Science World Journal Vol 14(No 1) 2019
www.scienceworldjournal.org
ISSN 1597-6343
Published by Faculty of Science, Kaduna State University

Full Length Research Article

QUANTUM MODELLING ANALYSIS OF SOME POTENT INDOLE DERIVATIVES ON NS5B POLYMERASE INHIBITORS

Bello A. S.[1*], Uzairu A.[2], Ibrahim M. T.[2], Gatugel Y. A.[3]

[1]Department of Science Laboratory Technology Adamawa State Polytechnic Yola, Adamawa state Nigeria.
[2]Department of Chemistry Ahmadu Bello University, P.M.B. 1044, Zaria, Kaduna State Nigeria.
[3]Department of Basic Science College of Agriculture Ganye, Adamawa State Nigeria.

*Corresponding Author Email Address: sadiqdanburam@gmail.com Phone: +2348039194040

ABSTRACT
Viral hepatitis C infection is one of the main causes of the hepatitis after blood transfusion, hepatitis C virus (HCV) infection is a global health threat. The Non-structural viral protein (NS5B) is one of the best-studied polymerase which emerged as an attractive target for the development of novel therapeutics against hepatitis C virus. Quantitative structure-activity relationship studies (QSAR) was carried out on a series of indole derivatives as anti-hepatitis C inhibitors. Density Functional Theory (DFT) quantum chemical calculation method was used to find the optimized geometry of the indole inhibitors. Five types of molecular descriptors were used to derive a quantitative relation between indole activity and structural properties. The relevant molecular descriptors were selected by Genetic Function Algorithm (GFA). The best model (model 1) was validated and found to be statistically significant with squared correlation coefficient of R^2 of 0.969, R^2adj value 0.917, and Q^2 LOO 0.612 and the external validation was found to be R^2 pred. = 0.815. The proposed model has good stability, robustness, and predictability on verifying with internal and external validation. The physicochemical parameters are to be considered when improving the inhibitory activities of the indole derivatives against an enzyme that causes HCV (NS5B polymerase).

Keywords: HCV, DFT, NS5Bpolymerase, QSAR.

INTRODUCTION

Hepatitis means inflammation of the liver. It is caused by several mechanisms, including certain infectious agents (Balavignesh *et al.*, 2013). Viral hepatitis is caused by different type of viruses such as hepatitis A, B, C, D and E. Jaundice is one of the characteristic features of hepatitis disease and proper diagnosis could be made by testing the patient's sera for the deduction of antiviral antibodies (Balavignesh *et al*., 2013). Many people with Hepatitis C do not have symptoms and do not know they are infected. If symptoms occur, they can include: fever, feeling tired, not wanting to eat, upset stomach, throwing up, dark urine, grey-colored stool, joint pain, and yellow skin and eyes (Rustgi, 2007). Hepatitis C Virus (HCV) is a RNA-type virus and belongs to family of flaviridae and genus hepacivirus (Frese *et al*., 2003). The genome of HCV is about 9.6 kilo base pairs in length encoding a polyprotein of over 3000 amino acids which is cleaved by HCV genome proteases (Chinnaswamy *et al*., 2010). The protein NS5B polymerase is an RNA dependent RNA polymerase (Balavignesh *et al*., 2013). It has been clinically validated that NS5B is essential for HCV replication (Balavignesh *et al*., 2013).
At present, HCV vaccine is unavailable (Fauvelle et al., 2013) and (Law *et al*., 2013) and the standard of care (SOC) involves the combination of a protease inhibitor with pegylated α-interferon (PEG-IFN-α) and the oral nucleoside antiviral agent ribavirin (RVB) (Lü and Xue, 2011). However, there are some malpractices of this therapy, such as poorly tolerated, unsuccessful cured for all patients, high cost, long treatment period, low sustained virological response rates, undesirable side effects and so on (Haudecoeur *et al*., 2013). Therefore, it is meaningful in the HCV research area to look for and develop a novel specifically targeted antiviral structure. At present, finding drugs directly targeting nonstructural proteins has been focused on intensive research (Summa *et al*., 2004) and (Gao *et al*., 2010).
Computer drug design has been extensively used for drug discovery and development of drugs due to their extrusive advantages of time-consuming, cost-reducing, the high efficiency in silico screening and prediction of candidate drugs with advancement in computer techniques and simulation Software (Mohammad and Zohreh, 2013) over the traditional wet laboratory method. QSAR is a way of finding a simple equation that can be used to calculate some property from the molecular structure of a compound. QSAR attempt to correlate structural molecular features (descriptors) with physicochemical properties such as biological activities for a set of compounds, by means of statistical methods. The process of building a QSAR model is similar, apart from what type of property is being predicted (Roy *et al*., 2015). QSAR plays an important role in lead structure optimization and it can be predicted that QSAR methods will become essential for handling the huge amount of data associated with combinatorial chemistry. Quantitative Structure Activity Relationships (QSARs) are also a relationships between structural properties of chemical substances and another property. This other property can be a physico-chemical property or a biological activity, including the ability to cause toxic effects (Roy *et al*., 2015). This research was aim to develop various QSAR models using Genetic Function Algorithm (GFA) method for predicting the activities of some selected indole derivatives.

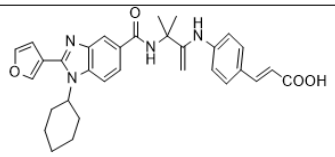
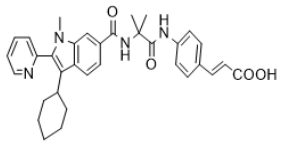
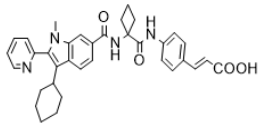
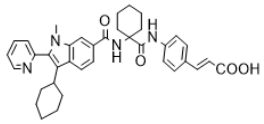
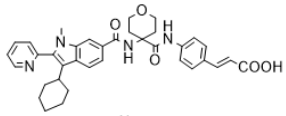
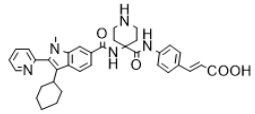
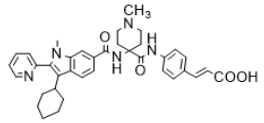
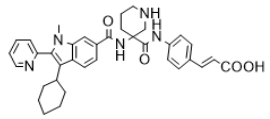
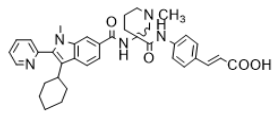
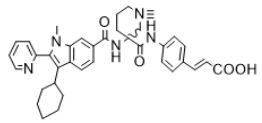
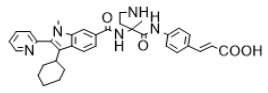
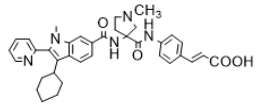
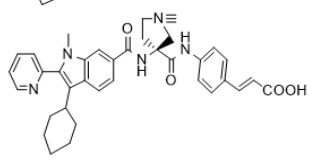
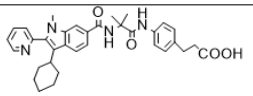
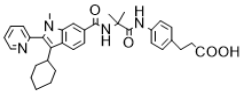
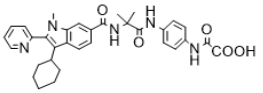
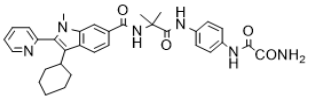
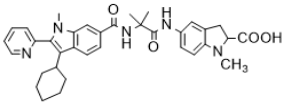
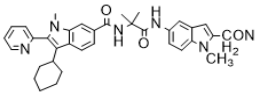
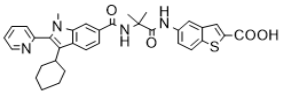
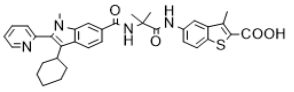
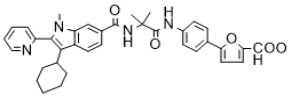
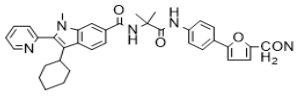
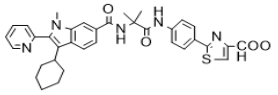
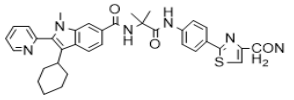
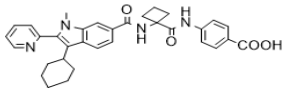
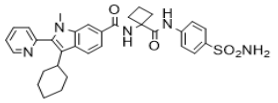
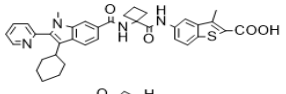
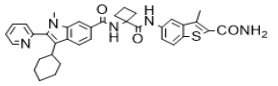
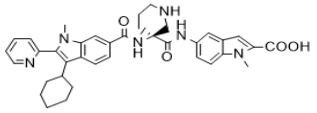
MATERIALS AND METHODS

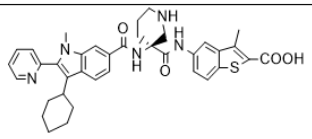
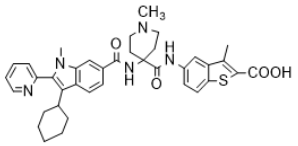
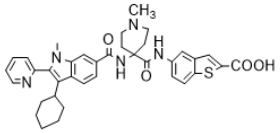
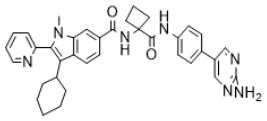
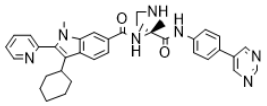
DATA SET.

Thirty-five 35 reliable indole derivatives were screen from the literature and used for the present research (Wei *et al*., 2016). The indole activities of the molecules measured as pIC_{50} (nM) were expressed in logarithmic scale as pIC_{50} (pIC_{50} = log 1/ pIC_{50}) and then used as dependent variable. Consequently, the data was correlated linearly to the independent variable/descriptors. The observed structures, the biological activities, predicted values and residuals value of the training and test sets of these compounds were presented in Table 1.

Science World Journal Vol 14(No 1) 2019
www.scienceworldjournal.org
ISSN 1597-6343
Published by Faculty of Science, Kaduna State University

Table 1. Biological activities, predicted value and Residual value of training and test set indole derivatives.

	Structure	pIC_{50} (nM)	Predicted values	Residual values
1		1.39	1.39	-0.00
2		1.55	1.69	-0.14
3		1.71	1.75	-0.04
4		2.28	2.07	0.20
5		1.83	1.46	0.36
6		1.11	1.34	-0.23
7		1.00	1.47	-0.47
8		1.60	1.50	0.09
9		1.63	2.01	-0.38
10		2.26	1.78	0.47
11		1.74	1.76	0.02
12		1.81	1.46	0.34
13		1.38	1.53	-0.15
14		1.36	1.67	-0.31
15		2.12	1.76	0.35
16		1.53	1.59	-0.06
17		2.09	1.99	0.09
18		1.87	1.67	0.19
19		2.21	2.14	0.06
20		1.36	1.34	0.01
21		1.54	1.60	-0.06
22		2.18	2.38	-0.20
23		2.51	2.57	-0.06
24		2.18	2.08	0.09
25		2.50	2.49	8.710
26		1.96	1.99	0.03
27		2.49	2.45	0.03
28		2.08	1.88	0.19
29		1.93	1.98	-0.05
30		1.64	1.67	-0.03

	Structure	pIC50 (nM)	Predicted values	Residual values
31		1.54	1.41	0.12
32		1.23	1.38	-0.15
33		1.43	1.34	0.08
34		2.87	2.93	0.06
35		1.80	1.80	-0.00

Molecular modeling

All structures were constructed using ChemDraw Ultra 12.0 Software and save as cdx file format, the structures were converted to 3D using Spartan 14.0 version 1.1.2 software, molecular mechanics force field (MM+) calculation was carried out to minimize the energy of the molecules prior to the quantum chemical calculations. Density functional theory with B3LYP/6-311G* was employed for complete geometry optimization of the drawn structures to obtain the lowest energy for all the inhibitors. The sdf format of the optimized structures that were from the Spartan'14 version 1.1.2 software package (Abdulfatai *et al.*, 2017) was converted to PaDEL-Descriptor version 2.18 toolkits (Yap, 2011) where the calculation of 1D, 2D, and 3D descriptors took place.

Computational method

For validated QSAR models, the descriptors (1D-3D) generated from the PaDEL version 2.18 toolkits (Yap, 2011) was divided into training and test sets. The training set was used to generate the model, while test set was used for external verification of the model (Kennard and Stone, 1969). The relationship between the activity values of the indole molecules against NS5B polymerase and calculated descriptors was obtained through correlation analysis using material studio software version 8. The Pearson's correlation matrix was used as a qualitative model in order to determine appropriate descriptors for regression analysis.

The descriptors from PaDEL version 2.18 toolkits (Yap, 2011) were analyzed for regression analysis with experimentally determined activities as the dependent variable and the selected descriptors as the independent variables using Genetic Function Algorithm (GFA) method in Material Studio software version 8. The models were registered based on Friedman's Lack of Fit (LOF). In GFA algorithm, the individual or model is represented as a one-dimensional bit. The characteristic of GFA is that it can create a population of models instead of a single model. GFA algorithm developed better models than those made using stepwise regression methods. Thus, the models were estimated using the LOF, which was measured using a slight formula of the original Friedman formula, so that the better score can be received. The revised formula of LOF (Khaled, 2011) is as in equation 1:

$$LOF = \frac{SSE}{\left(1-\frac{c+dp}{M}\right)^2} \quad (1)$$

SSE is the sum of squares of errors c is the number of terms in the model, unlike the fixed term d is a user-defined smoothing parameter, p is the total number of descriptors contained in all model terms (ignoring the constant term), and M is the number of samples in the training set.

Quality assurance of the model

The reliability and predictive power of QSAR models were evaluated by internal and external validation parameters.

Internal and external validations

The internal and external validation parameters were compared with the minimum recommended value for the evaluation of the quantitative QSAR model (Veerasamy *et al.*, 2011) as shown in Table 2. The R^2 describes the fraction of the total variation attributed to the model, equation 2 shows the formula for calculating the internal and external validation.

$$R^2 = 1 - \frac{\Sigma\left(Y_{obs}-Y_{pred}\right)^2}{\Sigma\left(Y_{obs}-\bar{Y}_{training}\right)^2} \quad (2)$$

where Yobs, Ypred, and Ytraining are the experimental property, the predicted property, and the mean experimental property of the samples in the training set respectively.(Veerasamy *et al.*, 2011). Adjusted R^2 value varies directly with the increase in a number of repressors i.e. descriptors; thus, R^2 cannot be a useful measure of the goodness of model fitness. Therefore R^2 is adjusted for the number of explanatory variables in the model. R^2 adj is defined as in equation 3.

$$R^2 adj = 1 - (1-R^2)\frac{n-1}{n-p-1} = \frac{(n-1)\,R2-p}{n-p+1} \quad (3)$$

Where n is the number of training compounds, p= number of independent variables in the model.

The leave one out cross validation coefficient (Q2) is given by equation 4.

$$Q^2 = 1 - \frac{\Sigma(Yp-Y)^2}{\Sigma(Y-Ym)^2} \quad (4)$$

where Yp and Y are the predicted and observed activity respectively of the training set and Ym is the mean activity value of the training set (Jalali-Heravi and Kyani, 2004).

Applicability domain

Applicability Domain (AD) is the chemical descriptor space incorporated by special training collection of chemicals. The applicability domain of the developed models was assessed in order to specify the scope of their proposed models by defining the model limitations with respect to its structural domain and response area. Leverage refers to the compound's distance from the centroid of X. The leverage of the compound in the defined original variable space is as follows:

$$hi = X_i^T (X^T X)^{-1} X^i \quad (5)$$

The warning leverage (h*) is defined as follows:

$$hi = \frac{3(P+1)}{N} \quad (6)$$

N is the number of training compounds, and p is the number of predictor variables. Xi is the descriptor vector of the considered

compound and X is the descriptor matrix derived from the training set descriptor values. Figure 2 shows that six of the test set compound fall outside the domain of the model (the warning leverage limit is 0.4), hence they are accepted as influential

Table 2. General minimum recommended value for the evaluation of the quantitative QSAR model

Name	Symbols	Value
R^2	Coefficient of determination	≥0.5
$P_{(95\%)}$	Confidence interval at 95% confidence level	<0.05
Q^2	Cross-validation coefficient	≥0.5
$R^2 - Q^2$	Difference between R^2 and Q^2	≤0.3
$N_{ext.\ Test\ set}$	Minimum number of external test set	≥5
R^2_{ext}	Coefficient of determination for external test set	≥0.5

The closer the value of R^2 is to 1.0, the better the regression equation explains the Y variable.

RESULTS AND DISCUSSION

Five developed QSAR models were generated out of which the best model (model 1) was identified and reported due to the statistical importance. Table 3 shows the name and definitions of the descriptors used in the QSAR model. Table 4 gives the result of the Genetic Function Algorithm (GFA) of model 1 produced from Material Studio software version 8. The minimum recommended value for validation of the generally acceptable QSAR model was consistent with the parameters of model 1. Based on the generated statistics, Model 1 was selected and reported as the best QSAR model.

pIC_{50} = 0.223637543 * **ATSm2** - 0.214761052 ***ATSm3** - 0.214761052 * **BCUTw-1l** + 0.229604208 * **MLFER_L**+ 28.226827661 * **PetitjeanNumber** + 19.85022, N = 35, $R^2_{ext.}$ = 0.81500, R^2 = 0.969577, R^2_{adj} = 0.917208, LOF = 0.210472, Min expt. Error for non-significant LOF (95%) = 0.173480

Table 3. List of some physiochemical descriptors used for the best model

S/N	Symbols	Name of descriptors	Class
1	ASTm2	ATs autocorrelation descriptors weighted by scale atomic mass.	2D
2	ASTm3	ATs autocorrelation descriptors weighted by scale atomic mass.	2D
3	MLFER_L	Overall or summation solute hydrogen bond acidity.	2D
4	BCUTw-1l	High lowest atom weighted BCUTS	2D
5	PetitjeanNumber	Petitjean topological shape index	2D

From Table 4, the square of the correlation coefficient (R^2) describes the fraction of the total variation attributed to the model. The closer value of R^2 is to 1.0, the better the regression equation explains the Y variable. R^2 is the most commonly used internal validation indicator.

The correlation matrix was performed on the descriptors of model 1 and found to be highly correlated which means that the descriptors used to build the model are good (Table 5).

Table 4: Validation of the genetic function approximation from Material Studio

Validation parameters	EQ 1	EQ 2	EQ 3	EQ 4	EQ 5
Friedman LOF	0.210472	0.210472	0.214247	0.214567	0.215109
R-squared	0.969577	0.969577	0.965444	0.965093	0.964500
Adjusted R-squared	0.917208	0.917208	0.912136	0.911705	0.910977
Cross validated R-squared	0.612300	0.612500	0.658755	0.658730	0.655919
Significant Regression	Yes	Yes	Yes	Yes	Yes
Significance-of-regression F-value	14.695280	14.695280	14.358828	14.330836	14.283646
Critical SOR F-value (95%)	2.684036	2.684036	2.684036	2.684036	2.684036
Replicate points	0	0	0	0	0
Computed experimental error	0.000000	0.000000	0.000000	0.000000	0.000000
Lack-of-fit points	22	22	22	22	22
Min expt. error for non-significant LOF (95%)	0.173480	0.173480	0.175029	0.175160	0.175381

Table 5: Pearson's correlation matrix for descriptors used in QSAR model for the activities of anti-hepatitis C

	ATSm2	ATSm3	BCUTw-1l	MLFER_L	PetitjeanNumber
ATSm2	1				
ATSm3	0.813173	1			
BCUTw-1l	-0.22463	-0.16066	1		
MLFER_L	0.266894	0.601166	0.442264	1	
topoShape	0.324588	0.595204	0.228893	0.801893	1

Figure 1 shows that the developed model is stable and the residuals on both sides of zero are randomly propagated. Figure 2 also shows the Williams' plot that is the plot of standardized residuals against leverages value for each compound in the dataset resulting in the discovery of 6 influentials compounds (pIC50 of 2.26, 1.63, 2.08, 2.49, 1.96, and 2.87). The influencials compound were out of the applicability domain of the model because their leverage values are higher than the warning leverage value (h* = 0.64), and the high leverage value is responsible for influencing the performance of the model.

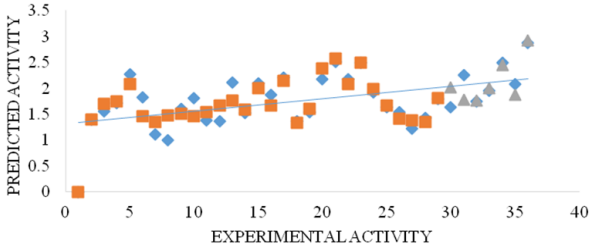
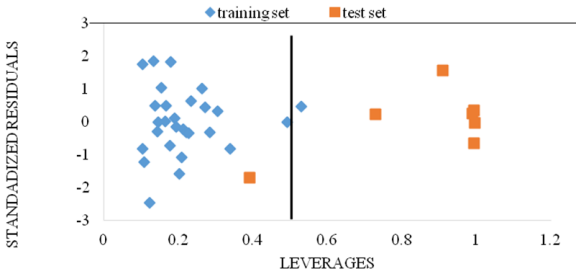


Figure 1: The plot of the Experimental and predictive activity of both training and test set of the best model (1)

Figure 2, Applicability Domain (AD) of the chemical descriptors space of both the training set and test set. The applicability domain of the developed models was assessed in order to specify the scope of their proposed models by defining the model limitations with respect to its structural domain and response space.

Science World Journal Vol 14(No 1) 2019
www.scienceworldjournal.org
ISSN 1597-6343
Published by Faculty of Science, Kaduna State University



h* = 0.64

Figure 2: Williams plot, the plot of the standardized residuals versus the leverage value of both the training set and test set of model 1

Table 6 shows the corresponding VIF values of the descriptors in which three of the variables have VIF values of less than 10, indicating that the obtained model has statistical significance and the descriptors were found to be reasonably orthogonal (Riu and Rius, 1996).

Table 6: Variance Inflation Factor (VIF) Statistic for the Descriptors in Model 4.1

S/NO	Dependent Variable	VIF
1	ASTm2	36.2038
2	ASTm3	45.80344
3	BCUTw-1l	1.139135
4	MLFER_L	3.992498
5	PetitjeanNumber	5.181799

Interpretation of descriptors in model 1

ASTm2 and **ASTm3** have are defined as 2D correlated descriptors ATs autocorrelation descriptor weighted by scale atomic mass. **MLFER_L** is 2D MLFER Descriptors which is defined as the overall or summation solute hydrogen bond acidity. **BCUTw-1l** is also 2D PaDEl rotatable bonds count Descriptors and is defined as the high lowest atom weighted BCUTS. **PetitjeanNumber** is another 2D Descriptor and is defined as Petitjean topological shape index.

The positive coefficients in model 1 implies that increase in physiochemical parameters of **BCUTw-1l** (High lowest atom weighted BCUTS), **PetitjeanNumber** (Petitjean topological shape index) and **MLFER_L** (Overall or summation solute hydrogen bond acidity) will increase the anti-Hepatitis C activities (pIC_{50}) against NS5B polymerase an enzyme responsible for Hepatitis C. Also, negative coefficients of **ATSm2** (ATs autocorrelation descriptors weighted by scale atomic mass) and **ATSm3** (ATs autocorrelation descriptors weighted by scale atomic mass) will decrease the inhibitory activities of indole derivatives against NS5B polymerase enzyme.

Comparison of observed and predicted PIC_{50} of model 1

The comparison of the predicted PIC_{50} of the model 1 with its experimental values is presented in Table 1. The low residual values confirmed the high predictive power of the model. The actual, predicted and residual pIC_{50} values of the test set compounds are presented in the Table 7, which shows a good agreement between the test set compounds.

Table 7: Calculation of predictive R^2 of model 1

pIC50	ATSm2	ATSm3	BCUTw-1l	MLFER_L	PetitjeanNumber	Predicted IC50	Residuals
1.63	54.322	78.868	11.9	25.962	0.909091	2.014101	0.384101
2.26	54.322	78.868	11.95281	25.573	0.909091	1.78088	-0.47912
1.74	52.156	75.868	11.99423	25.31	0.909091	1.767719	0.027719
1.96	48.824	70.536	11.9	23.742	0.9	1.996824	0.036824
2.49	58.055	78.148	11.9	23.864	0.9	2.452459	-0.03754
2.08	57.163	86.215	11.9	26.554	1	1.880029	-0.19997

The stability, reliability and robustness of the generated model 1 were confirmed by predictive R^2 (Table 8). The predictive R^2 for model 1 were in agreement with the standard minimum recommended value for the evaluation of the quantitative QSAR model.

Table 8: External validation of Model 1

CP/CP/NO	Y(te)	Ypred(te)	[Ypred(te) _ Y(te)]2	Ym(tr)	[Y(te) _ Ym(tr)]2
1.63	2.014101	0.384101	0.147534	1.7385	0.011772
2.26	1.78088	-0.47912	0.229556	1.7385	0.271962
1.74	1.767719	0.027719	0.000768	1.7385	2.25E-06
1.96	1.996824	0.036824	0.001356	1.7385	0.049062
2.49	2.452459	-0.03754	0.001409	1.7385	0.564752
2.08	1.880029	-0.19997	0.039988	1.7385	0.116622
			Σ =0.424241		Σ =1.280292

Therefore, Pred. R^2 = 1 - (0.4242/1.2802) = 0.926606.

The Y-randomization test shows that the model is not obtained by chance and it is robust because it has significantly low R^2 and Q^2 values for several trials (Table 9).

Table 9: Y-Randomization test of model 1.

Model	R	R^2	Q^2
Original	0.780578	0.609303	0.473827
Random 1	0.331494	0.109888	-0.32188
Random 2	0.214691	0.046092	-0.57872
Random 3	0.177125	0.031373	-0.36707
Random 4	0.546486	0.298647	-0.02131
Random 5	0.437796	0.191665	-0.27685
Random 6	0.310253	0.096257	-0.38744
Random 7	0.398603	0.158885	-0.57834
Random 8	0.303292	0.091986	-0.42802
Random 9	0.306169	0.09374	-0.30068
Random 10	0.279947	0.078371	-0.36735
Random Models Parameters			
Average r :	0.330586		
Average r^2 :	0.11969		
Average Q^2 :	-0.36277		
cRp^2 :	0.551961		

Conclusion

In this research, QSAR model was generated with descriptors (ASTm1, ASTm3, BCUTw-1l MLFER_L and PetitjeanNumber) which were correlated with biological activities of indole derivatives and have the optimum model exhibited statistically significant results of squared correlation coefficient (R^2) of 0.969, adjusted squared correlation coefficient (R^2 adj) value 0.917, and cross-validation coefficient (Q^2) LOO 0.612. The best model that is model (1) was subjected to external validation and was found to be R^2 pred. = 0.815 which shows the stability, reliability and robustness of the model. The low residual values of the test set compounds confirmed the high predictive power of the model. The VIF values of the descriptors shows that three of the variables have their values less than 10, indicating that the obtained model has statistical significance and the descriptors were found to be reasonably orthogonal. The Y-randomization test shows that the model is not obtained by chance. The result of QSAR model study provides a good approach for pharmaceutical and medicinal researchers to design new anti-hepatitis C agent against NS5B polymerase receptor.

REFERENCES

Abdulfatai, U., Uzairu, A., And Uba, S., (2017) Quantitative Structure-Activity Relationship And Molecular Docking Studies Of A Series Of Quinazolinonyl Analogues As Inhibitors Of Gamma Amino Butyric Acid Aminotransferase. *Journal Of Advanced Research, 8*(1), 33-43.

Arthur, D. E, (2017) Toxicity Modelling Of Some Active Compounds Against K562 Cancer Cell Line Using Genetic Algorithm-Multiple Linear Regressions. *Journal Of The Turkish Chemical Society, Section A: Chemistry, 4*(1), 355-374.

Balavignesh, V., Srinivasan, E., Ramesh Babu, N., And Saravanan, N., (2013) Molecular Docking Study On Ns5b Polymerase Of Hepatitis C Virus By Screening Of Volatile Compounds From Acacia Concinna And Admet Prediction. *International Journal Of Pharmacy & Life Sciences, 4*(4).

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 Adaptation And Treatment, 2*, 73-89.

Fauvelle, C., Lepiller, Q., Felmlee, D. J., Fofana, I., Habersetzer, F., Stoll-Keller, F., Baumert, T. F., And Fafi-Kremer, S., (2013) Hepatitis C Virus Vaccines–Progress And Perspectives. *Microbial Pathogenesis, 58*, 66-72.

Frese, M., Barth, K., Kaul, A., Lohmann, V., Schwärzle, V., And Bartenschlager, R., (2003) Hepatitis C Virus Rna Replication Is Resistant To Tumour Necrosis Factor-A. *Journal Of General Virology, 84*(5), 1253-1259.

Gao, M.; Nettles, R. E.; Belema, M.; Snyder, L. B.; Nguyen, V. N.; Fridell, R. A.; Serrano-Wu, M. H.; Langley, D. R.; Sun, J.-H.; And O'boyle Ii, D. R.; (2010)Chemical Genetics Strategy Identifies An Hcv Ns5a Inhibitor With A Potent Clinical Effect. *Nature, 465*(7294), 96.

Haudecoeur, R.; Peuchmaur, M.; Ahmed-Belkacem, A.; Pawlotsky, J. M.; And Boumendjel, A.; (2013) Structure–Activity Relationships In The Development Of Allosteric Hepatitis C Virus Rna-Dependent Rna Polymerase Inhibitors: Ten Years Of Research. *Medicinal Research Reviews, 33*(5), 934-984.

Jalali-Heravi, M., And Kyani, A., (2004) Use Of Computer-Assisted Methods For The Modeling Of The Retention Time Of A Variety Of Volatile Organic Compounds: A Pca-Mlr-Ann Approach. *Journal Of Chemical Information And Computer Sciences, 44*(4), 1328-1335.

Kennard, R. W., And Stone, L. A., (1969)Computer Aided Design Of Experiments. *Technometrics, 11*(1), 137-148.

Khaled, K., (2011) Modeling Corrosion Inhibition Of Iron In Acid Medium By Genetic Function Approximation Method: A Qsar Model. *Corrosion Science, 53*(11), 3457-3465.

Law, L. M. J., Landi, A., Magee, W. C., Tyrrell, D. L., And Houghton, M., (2013)Progress Towards A Hepatitis C Virus Vaccine. *Emerging Microbes & Infections, 2*(11), E79.

Lü, W., And Xue, Y., (2011)Prediction Of Hepatitis C Virus Non-Structural Proteins 5b Polymerase Inhibitors Using Machine Learning Methods. *Acta Physico-Chimica Sinica, 27*(6), 1407-1416.

Mohammad, H., And Zohreh, A., (2013)In-Silico Prediction Of Rgs4 Inhibitory Activity Of Sometiadiazolidinone. *Int J Med Pharm, 1*. Riu, J.; And Rius, F. X.; Assessing The Accuracy Of Analytical Methods Using Linear Regression With Errors In Both Axes. *Analytical Chemistry, 68*(11), 1851-1857.

Roy, K., Kar, S., And Das, R. N., (2015.) *Understanding The Basics Of Qsar For Applications In Pharmaceutical Sciences And Risk Assessment*: Academic Press 4(3): 45-49.

Rustgi, V. K., (2007)The Epidemiology Of Hepatitis C Infection In The United States. *Journal Of Gastroenterology, 42*(7), 513-521.

Summa, V., Petrocchi, A., Matassa, V. G., Taliani, M., Laufer, R., De Francesco, R., Altamura, S., And Pace, P., (2004) Hcv Ns5b Rna-Dependent Rna Polymerase Inhibitors: From A, Γ-Diketoacids To 4, 5-Dihydroxypyrimidine-Or 3-Methyl-5-Hydroxypyrimidinonecarboxylic Acids. Design And Synthesis. *Journal Of Medicinal Chemistry, 47*(22), 5336-5339.

Veerasamy, R., Rajak, H., Jain, A., Sivadasan, S., Varghese, C. P., And Agrawal, R. K., (2011) Validation Of Qsar Models-Strategies And Importance. *International Journal Of Drug Design & Discovery, 3*, 511-519.

Wei, Y., Li, J., Qing, J., Huang, M., Wu, M., Gao, F., Li, D., Hong, Z., Kong, L., And Huang, W., (2016) Discovery Of Novel Hepatitis C Virus Ns5b Polymerase Inhibitors By Combining Random Forest, Multiple E-Pharmacophore Modeling And Docking. *Plos One, 11*(2), E0148181.

Yap, C. W., (2011) Padel-Descriptor: An Open Source Software To Calculate Molecular Descriptors And Fingerprints. *Journal Of Computational Chemistry, 32*(7), 1466-1474.