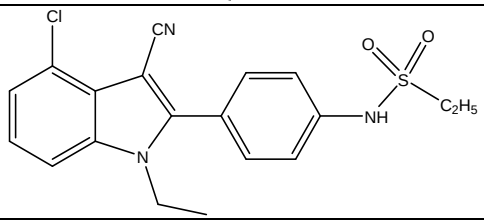
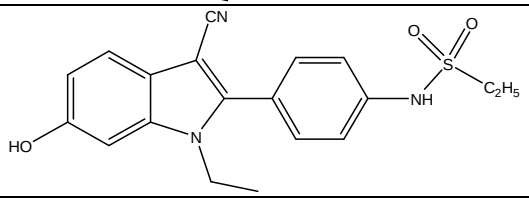
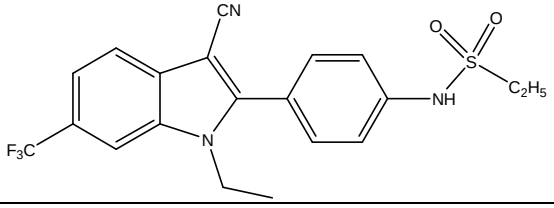
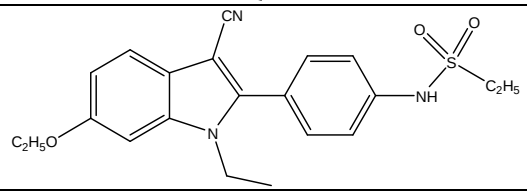
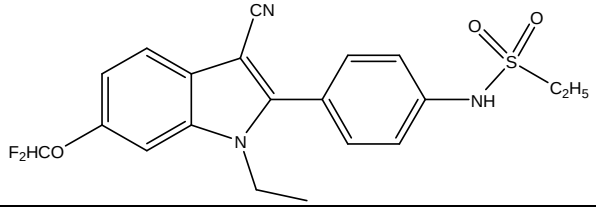
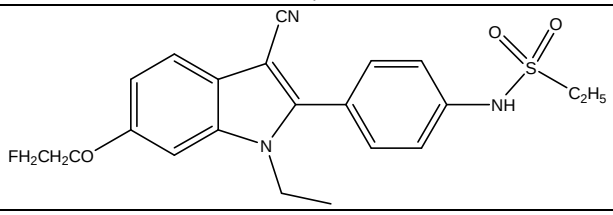
16		1.27	47.40	1.00	0.00
17		1.39	56.30	1.00	0.00
18[b]		**1.39**	**52.30**	**1.00**	**0.00**
19[c]		**1.42**	**55.30**	**1.00**	**0.00**
20		1.39	48.00	1.00	1.00
21		1.49	54.20	1.00	1.00

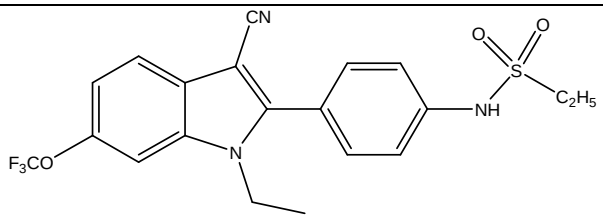
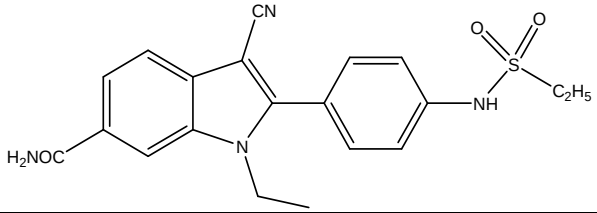
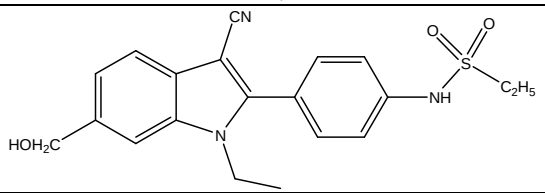
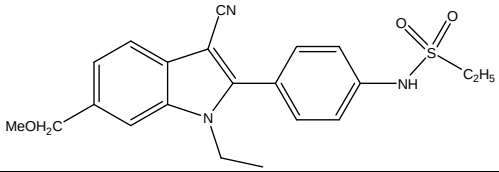
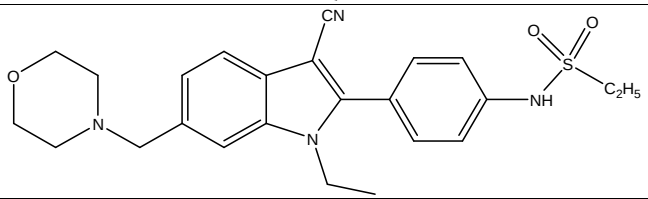
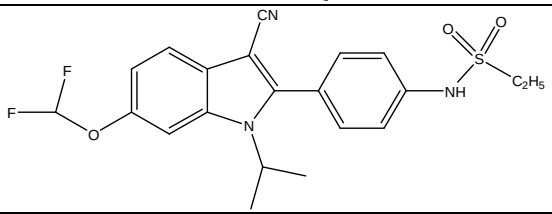
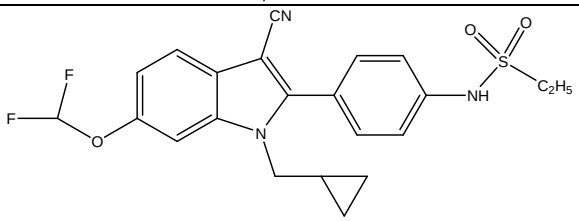
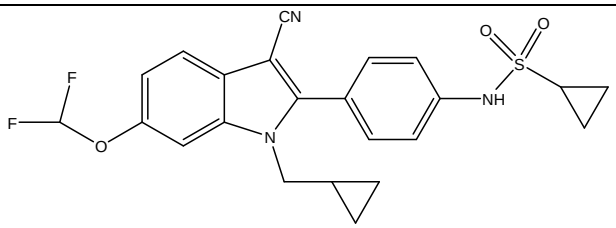
22[b]		**1.51**	**48.80**	**1.00**	**1.00**
23		1.55	59.10	1.00	1.00
24		1.42	47.50	1.00	1.00
25		1.53	46.30	1.00	1.00
26		1.53	46.30	1.00	1.00
27		1.56	50.47	1.00	1.00

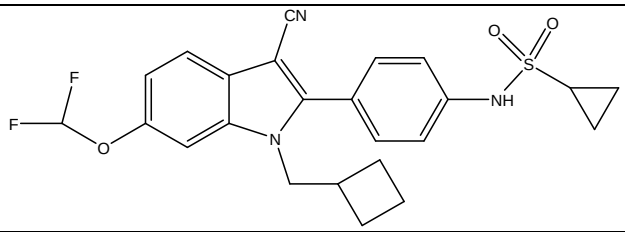
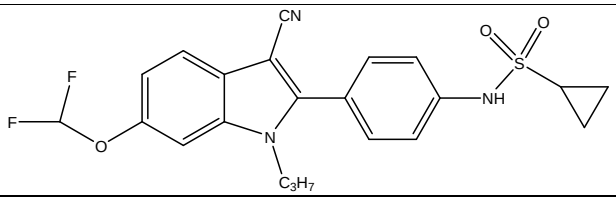
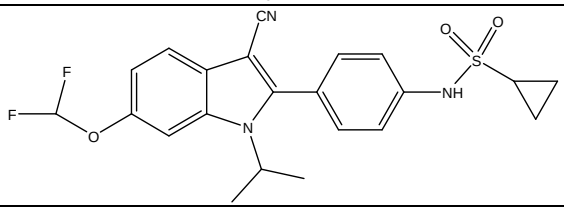
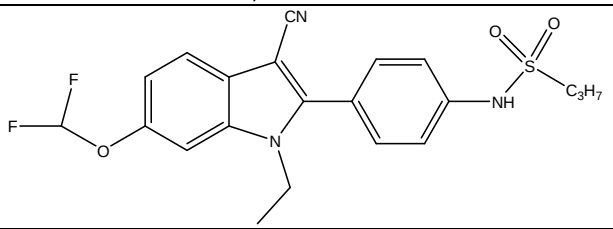
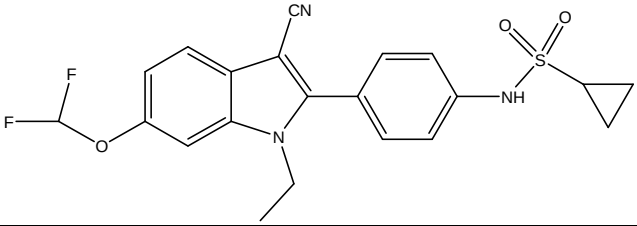
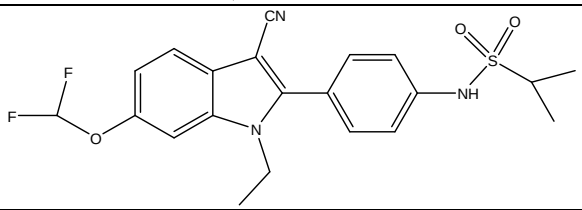
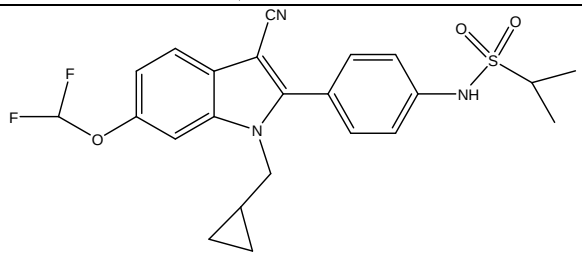
28		1.48	44.80	1.00	1.00
29		1.56	48.20	1.00	1.00
30[b]		**1.57**	**59.20**	**1.00**	**1.00**
31		1.56	57.80	1.00	0.00
32[c]		**1.49**	**52.00**	**1.00**	**0.00**
33		1.49	52.00	1.00	0.00
34[b]		**1.13**	**39.80**	**0.00**	**0.00**

35		1.18	44.30	0.00	0.00
36		1.11	39.40	0.00	0.00
37[b]		**1.18**	**43.50**	**0.00**	**1.00**
38[c]		**1.23**	**49.90**	**0.00**	**0.00**
39[b]		**1.24**	**46.30**	**0.00**	**0.00**
40		1.23	47.80	0.00	0.00
41[b]		**1.18**	**43.50**	**0.00**	**0.00**
42		1.19	43.50	0.00	0.00

43		1.25	47.30	0.00	0.00
44		1.31	52.30	0.00	1.00
45[b]		**1.37**	**55.60**	**0.00**	**1.00**
46		1.24	49.00	0.00	1.00
47		1.25	47.30	0.00	1.00
48[b]		**1.24**	**46.30**	**0.00**	**0.00**
49		1.24	45.30	0.00	1.00
50		1.26	45.90	0.00	1.00

51		1.32	50.10	0.00	1.00
52[b]		**1.29**	**47.10**	**0.00**	**1.00**
53		1.32	50.50	1.00	0.00
54		1.31	52.50	0.00	0.00
55		1.34	43.20	0.00	0.00
56[b]		**1.23**	**46.50**	**0.00**	**0.00**
57		1.34	45.10	0.00	0.00
58		1.27	46.00	0.00	0.00

59[b]		**1.36**	**42.38**	**0.00**	**0.00**
60[b]		**1.34**	**55.50**	**0.00**	**0.00**
61		1.29	51.30	0.00	0.00
62		1.23	46.50	0.00	0.00
63[b]		**1.28**	**50.90**	**0.00**	**0.00**
64		1.32	43.40	0.00	0.00
65		1.40	49.30	0.00	0.00
66		1.50	55.00	0.00	0.00

67		1.46	53.70	0.00	0.00
68[b]		**1.40**	**49.30**	**0.00**	**0.00**
69		1.41	48.10	0.00	0.00
70[b]		**1.32**	**44.50**	**0.00**	**0.00**
71		1.43	50.30	0.00	0.00
72[b]		**1.32**	**43.40**	**0.00**	**0.00**
73		1.38	47.30	0.00	0.00

[a]Taken from refs [9,10]. [b]Used for test set. [c]Not included in the derivation of Eq. (3) as they exhibited aberrant behavior.

When multiple regression analysis was performed on the compounds of training set, the following correlation was obtained.

$\log(1/EC_{50})$=7.235(±1.871)D-0.109(±0.041)ST+0.495(±0.363)I_4-0.362(±0.302)I_1
+1.832(±2.212)

n = 50, r = 0.862, r^2_{cv} = 0.59, r^2_{pred}= 0.637, s = 0.473, $F_{4,45}$ = 32.425(3.76) (3)

In this equation *n* is the number of data points used to derive it, *r* is the multiple correlation coefficient, *s* is the standard deviation, *F* is F-ratio between the variances of calculated and observed activities. The figure within parenthesis refers to standard F-value at 99% level, data within the parenthesis with ± sign are 95% confidence intervals. All these statistical parameters indicated the significant correlation exhibited by this equation between the HCV inhibition potency of the compounds and their density and surface tension. The positive coefficient of D suggests that increase in density of compound will increase its activity. It can also be said that activity of the compounds will be controlled by the size of the molecule, indicating that there is dispersion interaction between the molecule and receptor. The equation, however, also contains two indicator variables I_4 and I_1. The I_4 has been used with a value of unity for, sulphur atom of sulphonamide group placed adjacent to the aromatic ring and I_1 has been used with a value of 1 for a substitution of difluoromethoxy at 6-position of an indole ring. A positive coefficient of I_4 indicates that a sulphur atom attached to the phenyl ring is conducive to the activity while nitrogen atom attached to the ring is detrimental to it. The reason for the positive effect of it due to the absence of hydrogen bond interaction between sulphur and the receptor. On the other hand substitution of difluoromethoxy at 6-position of an indole ring is detrimental for the activity may be due to some electronic factors. The negative coefficient of surface tension may be attributed to its steric role that might hinder the interaction with the active site of the receptor. All the indicator variables and surface tension (ST) parameter play significant role in overall significance of the correlation expressed by Eq. (3). If they are eliminated one-by-one, a significant decrease in the significance of correlations is successively observed.

$\log EC_{50}$ = 6.956(±1.950) D -0.100(±0.042) ST +0.368(±0.365) I4 +1.697(±2.321)
n=50,r=0.842,s=0.497,$F_{3,46}$=37.356(4.24) (4)
$\log EC_{50}$ = 8.113(1.628) D -0.105(0.043) ST +0.502(2.060)
n=50,r=0.826,s=0.514,$F_{2,47}$ = 50.60 (5.10) (5)
$\log EC_{50}$ = 5.928(1.642) D -1.630(2.258)
n=50,r=0.723, s=0.623, $F_{1,48}$=52.705(7.19) (6)

The correlation expressed by Eq. (3) seems to have a quiet good predictive ability. The excellent agreement have been shown between the predicted and observed activity values for the both test and training set as shown in Table **2**. These observations can be also be visualized graphically in Fig.**1** and **2**.

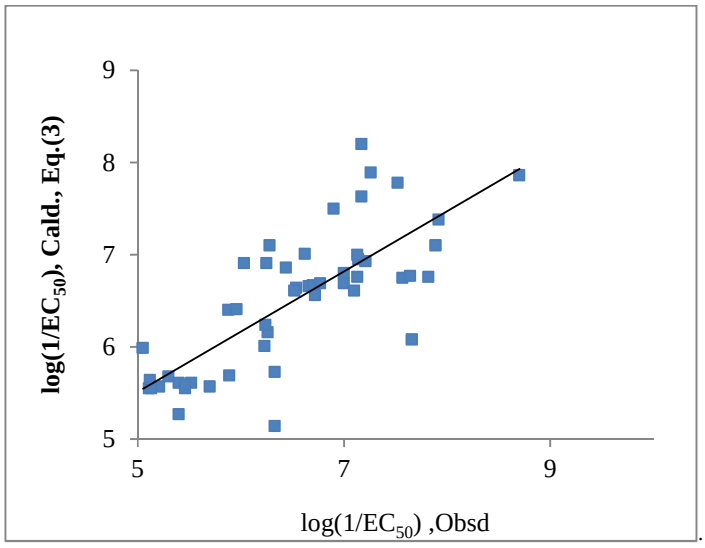


Fig. (1) A Plot of calculated activities *vs* observed activities for the compounds of training set

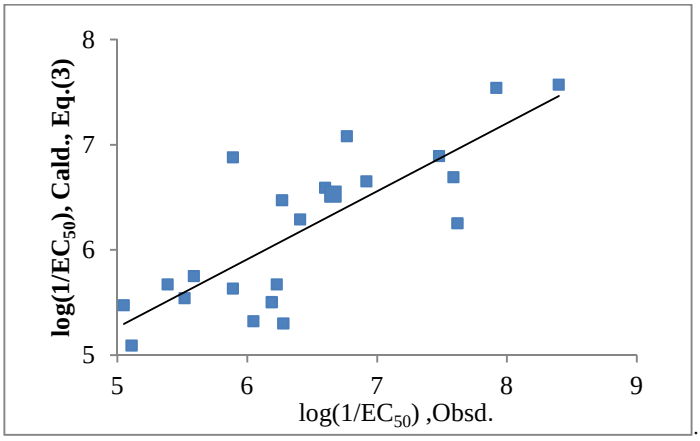


Fig. (2) A Plot of calculated activities *vs* observed activities for the compounds of test set

Using Eq. **3** some new compounds were predicted belonging to the series of Table **1** as listed in Table **3**. All these compounds have their activity greater than any compound in the existing series. In deriving Eq. (3), however, some compounds were deleted as they exhibited aberrant behaviors (large differences in their observed and predicted activities).

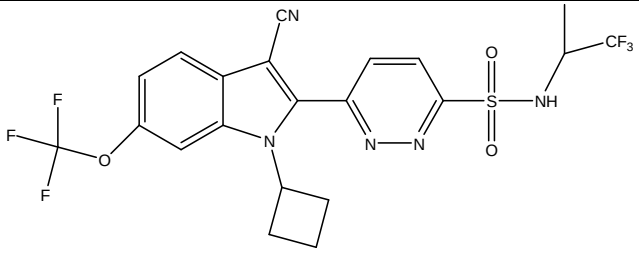
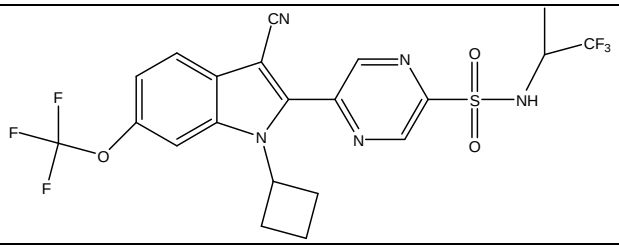
Table 2. Observed and Calculated anti-HCV Activity Values of Compounds of Table 1

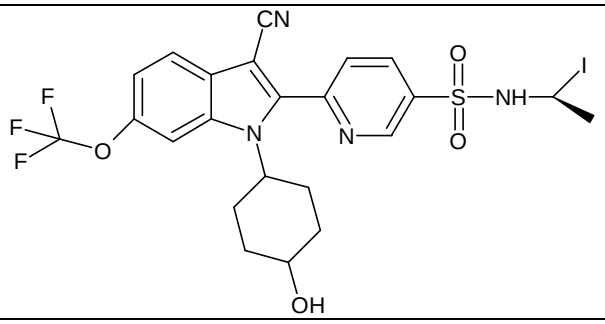
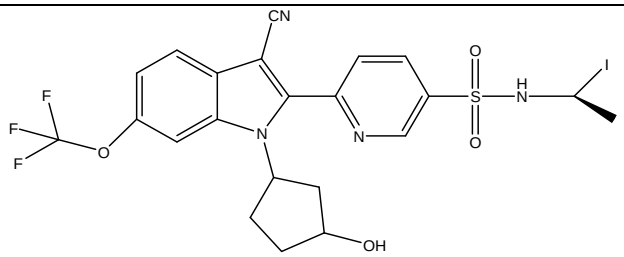
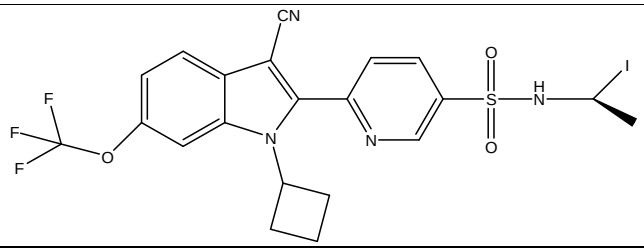
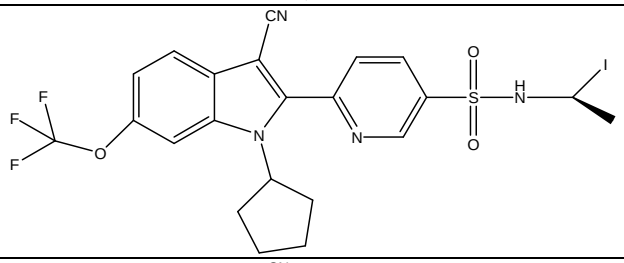
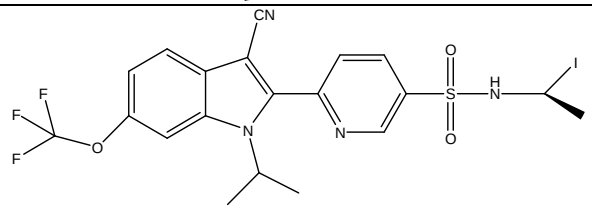
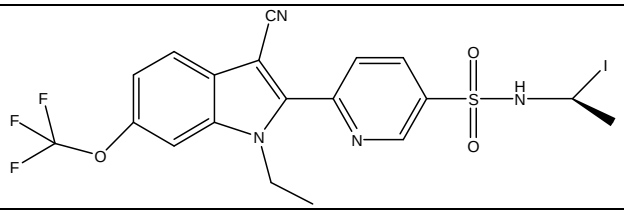
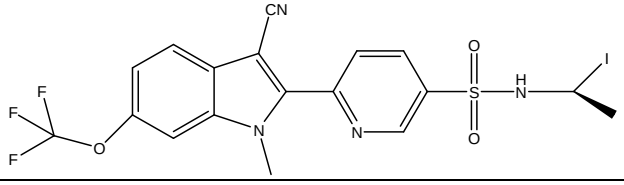
Compd.	log(1/IC$_{50}$)		
	Obsd[a]	Calcd, Eq.(3)	Pred.(LOO)
1[b]	**6.41**	**6.29**	-
2	7.00	6.80	6.80
3[b]	**7.48**	**6.89**	-
4	6.25	6.89	6.91
5	6.62	7.01	7.01
6	7.14	6.95	6.95
7	7.21	6.95	6.93
8	7.82	6.84	6.76
9	7.66	6.71	6.08
10	7.17	7.56	7.63
11[b]	**7.62**	**6.25**	-
12	6.26	6.18	6.16
13	7.89	7.15	7.10
14	5.88	6.33	6.40
15[b]	**6.27**	**6.47**	-
16	5.96	6.36	6.41
17	6.24	6.26	6.24
18[b]	**7.59**	**6.69**	-
19[c]	**7.55**		-
20	7.57	6.80	6.75
21	7.64	6.85	6.77
22[a]	**8.40**	**7.57**	-
23	6.03	6.75	6.91
24	6.28	7.07	7.10
25	7.17	7.99	8.20
26	8.70	7.99	7.86
27	7.52	7.76	7.78
28	7.26	7.80	7.89
29	6.66	6.69	6.66
30[b]	**5.89**	**6.88**	-
31	7.13	7.32	7.00
32[c]	**6.90**	**7.45**	-
33	7.92	7.45	7.50
34[b]	**5.39**[a]	**5.67**	-
35	5.21	5.55	7.38
36	5.52	5.57	5.57

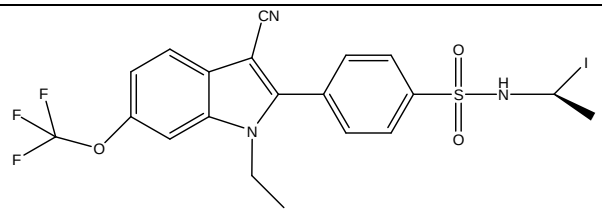
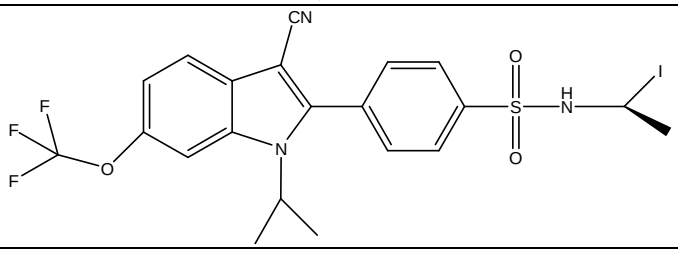
37[b]	**5.11**	**5.09**	-
38[c]	**5.40**	**5.30**	-
39[b]	**5.59**	**5.75**	-
40	5.11	5.53	5.61
41[b]	**5.89**	**5.63**	-
42	5.89	5.71	5.27
43	6.33	5.77	5.55
44	5.70	5.25	5.69
45[b]	**6.05**	**5.32**	-
46	5.05	5.12	5.73
47	6.33	5.70	5.99
48[b]	**6.19**	**5.50**	-
49	5.13	5.51	5.14
50	5.40	5.59	5.55
51	5.12	5.57	5.61
52[b]	**6.23**	**5.67**	-
53	5.46	5.52	5.64
54	5.70	5.59	5.55
55	6.44	6.82	5.57
56[b]	**6.28**	**5.30**	-
57	6.54	6.62	6.86
58	6.23	6.01	6.64
59[b]	**6.77**	**7.08**	-
60[b]	**5.05**	**5.47**	-
61	5.46	5.58	6.01
62	5.30	5.67	5.59
63[b]	**5.52**	**5.54**	-
64	6.70	6.66	5.68
65	6.52	6.59	6.67
66	6.77	6.70	6.61
67	6.72	6.55	6.69
68[b]	**6.60**	**6.59**	-
69	7.13	6.80	6.56
70[b]	**6.66**	**6.53**	-
71	7.00	6.70	6.76
72[b]	**6.92**	**6.65**	-
73	7.10	6.67	6.69

[a]Taken from refs [9,10]. [b]Used for test set. [c]Not included in the derivation of Eq. (1) as they exhibited aberrant behavior.

Table 3. Some Proposed Compounds Belonging to the series of Table 1 and Their Anti-HCV Activity Values Predicted from Eq. (3)

S.No	Compound	D	ST	I_4	I_1	log(1/IC_{50})
1		1.60	47.2	1.00	0.00	8.77
2		1.60	47.20	1.00	0.00	8.77

3		1.76	54.90	1.00	0.00	9.08
4		1.80	56.10	1.00	0.00	9.24
5		1.78	54.20	1.00	0.00	9.30
6		1.74	53.00	1.00	0.00	9.14
7		1.69	47.90	1.00	0.00	9.34
8		1.72	49.90	1.00	0.00	9.33
9		1.77	50.90	1.00	0.00	9.59

10		1.66	47.40	1.00	0.00	9.17
11		1.63	45.70	1.00	0.00	9.14

CONCLUSION

The result and discussion discussed above suggested that anti-HCV potency of *N*-(Indol-2-yl)phenyl/pyridine sulfonamides derivatives can be controlled by the shape and size of the substituents and also by various structural features. This means that some electronic interaction between the difluoromethoxy and the receptor, while sulphonamides (nitrogen is attached to phenyl ring) have hydrogen bond interaction results in detrimental effect**.**

REFERENCES

[1] M.P. Manns, G.R. Foster, J.K. Rockstroh, S. Zeuzem, F. Zoulim. *Nat. Rev. Drug Disc*., **2007**, 6, 991-1000.
[2] T. Adachi, H. Ago, N. Habuka, K. Okuda, M. Komatsu, S, Ikeda, K. Yatsunami. *Biochim. Biophys. Acta*, **2002**, 1601, 38-48.
[3] C.W. Shepard, M.J. *Alter*. *Lancet*, **2005**, 5, 524.
[4] H.B. El-serag. *Gastroenterology*, **2004**, 5, S27-S34.
[5] A.M. Bisceglie, J. Mc Hutchison, C.M. Rice, *Hepatology*, **2002**, 35, 224-231.
[6] J. Varshney, A. Sharma, P.K. Sharma. *Med. Chem. Res*., **2012**, 22, 1043-1048.
[7] S.L. Tan, A. Pause, Y. Shi. *Nat Rev Drug*, **2002**; 1: 867-881.
[8] P.L. Beaulieu. *Expert Opin Ther Pat* **2009**; 19: 145.
[9] X. Zhang, N. Zhang, G. Chen, **2013**; 23: 3947-3953.
[10] G. Chen, H. Ren. H. *Bioorg. Med. Chem. Lett*. **2013**; 23: 3942-3946.