International Journal of Interdisciplinary Research and Innovations (IJIRI)
Vol. 1, Issue 1, pp: (1-14), Month: October-December 2013, Available at: www.researchpublish.com

QSAR Study in Modeling Substituted Pyrimidines as HCV Replication Inhibitors Using 3D Morse and 2D-Autocorrelation Parameter

Shailendra Agarwal[1], Neha Singhal[2] and Anita K[3]

Department of Chemistry[1, 3]
Indian Institute of Science Education and Research (IISER) Bhopal
MP-462023, India

Abstract: **The present study deals with the investigation of HCV replication inhibitory activity of 60 compounds. Quantitative structure activity relationship (QSAR) was developed using a multiple linear regression (MLR) model. For this model, the squared correlation coefficient (R^2) is 0.81, the leave-one-out cross-validation correlation coefficient (Q_{LOO}) is 3.012. The multiple linear regression (MLR) shows that the best model is obtained using a 6 parametric model, containing GATS3e, Mor16m, Mor32u, RDF020e, RDF040u and RDF085v. These parameters are likely to influence the biological activity of these compounds. This study will pave the way for the further design, structural modification, and development of substituted pyrimidine derivatives as potent HCV NS5B inhibitors. This model has been tested using cross validation methods. The core finding of the work is given in the following research highlights-**

- **2D- QSAR studies of pyrimidine derivatives using 3D and auto-correlation parameters.**
- **Statistical analysis using a multiple linear regression method.**
- **Cross validation is done using Leave One Out method.**
- **Non colinearity and fidelity of the parameters are further checked by plotting VIF plots which confirm our results.**

Keywords: **HCV replication, 2D-autocorrelation parameters, MLR, cross validation, QSAR.**

1. Introduction

Hepatitis C virus (HCV) infection is a worldwide health hazard affecting more than 3% of the human population[1.] It is observed that HCV infection has increased the risk of developing liver cirrhosis, hepatocellular carcinoma and liver failure in humans[2]. Combinations of Ribavirin with α-interferon (IFN-α), or its polyethylene glycol modified form is the only recommended therapy but this is expensive and often causes side-effects[3].

HCV represents the only genus (hepacie virus) of flaviviridae virus family. It contains a single stranded positive sense RNA genome of 9.6KB which encodes a unique poly protein of approximately 3000 amino acids[4]. Several attempts were made to discover anti-virals which inhibits the viral replication. One of these attempts identified pyrimidine as a potential HCV replication inhibitor. Several potent and selective inhibitors of HCV replication, most of which target the NS5B RNA dependent RNA polymerase (RdRp) have been developed in recent years. Other viral (such as NS4A) and cellular targets (such as cyclophilins) involved in HCV life cycle are also being explored. The efficacy of a number of inhibitors has been or is being studied in patients chronically infected with HCV.

Both nucleoside and non-nucleoside inhibitors of HCV RdRp have been reported. Nucleoside HCV polymerase inhibitors act as premature chain terminator following conversion to their 5′-triphosphate metabolite and incorporation in the viral genome. In these studies, we have taken sixty (60) such molecules with their associated log EC_{50} activities as is reported in the literature[5].

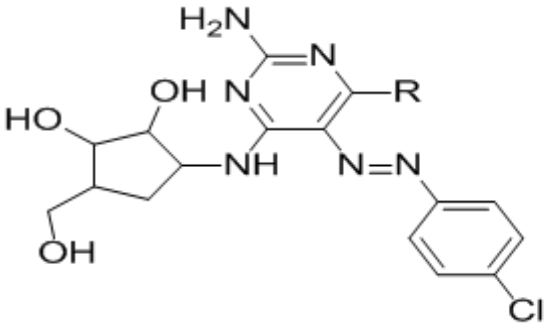
Quantitative structure activity relationships (QSAR) have often been used to find correlation between the biological activities and the molecular descriptors of different classes of compounds[6]. The QSAR model was obtained using the MLR (multiple linear regressions) statistical methods. In the present study the main goal was to build a QSAR model for the description and prediction of HCV NS5B inhibitory activity of substituted pyrimidines, using the MLR technique. This QSAR model will guide the synthesis of potentially new pyrimidine derivatives as HCV inhibitors.

2. Materials & methods

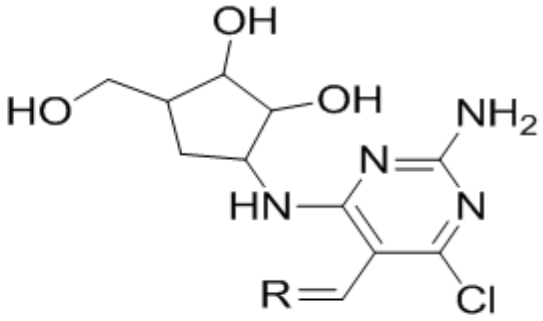
Data Set:

As a part of ongoing efforts to design novel molecules with potent HCV activity, a QSAR analysis was performed to relate HCV NS5B inhibitory activity of substituted pyrimidine derivatives to various physico-chemical properties. The selected series consisted of a total of 60 compounds with their biological activities, expressed in terms of the effective concentration (EC_{50}). For the purposes of correlation, these EC_{50} values have been reported in the literature in terms of their micro molar concentration (µM). The reported EC_{50} values were converted to their molar units and subsequently to the free energy related negative logarithmic state, i.e., Log (1/ EC_{50}). Scheme 1-3 depicts the chemical structure of the compounds. These compounds along with their inhibition data are presented in Table 1.

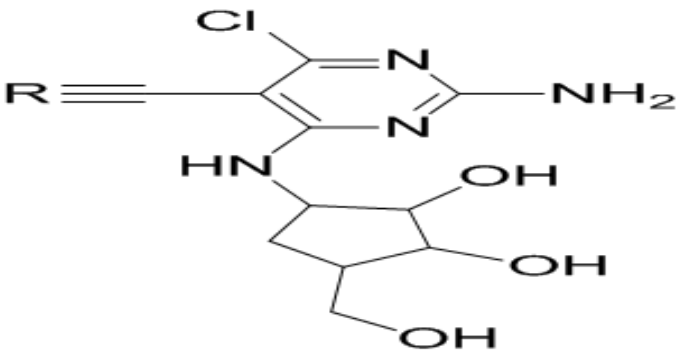
Scheme 1 Structure of the R group for (Compounds 1-13)



Scheme 2- Structure of the R group for compounds (14-30)



Scheme 3- Structure of the R group for (Compounds 31-60)



The structures of the compounds in the selected series were sketched using ChemDraw module of ChemOffice 2009and the sketched structures were then transferred to Chem3D module for generation of their three dimensional structures[7]. The geometries of the generated 3D structures were pre-optimized using MM2 force field as is implemented in the Chem3D module of CS ChemOffice 2009. All the molecular geometries were optimized using the quantum chemical program package MOPAC 7.1 applying the AM1 parameterization. The gradient norm 0.001 kcal/Å was used to calculate the electronic, geometric and energy parameters for these isolated molecules[8]. The optimized geometries of the molecules were used to compute the necessary quantum chemical descriptors available in the MOPAC server of Chem3D module. Furthermore, the molecular output was also used for the calculation of some selected descriptors available in the software E-DRAGON applet of VCClab server[9-10]. Here, for the purposes of computation of the descriptors, the valence satisfied (free valence was satisfied with hydrogen) structural fragments of the R group of the compound have been drawn in ChemDraw using the standard procedures. The molecular descriptors employed in the present study are summarized in Table 2. Empirical, constitutional, topological and functional group descriptors for all molecules were calculated using the QSAR software, Dragon 2005 and correlation between the biological activity and the molecular descriptors was found through forward stepwise multiple regression analysis using the method of least squares adopted by the statistical program NCSS[11].

3. Model Development

The statistical quality of the generated model was gauged by the parameters like correlation coefficient (R), squared correlation coefficient (R^2), or coefficient of multiple determination, which is a relative measure of the quality of fit, standard error of estimate (SEE) representing the absolute measure of the quality of fit, Fischer's value (F), which represent the F-ratio between the variance of calculated and observed activity, and chance statistics assuring that the results are not merely based on chance correlations. Best models were selected on the basis of their statistical significance.

Table-3 depicts the orthogonality of the descriptors in the selected QSAR models which was checked by the calculation of the overall correlation matrix. Regression analysis was performed by NCSS software. Statistically significant regression models were predicted in terms of maximum R^2 methods. All the models along with their quality has been summarized in Table-4. The selected models were validated by Leave One Out (LOO) and the validation parameters (Cross validated squared correlation coefficient (Q^2), R^2_{pred}, standard deviation of the sum of the squares of the differences between the predicted and the observed values (S_{PRESS}) were calculated for the generated model[12-13].

The Z score method was adopted for the detection of outliers. The Z score can be defined as the absolute difference between the value of the model and the activity field, divided by the square root of the mean square error of the data set. Any compound which shows a value of Z score higher than 2.5, during the generation of a particular QSAR model, is considered as an outlier.

Finally, the derived QSAR model were used for the prediction of the activity value of the compounds and the external validation parameter [predictive R^2 (R^2_{pred})] was calculated for evaluating the predictive capacity of the model. A value of R^2_{pred}greater than 3 indicates a good predictive capacity of the QSAR model.

4. Results & Discussion

In this work, six descriptors were selected, namely, Mor16m, Mor32u, GATS3e,RDF085v, RDF020E and RDF040u for the prediction ofEC_{50} values. A brief description of the descriptors is represented in Table 2. The methods for the calculations of these descriptors and their meaning have been explained in the Handbook of Molecular Descriptors byTodeschini et al[14].

In the present study, efforts have been made to find the structural requirements for the inhibitory activity of substituted pyrimidine analogs against Hepatitis C virus. We have tried to develop the best QSAR model to explain the correlation between the various autocorrelation, 3D and Getaway parameters and HCV NS5B inhibitor activities of such pyrimidine

derivatives. Table 3 demonstrates the correlation matrix showing the inter correlation amongst all the parameters. For the set of 60 compounds, the one-variable model shows the following results.

One-variable model (Model-29)

pIC_{50}=0.1817± (0.0189) RDF085v-0.9394

N=57, R^2=0.6270, R^2_A=0.6202, Se=0.4032, F=92.452, Q=1.963872

Successive regression indicated that the quality of the above model is improved by the addition of different parameters. The value of R^2 and R^2_{adj} is also increased in each case.

Two –variable model (Model-33)

$pIC_{50=}$0.0603± (0.0215) RDF020e-0.1649± (0.0188) RDF085v-1.3759

N=57, R^2=0.6744, R^2_A=0.6624, Se=0.3802, F=55.929, Q=2.1599

Three-variable model (Model-36)

$pIC_{50=}$-0.3781± (0.2047) Mor16m+0.0613± (0.0210) RDF020e+0.1707± (0.0187) RDF085v-1.5319

N=57, R^2=0.6941, R^2_A=0.6768, Se=0.3720, F=48.089, Q=2.239

Four-variable model (Model-37)

$pIC_{50=}$-0.9591±(0.8556)GATS3e-0.3836±(0.2043)Mor16m+0.0534±(0.0222)RDF020e+0.1589± (0.0214) RDF085v-0.2812

N=57, R^2=0.7013, R^2_A=0.6784, Se=0.3711, F=30.527, Q=2.2566

Five-variable model (Model-41)

$pIC_{50=}$-1.8727±(0.7828)GATS3e+1.0371±(0.2383)Mor32u+0.1038±(0.0219)RDF020e-
0.0339±(0.0131)RDF040u+0.1778±(0.0187)RDF085v+1.5526

N=57, R^2=0.7879, R^2_A=0.7671 Se=0.3157, F=37.897 Q=2.811

Six-variable model (Model-42)

$pIC_{50=}$-2.0375±(0.7447)GATS3e-0.4363±(0.1678)Mor16m+1.0018±(0.2263)Mor32u+0.1079±(0.0208)RDF020e-0.0399±(0.0126)RDF040u+0.1839±(0.0179)RDF085v+1.6489

N=57, R^2=0.8132, R^2_A=0.7908 Se=0.2993, F=36.273, Q=3.0125

We have observed that in all the other cases,(except for the compounds 2, 27 and 31) the estimated activities are very near to the experimental activities. Hence these compounds were considered as outliers. The quality of the regression model after the deletion of these three compounds are shown in Table-4b.The values of the descriptors in the QSAR model derived for the series are tabulated in Table -5.

The selected series of compounds were modeled using a wide range of descriptors. Amongst them, RDF020e, RDF040u, RDF085v, GATS3e (2D auto correlation parameter), Mor16m, Mor32u (3D Morse) are some of the highly correlated descriptors. Here, the RDF code considers the atoms as virtual spheres from 1.0 to 15.5 Å in diameter. Interestingly, the optimum model only includes contributions from the inner part of the inhibitor structures. The obtained MLR model only takes into account the atoms which are less than 4.0 Å in diameter, excluding atoms at the most external spheres. RDF040u considers the atoms around a distance of 4.0Å and has a negative effect. RDF020e considers the atoms around 2.0 Å in the Sanderson atomic electro negativity weighting scheme, and RDF085v considers the atoms around 8.5 Å in the atomic volume

weighting scheme. The radial distribution function descriptors are difficult to interpret. They cannot estimate the specific positions of the substituent as these encode mathematically defined information, although the inclusion of the atomic property weighting scheme provides greater applicability. We, therefore, conclude that the structural information obtained showed that an adequate distribution of Sanderson's electro negativities and the atomic volume has an important influence on the HCV-NS5B inhibitory activities of the studied compounds[15-17].

Mor16m is a 3D MoRSE descriptors (3D Molecule Representation of Structures based on Electron diffraction) which are derived from Infrared spectra simulation using a generalized scattering function. The 3D-MoRSE code as a molecular transform developed here shows a great potential for the representation of molecular structures. It has several merits, for example, (a) the number of values is independent of the size of the molecule and thus allows the study of datasets of greater structural variety; (b) the number of these values can be changed and thus the resolution in the representation of a molecular structure can be scaled. Different atomic properties such as atomic number, mass, partial charge, polarizability, etc. can be considered providing greater flexibility in the representation of molecules. Thus, 3D-MoRSE code can reveal the skeleton and information on the substituents for a given molecule. 3D-MoRSE—signal 32/unweighted (Mor32u) has a positive influence on the activity whereas 3D-MoRSE—signal 16/weighted by atomic masses (Mor16m) has a negative coefficient indicating that the low molecular weight compounds will prove detrimental for the inhibitory activity.

The descriptor GATS3e belongs to the 2D-AUTO class of descriptors. The 2D-AUTO descriptors have their origin in the autocorrelation of the topological structure of Broto-Moreau (ATS), of Moran (MATS) and of Geary (GATS)[18-21]. The computation of these descriptors involves the summation of different autocorrelation functions corresponding to the different fragment lengths and leads to different autocorrelation vectors corresponding to the lengths of the structural fragments. The function of autocorrelation is a summation of the function value products calculated at x and x+1, where l is the lag. From the above regression model, GATS3eindicates that the presence of an electronegative atom at topological distance equals to 3 contributes negatively to the inhibitory activity. It suggests that the presence of a less electronegative atom at these topological distances contributes significantly to the inhibitory activity.

The difference between the observed and the calculated activity (residual) is given in Table-6 which is least for model 42. This shows that it is the most appropriate model for modeling the $\log EC_{50}$ value forthe present set of compounds. The predictive potential of this model has also been obtained by plotting a graph between observed $\log EC_{50}$ and estimated $\log EC_{50}$ values and is depicted in Figure 1.

The statistical details of the QSAR model given above accounts for its good statistical quality. The R^2 of the model shows a better improvement from 5 parametric models to 6 parametric model and also R^2_A shows a significant increase. The predictive potential of the proposed models can be determined by the calculation of Pogliania quality factor (Q) and cross validation values[22-24]. The calculated values of Q for the proposed models increase from monoparametric to 6-parametric model. Increase in the R^2_A value indicates that the added parameter is favorable for exhibiting the desired activity.

5. Model Validation

To validate the model, cross validation parameters have been calculated and are reported in Table-7. It is a known fact that PRESS is a good estimate of the real predictive power of the model. If PRESS is smaller than SSY, the model is considered to be statistically significant. From Table 7, it is clear that all the models proposed by us are statistically significant. To be a reasonably good QSAR model, the ratio of PRESS/SSY should be smaller than 0.4. The model proposed by us (Model-42) satisfies this ratio, and thus model 42 has an excellent predictive power. The models are cross validated by Leave-One-Out method. Another cross validated parameter which is related to the uncertainty of prediction, PSE has also been calculated. The lowest values of PSE for the model support the highest predictive potential. The low value of PSE and S_{PRESS} and high value

of R^2_{cv} suggests that the six parametric model is the most appropriate in predicting logEC_{50} values of the present set of compounds.

For any kind of possible defect, we have calculated variants inflation factor, tolerance and condition number for various parameters using VIF plot (Figure 2). The effects of inter correlation of descriptors were checked through variance inflation factor (VIF). VIF value is calculated using the formula $1/(1 - r^2)$, where r^2 is the multiple correlation coefficient of one descriptor's effect regressed on the remaining molecular descriptors. If VIF value is larger than 10, the information of the descriptor might overlap with the other descriptors. In the model, the VIF values of these descriptors are positioned in the range of 1.08 to 1.91 (Table 8). Therefore, from VIF analysis, it is clear that the descriptors used in models are fully self-governing. Therefore, the model is free from any kind of defects. The ridge trace suggest that there is no colinearity in the model.

Table 1. Structure of the compound in Scheme 1-3 along with their biological activity

Compound No.	R	log EC_{50}
1	4-Cl	0.522879
2	3-Cl	0.49485
3	2-Cl	0.958607
4	H	1
5	4-F	0.69897
6	3-F	1.154902
7	2-F	1
8	4-CN	0.079181
9	3-CN	0.221849
10	4-NO_2	0.579784
11	3-NO_2	0.045757
12	4-Me	1.221849
13	4-Ome	1.69897
14	Ph	0.045757
15	4-MePh	0.278754
16	4-MeOPh	0.278754
17	3-MeOPh	0.221849
18	2-MeOPh	0.39794
19	3,4-DiMeOPh	0.920819
20	4-ClPh	0.518514
21	2-ClPh	0.522879
22	3-PPh	0.69897
23	2-PPh	0.30103
24	4-CF3Ph	0.623249
25	4-CNPh	0.740363
26	3-CNPh	0.221849
27	4-Pyridyl	0.531479
28	trans-3-pyridyl	0
29	cis-3-pyridyl	0.60206
30	2-Pyridyl	0.361728
31	Ph	0.69897
32	4-MePh	0.853872
33	4-MeOPh	0.886057
34	4-PPh	0.468521
35	2-PPH	0.638272
36	(2-CO2Me)Ph	0.886057
37	(2-CO2Et)Ph	1
38	(2-CO2iPr)Ph	0.522879
39	(2-CONH2)Ph	0.69897
40	(4-CONH2)Ph	0.69897
41	(2-CONHEt)Ph	0.69897
42	(2-SO2Me)Ph	1
43	(4-SO2Me)Ph	0.278754
44	(2-SO2NH2)Ph	0.60206
45	2-Thiazolyl	0.69897
46	2-Thiophenyl	0.823909
47	3-Thiophenyl	1
48	2-Imidazolyl	0.30103
49	2-Pyridyl	0.09691
50	3-Pyridyl	0.39794
51	4-Pyridyl	0.045757
52	(2-CO2Me)-3-pyridyl	1.39794
53	(2-CO2Et)-3-pyridyl	1.69897
54	(2-Me)-3-pyridyl	1.154902
55	(4-Me)-3-pyridyl	0.823909
56	(2,4-diMe)-3-pyridyl	1.30103
57	(4-MeO)-3-pyridyl	1.522879
58	(4-EtO)-3-pyridyl	1.69897
59	(4-MeO-2-Me)-3-pyridyl	2
60	(4-EtO-2-Me)-3-pyridyl	1.522879

Table 2. Descriptor classes used for the analysis of HCV activity

Symbol	Definition
RDF040u	Unweighted radial distribution function at 4.0Å
RDF020e	Radial distribution function at 2.0Å weighted by Sanderson electro negativities
RDF085v	Radial Distribution Function - 8.5 Å / weighted by atomic Vander Waals volumes
Mor16m	3D-MoRSE - signal 16 / weighted by atomic masses
Mor32u	3D-MoRSE - signal 32 / unweighted
GATS3e	Geary autocorrelation - lag 3 / weighted by atomic Sanderson electro negativities

Table 3. Correlation Matrix between the experimental $LogEC_{50}$ value and different molecular descriptors used to calculate the QSAR model

	GATS3e	RDF040u	RDF085v	RDF020e	Mor32u	Mor16m	$LogEC_{50}$
GATS3e	1.0000						
RDF040u	-0.5551	1.0000					
RDF085v	-0.5334	0.3770	1.0000				
RDF020e	-0.4214	0.5758	0.2517	1.0000			
Mor32u	0.3705	-0.2898	-0.3238	-0.3877	1.0000		
Mor16	-0.1334	-0.0319	0.2726	0.0311	-0.1140	1.0000	
$LogEC_{50}$	-0.5012	0.3496	0.5809	0.4863	-0.1552	0.0642	1.0000

Table 4. Regression parameters and quality of correlation

Model No.	Parameters used	Ai=(1……7)	B	Se	R^2	R^2_A	F-ratio	Q=R/Se
1	Mor16m	-0.0200±(0.3312)	0.7179	0.6521	0.0001	0.0000	0.265	0.015335
2	Mor32u	-0.4384±(0.3924)	0.4622	0.6452	0.0211	0.0042	1.249	0.225137
3	RDF040u	0.0476±(0.0183)	-0.3736	0.6172	0.1044	0.0890	6.7620	0.523509
4	RDF020e	0.1205±(0.0302)	-0.4514	0.5777	0.2153	0.2018	15.914	0.803193
5	GATS3e	-4.7433±(0.9460)	5.9920	0.5447	0.3024	0.2904	25.141	1.009563
6	RDF085v	0.1522±(0.0197)	-0.6801	0.4579	0.5070	0.4985	59.639	1.555011
7	RDF085v Mor32u	0.1611±(0.0210) 0.3601±(0.2961)	-0.5485	0.4560	0.5194	0.5026	30.806	1.58047
8	GATS3e RDF085v	-2.0166±(0.9146) 0.1252±(0.0227)	1.8086	0.4434	0.5457	0.5298	34.236	1.666024
9	Mor16m RDF085v	-0.5568±(0.2326) 0.1652±(0.0197)	-0.9329	0.4403	0.5520	0.5363	35.115	1.687411
10	RDF020e RDF085v	0.0768±(0.0229) 0.1534±(0.0188)	-1.2738	0.4221	0.5883	0.5738	40.723	1.817122
11	GATS3e RDF020e RDF085v	-1.1457±(0.9232) 0.0664±(0.0243) 0.1223±(0.0215)	0.2204	0.4201	0.5993	0.5778	27.919	1.842763
12	RDF020e RDF040u RDF085v	0.0982±(0.0267) -0.0240±(0.0159) 0.1445±(0.0196)	-1.0138	0.4175	0.6043	0.5831	28.510	1.861958
13	Mor16m RDF020e RDF085v	-0.5335±(0.2135) 0.0751±(0.0219) -0.1482±(0.0187)	-1.5025	0.4039	0.6296	0.6097	31.727	1.964529
14	Mor32u RDF020e RDF085v	0.7542±(0.2757) -13.1777±(2.5639) -5.2853±(0.8226)	-1.1617	0.4000	0.6368	0.6173	32.728	1.994994

15	GATS3e Mor16m RDF020e RDF085v	-1.1441±(0.8823) -0.5333±(0.2122) 0.0647±(0.0232) 0.1352±(0.0212)	-0.0104	0.4015	0.6406	0.6144	24.505	1.993462
16	Mor32u RDF020e RDF040u RDF085v	0.7531±(0.2721) 0.1192±(0.0264) -0.0239±(0.0150) 0.1586±(0.0192)	-0.9030	0.3947	0.6527	0.6274	25.842	2.046867
17	GATS3e Mor32u RDF020e RDF085v	-1.5065±(0.8721) 0.8195±(0.2736) 0.0862±(0.0236) 0.1335±(0.0205)	0.8128	0.3931	0.6555	0.6304	26.162	2.059602
18	Mor16m Mor32u RDF020e RDF085v	-0.5086±(0.2023) 0.7244±(0.2638) 0.0955±(0.0220) 0.1611±(0.0184)	-1.3842	0.3822	0.6742	0.6506	28.460	2.148343
19	GATS3e Mor16m RDF020e RDF040u RDF085v	-1.9709±(0.8881) -0.6252±(0.2036) 0.0948±(0.0246) -0.0425±(0.0156) 0.1442±(0.0203)	1.4897	0.3799	0.6841	0.6548	23.387	2.177161
20	GATS3e Mor32u RDF020e RDF040u RDF085v	-2.2384±(0.8890) 0.8496±(0.2626) 0.1133±(0.0253) -0.0369±(0.0153) 0.1399±(0.0198)	2.1724	0.3768	0.6892	0.6604	23.948	2.20324
21	GATs3e Mor16m Mor32u RDF020e RDF085v	-1.4917±(0.8315) -0.5061±(0.1983) 0.7892±(0.2611) 0.0838±(0.0226) 0.1453±(0.0200)	0.5719	0.3747	0.6926	0.6641	24.330	2.221046
22	Mor16m Mor32u RDF020e RDF040u RDF085v	-0.5743±(0.1985) 0.7192±(0.2558) 0.1221±(0.0248) -0.0303±(0.0143) 0.1742±(0.0188)	-1.0843	0.3706	0.6993	0.6715	25.116	2.256453
23	GATS3e Mor16m Mor32u RDF020e RDF040u RDF085v	-2.3699±(0.8216) -0.6011±(0.1866) 0.8197±(0.2425) 0.1160±(0.0234) -0.0445±(0.0143) 0.1552±(0.0189)	2.1634	0.3478	0.7401	0.7107	25.154	2.473521

After deletion of the three compounds (compd. no 2, 27 and 31)

Model No.	Parameters used	Ai=(1……7)	B	Se	R^2	R^2_A	F-ratio	Q=R/Se
24	Mor16m	0.0397±(0.3568)	0.7402	0.6601	0. 0002	0. 0000	0.012	0.021424
25	Mor32u	-0.5116±(0. 4184)	0.4300	0.6514	0. 0265	0. 0088	1.495	0.249905
26	RDF040u	0.0479±(0.0189)	-0.3755	0.6246	0.1049	0.0886	6.444	0.518544
27	RDF020e	0.1205±(0.0314)	-0.4505	0.5865	0.2107	0.1964	14.683	0.782644
28	GATS3e	-5.0175±(0.9777)	6.3052	0.5429	0.3238	0.3115	26.337	1.048138
29	RDF085v	0.1817±(0.0189)	-0.9394	0.4032	0.6270	0.6202	92.452	1.963872
30	Mor16m RDF085v	-0.3625±(0.2184) 0.1876±(0.0189)	-1.0819	0.3969	0.6451	0.6320	49.081	2.023636
31	GATS3e RDF085v	-1.5729±(0.8623) 0.1586±(0.0224)	1.0208	0.3949	0.6486	0.6356	49.846	2.039395
32	Mor32u RDF085v	0.6274±(0.2742) 0.2007±(0.0200)	-0.7452	0.3885	0.6600	0.6474	52.403	2.09113
33	RDF020e RDF085v	0.0603±(0.0215) 0.1649±(0.0188)	-1.3759	0.3802	0.6744	0.6624	55.929	2.159965
34	GATS3e	-0.9210±(0.8755)	-0.1727	0.3798	0.6811	0.6630	37.728	2.172954

	RDF020e RDF085v	0.0527±(0.0227) 0.1535±(0.0217)						
35	RDF020e RDF040u RDF085v	0.0815±(0.0247) -0.0239±(0.0144) 0.1739±(0.0193)	-1.1153	0.3742	0.6904	0.6729	39.395	2.220479
36	Mor16m RDF020e RDF085v	-0.3781±(0.2047) 0.0613±(0.0210) 0.1707±(0.0187)	-1.5319	0.3720	0.6941	0.6768	48.089	2.239588
37	GATS3e Mor16m RDF020e RDF085v	-0.9591±(0.8556) -0.3836±(0.2043) 0.0534±(0.0222) 0.1589±(0.0214)	-0.2812	0.3711	0.7013	0.6784	30.527	2.256633
38	GATS3e Mor16m RDF020e RDF040u RDF085v	-1.8484±(0.8686) -0.4808±(0.1957) 0.0822±(0.0234) -0.0405±(0.0147) 0.1647±(0.0203)	1.2826	0.3496	0.7399	0.7144	29.020	2.460453
39	GATS3e Mor16m Mor32u RDF020e RDF085v	-1.1638±(0.7502) -0.3402±(0.1790) 1.0099±(0.2454) 0.0797±(0.0204) 0.1784±(0.0193)	0.1128	0.3246	0.7758	0.7538	35.293	2.713478
40	Mor16m Mor32u RDF020e RDF040u RDF085v	-0.3972±(0.1775) 0.9663±(0.2398) 0.1124±(0.0220) -0.0271±(0.0124) 0.2029±(0.0175)	-1.1403	0.3177	0.7852	0.7642	37.288	2.789157
41	GATS3e Mor32u RDF020e RDF040u RDF085v	-1.8727±(0.7828) 1.0371±(0.2383) 0.1038±(0.0219) -0.0339±(0.0131) 0.1778±(0.0187)	1.5526	0.3157	0.7879	0.7671	37.897	2.811648
42	GATS3e Mor16m Mor32u RDF020e RDF040u RDF085v	-2.0375±(0.7447) -0.4363±(0.1678) 1.0018±(0.2263) 0.1079±(0.0208) -0.0399±(0.0126) 0.1839±(0.0179)	1.6489	0.2993	**0.8132**	0.7908	36.273	3.01295

Table 5. Calculated values of the Autocorrelation parameters of HCV-NS5B Inhibitors with 3D Morse values and their experimental pEC_{50} values.

S.No.	GATS3e	RDF040u	RDF085v	RDF020e	Mor32u	Mor16m	logEC50
1	1.156	22.66	8.03	6.39	-0.45	-0.23	0.522
2	1.15	22.64	12.38	7.67	-0.52	0.05	0.49
3	1.13	21.94	11.38	7.8	-0.53	-0.16	0.95
4	1.14	21.209	10.77	7.85	-0.54	-0.35	1
5	1.09	19.711	12.12	7.66	-0.49	-0.42	-0.69
6	1.09	20.84	9.12	7.66	-0.579	-0.26	1.15
7	1.12	20.8	12.12	7.69	-0.57	-0.26	1
8	1.12	22.09	6.9	7.81	-0.571	-0.493	0.07
9	1.12	22.56	6.12	8.03	-0.48	-0.31	0.22
10	1.16	23.71	8.26	8.61	-0.32	-0.47	0.57
11	1.16	23.86	5.73	7.89	-0.38	-0.64	0.04
12	1.11	21	11.01	8.02	-0.51	-0.22	1.22

13	1.108	21.624	11.31	9.76	-0.46	-0.22	1.69
14	1.134	24.33	7.36	9.88	-0.64	-0.27	0.04
15	1.1	25.52	8.57	8.84	-0.57	-0.32	0.27
16	1.088	22.43	7.98	10.46	-0.72	-0.21	0.27
17	1.088	24.45	8.19	10.56	-0.57	-0.34	0.22
18	1.13	25.52	7.2	11.52	-0.66	-0.45	0.39
19	1.03	26.59	8.83	12.95	-0.65	-0.36	0.92
20	1.13	24.46	6.91	8.57	-0.48	-0.16	0.51
21	1.16	24.93	7.85	8.66	-0.41	0.05	0.52
22	1.06	22.76	7.56	8.51	-0.52	-0.31	0.69
23	1.16	23.35	7.69	8.73	-0.6	-0.45	0.3
24	1.006	24.79	8.72	8.63	-0.67	0.05	0.62
25	1.11	20.92	8.45	8.42	-0.58	-0.25	0.74
26	1.11	21.09	6.64	8.33	-0.34	-0.26	0.22
27	1.03	22.73	12.25	9.02	-0.85	0.011	0.53
28	1.17	22.04	7.16	9.35	-0.63	-0.22	0
29	1.17	20.137	4.09	13.76	-0.61	-0.42	0.6
30	1.17	23.3	5.74	8.96	-0.42	-0.32	0.36
31	1.13	20.16	4.74	9.14	-0.8	-0.45	0.69
32	1.1	21.03	10.05	10.46	-0.65	-0.12	0.85
33	1.088	22.39	7.54	11.28	-0.68	-0.24	0.88
34	1.06	19.78	8.53	9.72	-0.72	-0.16	0.46
35	1.16	19.25	9.94	9.88	-0.76	-0.09	0.63
36	1.099	25.3	9.43	12.31	-0.54	0.01	0.88
37	1.099	27.4	10.95	11.26	-0.57	-0.11	1
38	1.099	28.93	10.8	10.63	-0.78	-0.38	0.52
39	1.07	23.78	8.98	10.86	-0.52	-0.07	0.69
40	1.09	22.67	8.93	9.56	-0.43	-0.06	0.69
41	1.12	28.23	11.05	12.33	-0.67	-0.06	0.69
42	1.06	23.21	11.67	9.89	-0.7	0.29	1
43	1.08	22.33	9.57	10.03	-0.72	0.17	0.27
44	1.1	22.82	10.48	9.67	-0.59	0.07	0.6
45	1.24	17.89	8.41	8.39	-0.51	-0.441	0.69
46	1.11	22.36	8.14	10.63	-0.29	-0.47	0.82
47	1.13	24.37	10.211	10.87	-0.26	-0.299	1
48	1.24	16.18	7.2	9	-0.49	0.03	0.3
49	1.17	19.84	7.87	8.75	-0.69	-0.21	0.09
50	1.17	19.44	8.11	8.68	-0.77	-0.24	0.39
51	1.15	20.97	7.72	8.85	-0.76	-0.2	0.04
52	1.04	21.05	10.63	10.77	-0.37	-0.19	1.39
53	1.03	20.49	11.27	10.5	-0.68	-0.28	1.69
54	1.11	19.02	10.73	8.88	-0.52	-0.33	1.15
55	1.11	19.77	10.05	8.93	-0.55	-0.41	0.82
56	0.96	30.69	12.68	10	-0.83	-0.43	1.3

57	1.04	29.2	13.466	14.25	-0.95	-0.199	1.52
58	1.05	29.24	13.53	14.04	-0.99	-0.44	1.69
59	1.01	30.224	11.6	14.25	-0.52	-0.19	2
60	1.018	30.399	13.433	12.48	-0.84	-0.27	1.52

Table 6. Observed and Estimated values of PIC_{50} using Model no.42

Comp No.	Actual $LogEC_{50}$	Predicted $LogEC_{50}$	Residual
1	0.523	0.200	0.323
2	0.959	0.930	0.029
3	1.000	0.909	0.091
4	0.699	0.832	-0.133
5	1.155	1.184	-0.029
6	1.000	1.005	-0.005
7	0.079	0.214	-0.135
8	0.222	0.144	0.078
9	0.580	0.676	-0.096
10	0.046	0.136	-0.090
11	1.222	1.007	0.215
12	1.699	1.297	0.402
13	0.046	0.263	-0.218
14	0.279	0.475	-0.196
15	0.279	0.502	-0.223
16	0.222	0.679	-0.457
17	0.398	0.419	-0.021
18	0.921	1.007	-0.086
19	0.519	0.156	0.363
20	0.523	0.235	0.288
21	0.699	0.492	0.207
22	0.301	0.300	0.001
23	0.623	0.443	0.180
24	0.740	0.530	0.211
25	0.222	0.424	-0.202
26	0.000	0.169	-0.169
27	0.602	0.261	0.341
28	0.362	0.071	0.291
29	0.854	0.937	-0.083
30	0.886	0.565	0.321
31	0.469	0.661	-0.192
32	0.638	0.685	-0.046
33	0.886	0.920	-0.034
34	1.000	1.024	-0.024
35	0.523	0.773	-0.251
36	0.699	0.840	-0.141
37	0.699	0.774	-0.075
38	0.699	0.944	-0.245
39	1.000	0.939	0.061
40	0.279	0.600	-0.321
41	0.602	0.841	-0.239
42	0.699	0.521	0.178
43	0.824	1.051	-0.227
44	1.000	1.285	-0.285
45	0.301	0.248	0.053
46	0.097	0.263	-0.167
47	0.398	0.246	0.152
48	0.046	0.165	-0.119
49	1.398	1.514	-0.116

50	1.699	1.365	0.334
51	1.155	1.169	-0.014
52	0.824	1.023	-0.199
53	1.301	1.216	0.085
54	1.523	1.499	0.024
55	1.699	1.552	0.147
56	2.000	1.603	0.397
57	1.523	1.455	0.068

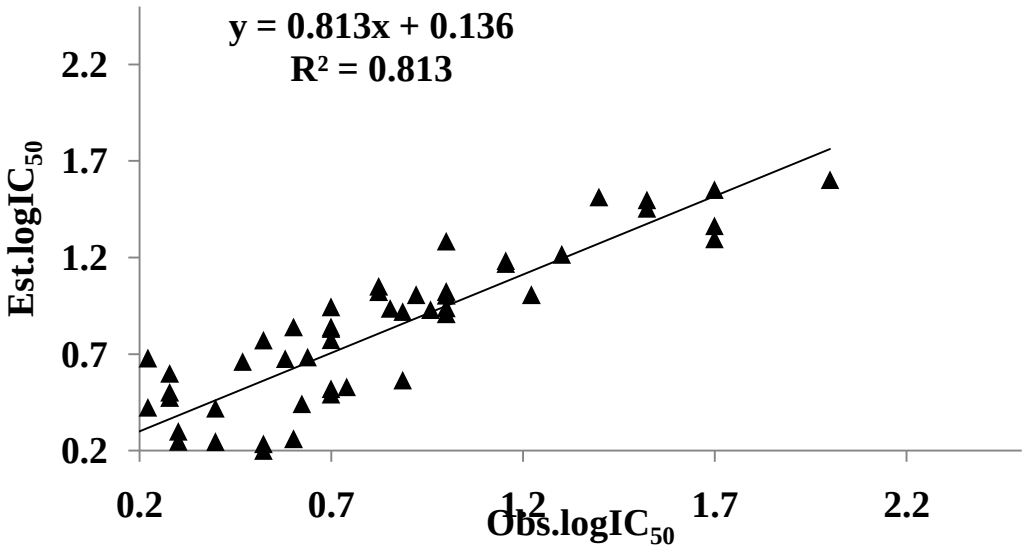


Fig: 1: Scatter plot between the observed and predicted activities of Model -42.

Table 7. Cross validation parameters for the proposed models

Model No.	Parameters used	PRESS	SSY	PRESS/SSY	R^2CV	S_{PRESS}	PSE
29	RDF085	4.7711	8.0201	0.5949	0.4051	0.2945	0.2893
33	RDF020e,RDF085v	4.1646	8.6267	0.4828	0.5172	0.2777	0.2703
36	RMor16m,DF020e,RDF085v	3.9136	8.8786	0.4408	0.5592	0.2717	0.2620
37	Mor16m,Mor32u,DF020e,RDF085v	3.0032	9.7880	0.3068	0.6932	0.2403	0.2295
41	Mor16m,Mor32u,DF020e,RDF040u, RDF085v	2.7474	10.0438	0.2735	0.7265	0.2321	0.2195
42	GATS3e,Mor16m,Mor32u,DF020e,RDF040u,RDF085v	2.3896	10.4016	0.2297	0.7703	0.2186	0.2048

Table 8. Variants Inflation parameters for the proposed models

Names of independent variables	Variance inflation	Tolerance	Eigen value	condition no.
GATS3e	1.91	0.52	2.80	1.00
RDF040u	1.95	0.51	1.06	2.65
RDF085v	1.62	0.62	0.73	3.85
RDF020e	1.69	0.59	0.68	4.15

Mor32u	1.39	0.72	0.40	7.05
Mor16m	1.08	0.93	0.34	8.28

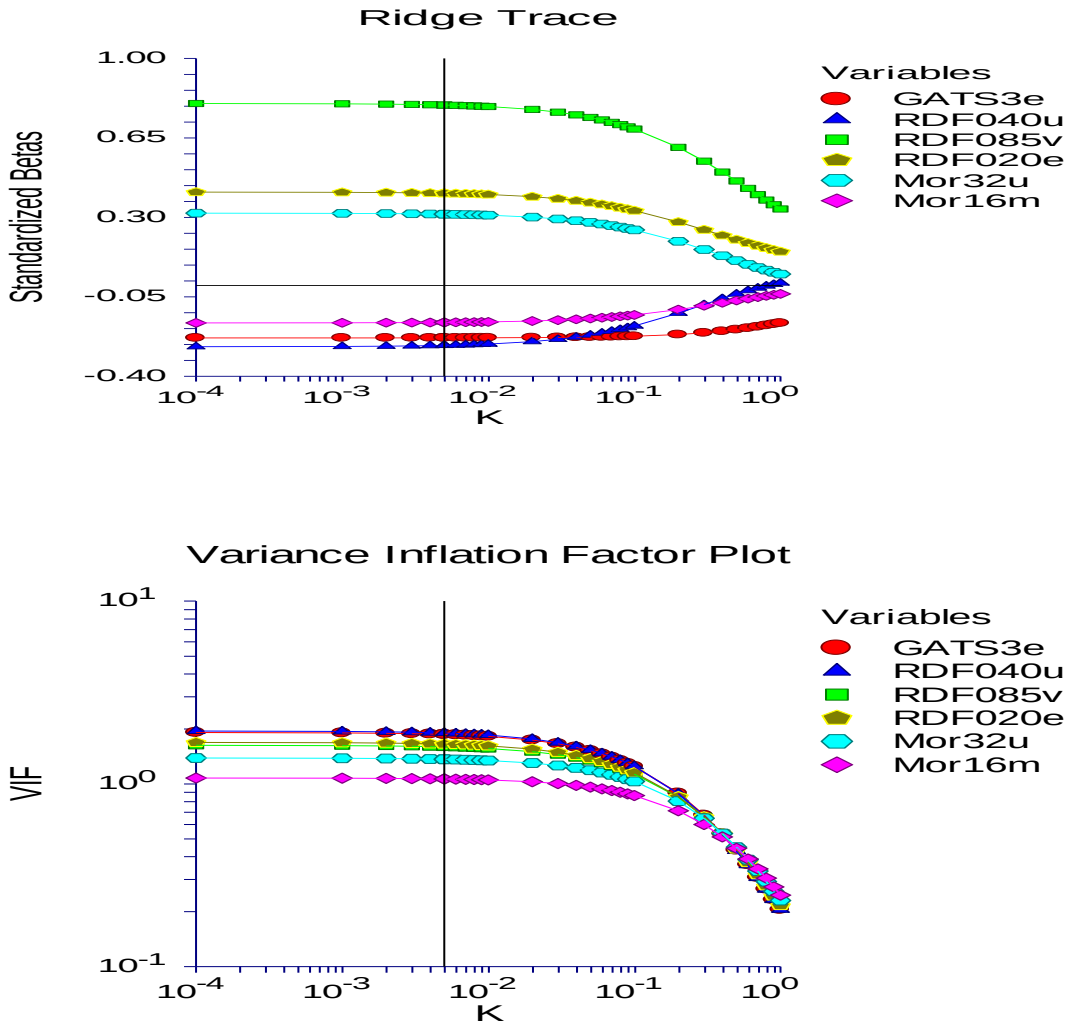


Fig: 2[a]Ridge trace and variance inflation factor plot for model-42.

[a]=J. Hintze, 2007. NCSS 2007. NCSS, LLC. Kaysville, Utah, USA. www.ncss.com

6. Conclusion

Molecular modeling studies were performed to design new more potent compounds to inhibit substituted pyrimidine for HCV NS5B inhibitory activity. Results reveal that the 2D-QSAR studies signify a positive contribution of Mor32u, RDF020e and RDF085v toward the biological activity, whereas negative contribution of Mor16m, GATS3e and RDF040u towards inhibitory activity, which means that low molecular weight compounds and low electronegative atoms should be considered for modeling the HCV NS5B inhibitors. Furthermore, visualization of the QSARmodel based on the 3D structure of the molecules under study provided details of the relationship between structure and activity. Furthermore, we hope that the current study provides better insights into the design of more potent pyrimidine analogs as NS5B inhibitory agent in the future before their synthesis.

Acknowledgement

We would like to thanks to Dr. V K Agarwal for providing softwares (Dragon 5.5) for statistical analysis. DC thanks DST-Fast Track for research funding.

References

1) Lavanchy, The global burden of hepatitis C. Liver Int. 29, (2009), 74.

2) F. X.Bosch, J.Ribes, RCleries, MDiaz,Epidemiology of hepatocellular carcinoma.Clin. Liver Dis. 9,(2005),191.

3), K. AMcGlynn, W. TLondon,Epidemiology and natural history of hepatocellular carcinoma.Best Pract. Res. Clin.Gastroenterol.19,(2005)

4) EASchaefer, RTChung,Anti-hepatitis C virus drugs in development.Gastroenterology. 142, (2012), 1340-1350.

5) C. D. Kwong et al,Novel substituted pyrimidines as HCV replication (replicase) inhibitors.Bioorg. Med. Chem. Lett.22, (2012),1160–1164.

6) Anamaya (New Delhi) and Springer VerlagQSAR and Molecular Modeling, (2011).

7) CS chem office version 7.0, Cambridge soft cooperation, software publisher association,1730 M Street, NW, Suite 700, Washington DC.

8) J. J. P. Stewart et al, MOPAC 2006 Version 7.1, Fujitsu Limited, Tokyo, Japan.

9), I. VTetko, J.Gasteiger, R.Todeschini, AMauri, D.Livingstone, P.Ertl, V. A.Palyulin, E. V.Radchenko, N. S.Zefirov, A. S.Makarenko, V. Y.Tanchuk, V. VProkopenko, Virtual computational chemistry laboratory - design and description, J. Comput. Aid. Mol. Des.19, (2005), 453-63.

10) VCCLAB, Virtual Computational Chemistry Laboratory, http://www.vcclab.org, 2005.

11) J. Hintze, 2007. NCSS 2007. NCSS, LLC. Kaysville, Utah, USA. www.ncss.com.

12) Naveen K Mahobia, Roshan DPatel, Naheed W.Sheikh, Sudarshan K.Singh, AchalMishra, RavindraDhardubey,Validation Method Used In Quantitative Structure Activity Relationship. Pharma Chemica. 2, (2010), 260-271.

13) RavichandranVeerasamy, HarishRajak, AbhishekJain, ShaliniSivadasan, Christopher P.Varghese, Ram KishoreAgrawal,Int. J. Drug. Des. Disc. 2, (2011), 511-519.

14) RTodeschini,et al .Handbook of Molecular Descriptors. Weinheim, Germany: Wiley-VCH; 2002.

15) J.H. Schuur, P. Selzer, J. Gasteiger,The Coding of the Three-Dimensional Structure of Molecules by Molecular Transforms and Its Application to Structure-Spectra Correlations and Studies of Biological Activity. J. Chem. Inf. Comput. Sci.36, (1996), 334-344.

16) JGasteiger, PSasowski, LSelzer Steinhauer, V. Steinhauer, Chemical Information in 3D Space. J. Chem. Inf. Comput. Sci.36, (1996), 1030-1037.

17) M.CHemmer, VSteinhauer,J. Gasteiger,Deriving the 3D structure of organic molecules from their infrared spectra. J. Vibrat. Spect. 19, (1999), 151 -164.

18) R.C Geary, lncorp.Statist. 5, (1954), 115-145.

19) P. Broto, G. Moreau, C. Vandicke, Eur. J. Med. Chem. 19, (1984),66-70.

20) G. Moreau, P. Broto, Nouv. J. Chim.4, (1980),359-360.

21) P.A.P. Moran, Biometrika. 37, (1950), 17-23.

22) SChaterjee, A. SHadi, BPrice, Regression Analysis by Examples,3rd Ed Wiley: New York, 2000

23) L. Pogliani,Amino Acids.6, (1994),141.

24) L. Pogliani, J. Phys. Chem. 100, (1996), 18065.