Bayero Journal of Pure and Applied Sciences, 11(2): 57 - 71
Received: April, 2018
Accepted: November, 2018
ISSN 2006 – 6996

QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP AND MOLECULAR DOCKING STUDIES OF SOME SERIES OF IMIDAZOLE DERIVATIVES AS ANTI-HEPATITIS C DRUG

Bello, A.S.[1*], Uzairu, A.[1], Shallangwa, G.A.,[1] Ibrahim, A. and [1]Gatugel, Y. A.[2]
[1]Department of chemistry Ahmdu Bello University, P.M.B. 1044, Zaria, Kaduna state Nigeria.
[2]Department of basic science Adamawa State Collage of Agriculture Ganye, Nigeria
Corresponding author: sadiqdanburam@gmail.com, 08039194040.

ABSTRACT
Hepatitis C virus (HCV) NS5B RNA-depended-RNA-polymerase (RdRp) is an essential enzyme in HCV viral replication and has no functional equivalent in mammalian cells. In silico study was carried out to develop a Quantitative structure activity-relationship and molecular docking on some selected imidazole derivatives as anti-hepatitis C compounds. Density functional theory with B3LYP/6-311G* was employed for complete geometry optimization Five QSAR models were generated using Genetic Function Algorithm (GFA) of the material studio software version 8, in which model one (1) was selected as the best model and reported based on the validation parameter with the squared correlation coefficient (R^2) of 0.7114. Adjusted squared correlation coefficient (R^2 adj) value 0f 0.6458 and cross-validation coefficient (Q^2) LOO 0.5810. The best model that is model (1) was subjected to external validation and was found to be R^2 pred. = 0.5729. The result obtained from molecular docking studies shows that the compound with the best binding affinity of -10.7 Kcal/mol formed hydrogen bonding of (GLN 446 and GLU143) and hydrophobic interaction with the amino acid residues of the Non-Structural 5B polymerase(NS5B polymerase) receptor. The QSAR model and molecular docking results propose the direction of designing new imidazole derivatives agent with better activity against the NS5B polymerase target site.
Keywords: Binding affinity, HCV, imidazole, molecular docking, NS5B polymerase, QSAR.

INTRODUCTION

Hepatitis C Virus (HCV) infection is a global health threat. Currently about 180 million individuals have been chronically infected with HCV according to the database of World Health Organization (WHO) (Lavanchy, 2009) and more than 350,000 people die every year from hepatitis-related liver diseases caused by HCV infection, such as cirrhosis (Marcellin *et al.*, 2002), liver failure, and hepatocellular carcinoma (HCC) (Mas *et al.*, 2009). Its progression is slow, and symptoms are mild, those features make it a stealth epidemic, and most infections progress to the chronic state that persists for decades. (Shepard *et al.*, 2005). The HCV NS5B polymerase, and RNA dependent RNA polymerase (RdRp) are central enzyme in the replication of the virus, so the research and development of efficient NS5B polymerase inhibitors provide a comprehensive strategy for antiviral therapy against HCV infection.

Currently, a vaccine against HCV is unavailable (Fauvelle *et al.*, 2013) and (Zhao *et al.*, 2015). The traditional method named as the standard of care (SOC), is a combination of regular pegylated α-interferon (PEGeIFNea) injections with oral Ribavirin (RVB) and is analytically used for the treatment of HCV infection. However, it is associated with various side effects and yields at best a 50% sustained virological response (SVR) only for genotype 1 infected patients (Hadziyannis *et al.*, 2004). Therefore, it is very important to develop new anti-HCV drugs with promising activity and less toxic.

Computer drug designs have been extensively used for drug discovery and development 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.

Quantitative structure-activity relationship (QSAR) are predictive mathematical models correlating one or more piece of response data about chemicals, with the information numerically encoded in the form of descriptors. Various statistical tools, including regression and classification-based strategies, are used to analyze the response and chemical data and their relationship (Roy *et al.*, 2015). Molecular docking is one of the most frequently used methods in structure-based drug design SBDD because of its ability to predict, with a substantial degree of accuracy, the confirmation of small-molecule ligands within the appropriate target binding site (Meng *et al.*, 2011). This research was aim to develop various QSAR models using Genetic Function Algorithm (GFA) method for predicting the activities of some selected imidazole derivatives and to predict the strength of interactions between imidazole derivatives (inhibitors) and NS5B RpRd (PDB code 1CSJ), an enzyme that is responsible for Hepatitis C.

MATERIALS AND METHODS

Thirty-five (35) imidazole derivatives compounds were selected from the literature (Zhao *et al.*, 2015). The compounds were divided into a training set (80) and test set (20) to validate the model. The activities of imidazole compounds were measured in IC_{50} (nM) were expressed as logarithmic scale as pIC_{50} (pIC_{50} =log1/IC_{50}) was used as dependent variable, consequently correlating the data linearly with the independent variable/ descriptors. The observed structures and the biological activities of these compounds are presented in Table 1.

Structure Optimization

ChemDraw software version 12.0.2 (Li *et al.*, 2004) were used to draw the structure of the molecules as presented in Table 1. Spartan 14 Version 1.1.4 software was used to optimize the imidazole compound in which 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.

Molecular descriptor calculation

Padel descriptor software version 2.18toolkits (Yap, 2011) was used to generate molecular descriptors for all the thirty-five (35) molecules of the inhibitory compound.

Dataset Division

Kennard–Stone Algorithm was used to divide the dataset into the training set and test set (Kennard and Stone, 1969). The training set contains 80% of the dataset which was used to build the model and the remaining 20% which is the test set was used to validate the build model.

Model Building

Genetic Function Algorithm (GFA) of material studio software version 8 was used to determine the internal validation parameters. The number of descriptors in the regression equation was 5, and Population and Generation were set to 1000 and 1500 respectively. The models were scored based on Friedman's Lack of Fit (LOF) which measured the fitness score of the model. The revised formula of LOF (Khaled, 2011) is as follows:

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

Where SSE is the sum of squares of errors, c is the number of terms in the model, other than the constant 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 (Khaled, 2011).

Quality assurances of the model

The fitness, reliability, and predictability of the developed QSAR models were evaluated by the internal and external validation parameters. Table 2 shows the minimum recommended values for the validation parameters (internal and external) for generally acceptable QSAR model.

Internal and external validations

The square of the correlation coefficient (R^2) describes the fraction of the total variation attributed to the model. The closer the R^2 value to 1.0, the better the model generated. R^2 is expressed as:

$$R^2 = 1 - \frac{\Sigma\,(\text{Yobs}-\text{Y pred})2}{\Sigma\,(\text{Yobs}-\text{Ytraining})2} \qquad (2)$$

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 (R^2 adj) 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. The adjusted R^2 is defined as follows:

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

Where p and n are a number of descriptors in the model and number compounds that made up the training set (Abdulfatai *et al.*, 2017).

The cross-validation coefficient (Q^2_{cv}) was used to determine the strength of the QSAR model to predict the activity of a new compound. The cross-validation coefficient (Q^2_{cv}) is defined as:

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

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

External validation of the developed model was assessed by the value R_{test}^2value. The R_{test}^2 is defined by as:

$R_{test}^2 = 1 - \frac{\Sigma(Ypred_{test}-Yexp_{test})^2}{\Sigma(Ypred_{test}-\bar{Y}_{training})^2}$ (5)

Where $Ypred_{test}$and $Yexp_{test}$are the predicted and experimental activity test set. While $\bar{Y}_{training}$is the training set mean values of the experimental activity.

Applicability domain

The model built by QSAR was evaluated based on the approach of the applicability domain in order to establish the model that is robust and reliable and that predict the activities of the inhibitory compounds.(Tropsha *et al.*, 2003). Leverage indicates a compound's distance from the centroid of X. The leverage of a compound in the original variable space is defined as follows:

$hi = X_i^T (X^T X)^{-1} X^i$ (6)

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

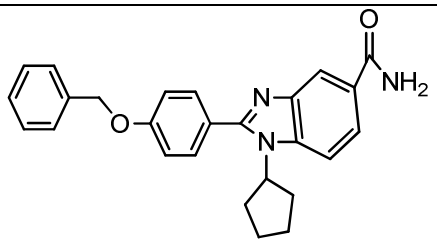
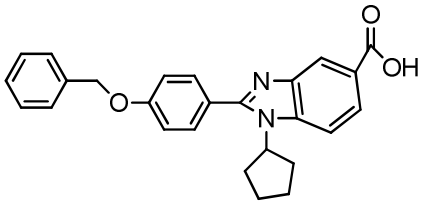
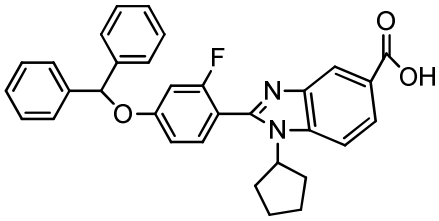
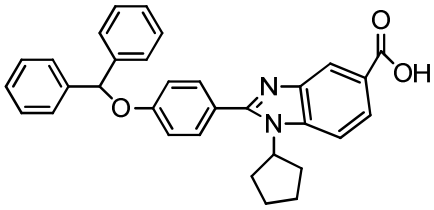
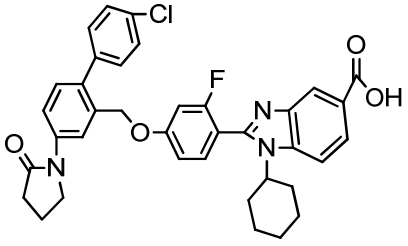
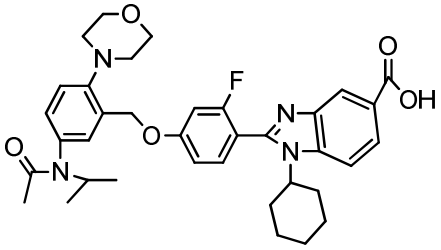
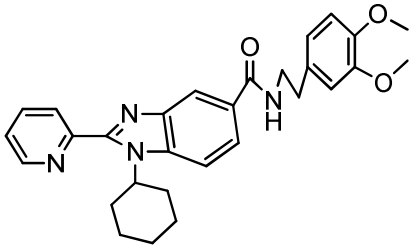
$Hi = \frac{3(P+1)}{N}$ (7)

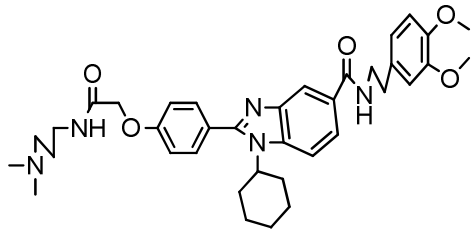
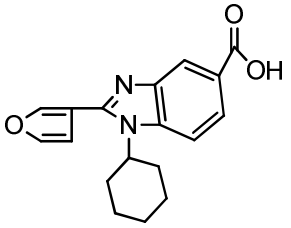
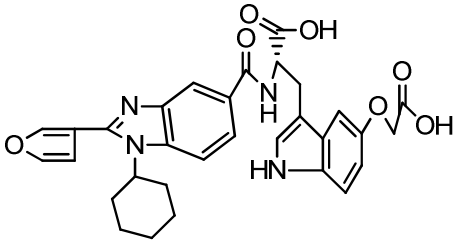
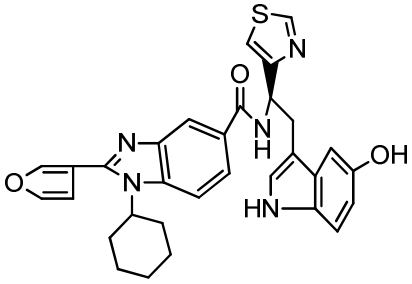
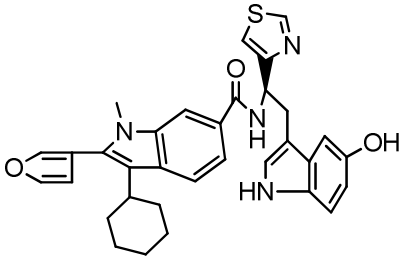
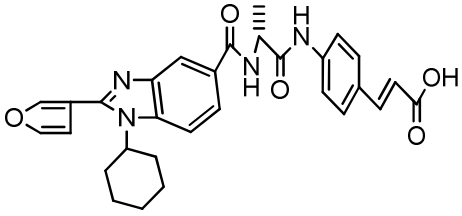
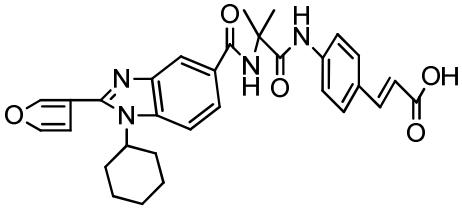
Where N is the number of training compounds, and p is the number of predictor variables. Where Xi is the descriptor vector of the considered compound and X is the descriptor matrix derived from the training set descriptor values. Fig. 3, shows only three (3) of the test set fall inside the domain of the model (the warning leverage limit is 0.7), hence they are accepted as Y influential.

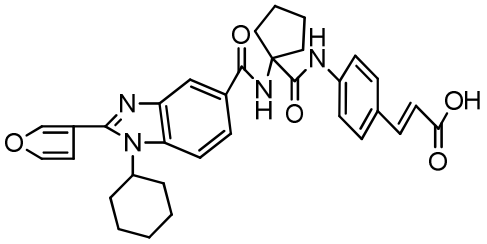
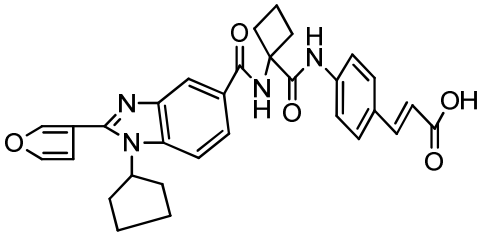
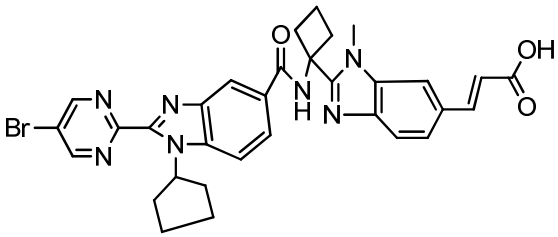
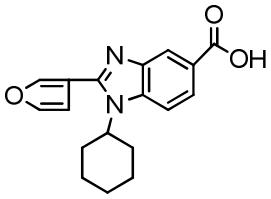
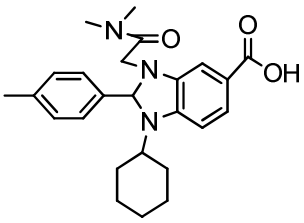
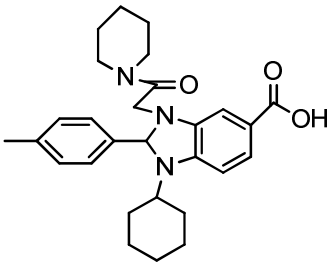
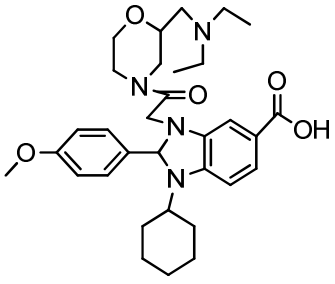
Molecular docking

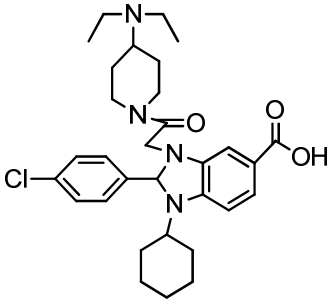
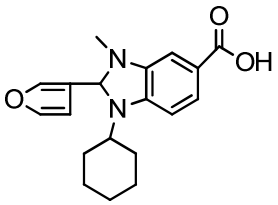
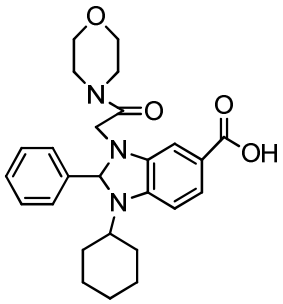
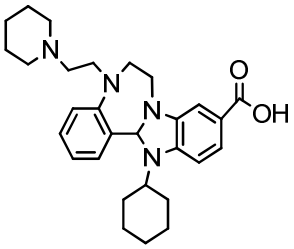
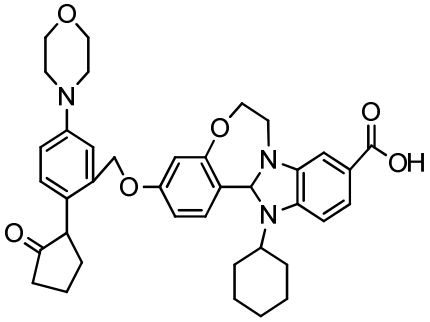
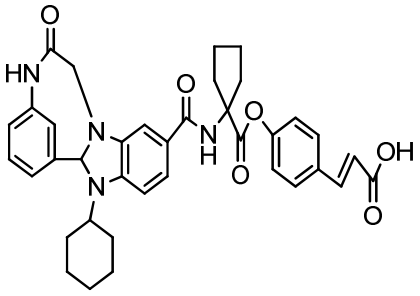
Molecular docking studies between the targets protein (1CSJ) and the imidazole derivatives (ligands) were carried out. The crystal structure of the target protein 1CSJ was retrieved from the protein data bank website (PDB). Initially, the optimized structure of imidazole derivatives was saved as SDF files which were letter converted to PDB files by Spartan 14 Version 1.1.4 software. Autodock Vina incorporated in Pyrx software were used for docking the prepared ligands with the prepared structure of 1CSJ receptor. The docked results were visualized and analyzed using Discovery Studio Visualizer.

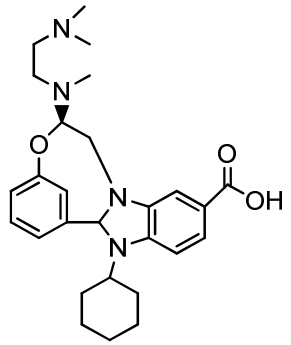
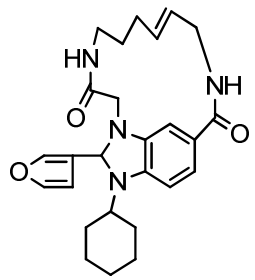
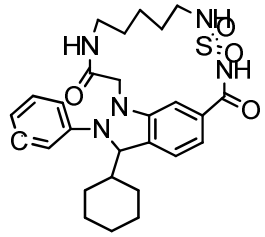
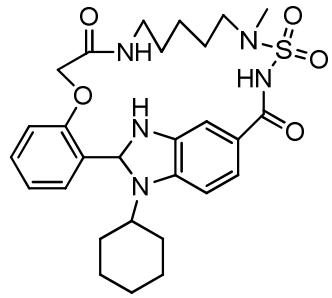
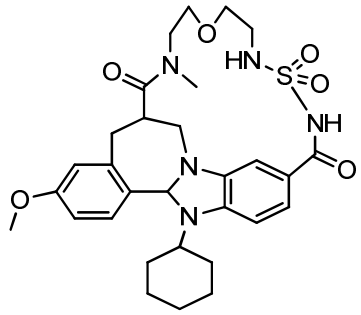
Table 1 Biological activities of training and test set derivatives.

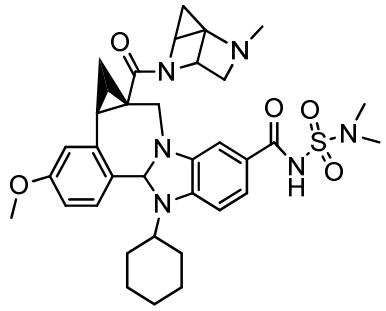
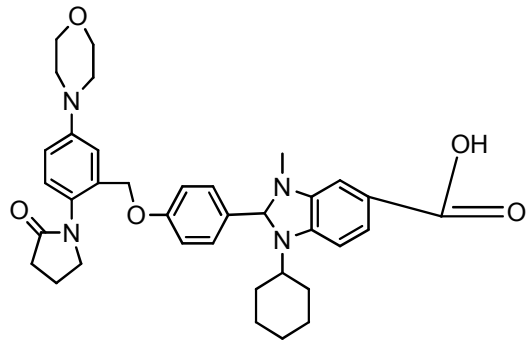
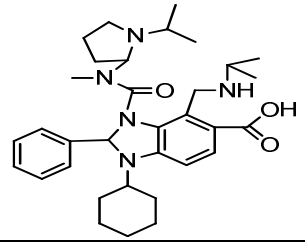
S/n	Structure	pIC_{50} (μM)	Pred. pIC_{50}
1Y		1.00	1.13045800
2Y		0.17	0.92096100
3Y		1.01	0.83937600
4X		0.52	-0.15048
5X		0.49	0.97861
6Y		1.37	0.93049200
7Y		1.07	1.02819700

8Y		1.14	1.23255300
9Y		0.20	0.97875700
10X		0.56	0.679371
11X		0.52	0.872689
12Y		1.30	0.97119100
13X		0.88	0.786908
14Y		1.39	1.23001200

15X		0.25	0.845448
16Y		1.22	0.87200400
17Y		1.36	1.56461200
18Y		1.79	0.98368900
19Y		1.76	1.85277500
20Y		1.95	1.38005900
21Y		2.22	2.14203100

22Y		2.39	2.07264100
23Y		1.35	0.93151900
24Y		0.09	0.99231900
25Y		2.04	1.85480200
26Y		2.14	2.23238400
27Y		1.69	1.18567800

28Y		2.22	2.20861500
29Y		0.41	0.31030200
30Y		0.12	0.46164800
31X		0.74	0.902028
32Y		1.11	1.14711500

33Y		2.52	2.84826100
34Y		2.00	0.47473500
35Y		1.95	2.19281400

Y= training set, X= test set.

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 R2 and Q2	≤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

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

S/N	Symbols	Name of descriptors	Class
1	Energy (aq)	Sum of the base energy.	2D
2	Min loclonpot	ATs autocorrelation descriptors weighted by scale atomic mass.	2D
3	ATSc4	ATS autocorrelation descriptors, weighed by charge	2D
4	Globaltopochargeindex	Global topological charge index	2D
5	WTPT-4	Sum path lengths starting from oxygen	2D

Table 4 Validation of the genetic function approximation from the material studio.

Validation parameters	EQ 1	EQ 2	EQ 3	EQ 4	EQ 5
Friedman LOF	0.90378100	0.94202700	0.94716900	0.94924400	0.95035800
R-squared	0.71144500	0.69923400	0.69759200	0.69693000	0.69657400
Adjusted R-squared	0.64586400	0.63087800	0.62886300	0.62805000	0.62761300
Cross validated R-squared	0.58109700	0.56223700	0.54578600	0.56179900	0.53685200
Significant Regression	Yes	Yes	Yes	Yes	Yes
Significance-of-regression F-value	10.84838200	10.22931100	10.14988400	10.11809000	10.10105900
Critical SOR F-value (95%)	2.68403600	2.68403600	2.68403600	2.68403600	2.68403600
Replicate points	0	0	0	0	0
Computed experimental error	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
Lack-of-fit points	22	22	22	22	22
Min expt. error for non-significant LOF (95%)	0.35948800	0.36701600	0.36801600	0.36841900	0.36863500

Table 5: Calculated descriptors for the test set and predicted Activity.

Energy (aq)	Min loclonPot	ATSc4	globalTopoChargeIndex	WTPT-4	Yp
-2137.3	38.71	-0.18462	1.759699	13.07891	-0.15048
-1873.18	33.83	-0.00177	1.740706	12.93923	0.97861
-1756.45	33.9	-0.00242	1.694046	12.87175	0.679371
-1571.5	33.72	0.021936	1.657272	8.046833	0.872689
-2455.01	34.13	-0.1449	1.771896	10.52827	0.786908
-1945	33.98	0.180821	1.729944	18.25107	0.845448
-2096.42	33.41	-0.04923	1.704878	8.138622	0.902028

Table 6 Pearson's correlation matrix for descriptors used in QSAR model for the activities of anti-hepatitis C molecules.

	Energy (aq)	Min loclonPot	ATSc4	globalTopoChargeIndex	WTPT-4
Energy (aq)	1				
Min loclonPot	-0.28228	1			
ATSc4	0.576632	-0.57935	1		
globalTopoChargeIndex	-0.84444	0.469726	-0.45897	1	
WTPT-4	-0.02002	0.202936	0.489569	0.392881	1

Table 7. Binding Affinity, Hydrogen bond interaction and hydrophobic interaction formed between ligands with best binding energy and the active site of the 1CSJ receptor.

Ligands	Binding energy	Residual interaction	Hydrogen bond	Hydrogen bond distance.
27	-10.1	TRP397, CYS14 ALA15, LYS141 VAL144, MSE139 ILE160	ASN142	2.27054
33	-9.8	PHE193, LEU384 CYS366, CYS366 PHE193	ASN411, ASP318 GLY410	3.63321, 3.64693 3.37238
32	-10.2	PHE193, PHE193	ASN316, TYR415 ASN411, TYR195 GLY449	2.11111, 3.53307 3.41174, 3.60541 2.82589
31	-10.7	GLU398, VAL144 TRP397, TRP397 ALA39, PRO404 ARG394, B:TRP397	GLN446, GLU143	2.03685, 3.6066

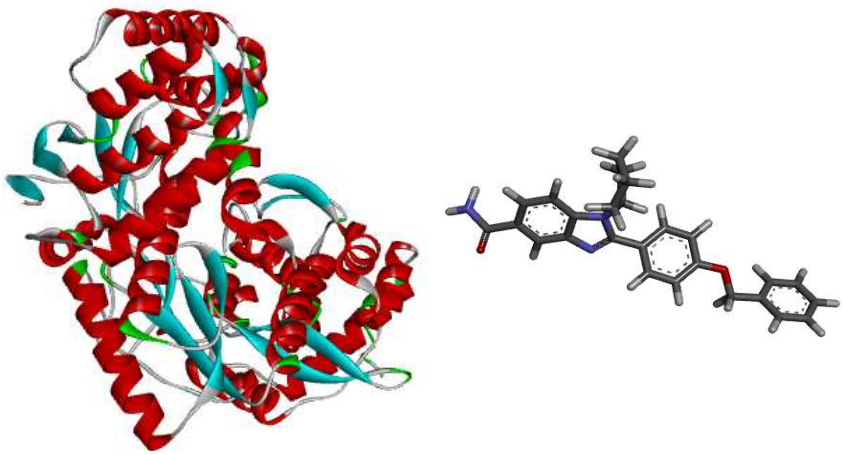


Figure 1: (A) Prepared structure of 1CSJ receptor (B) 3D structure of the prepared ligand.

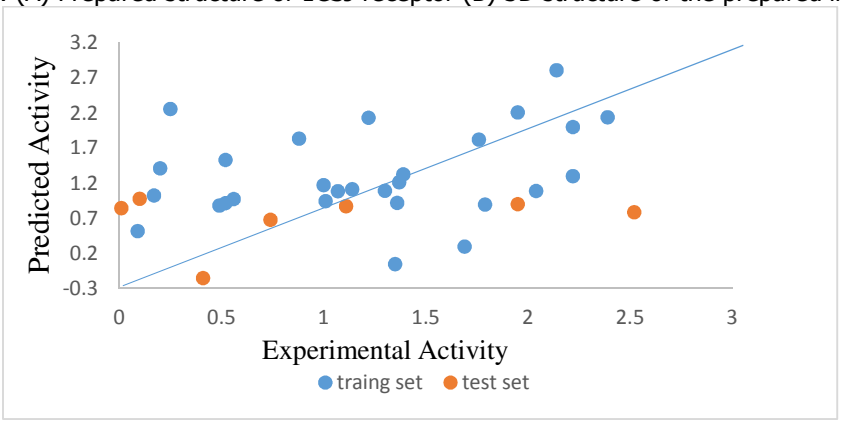
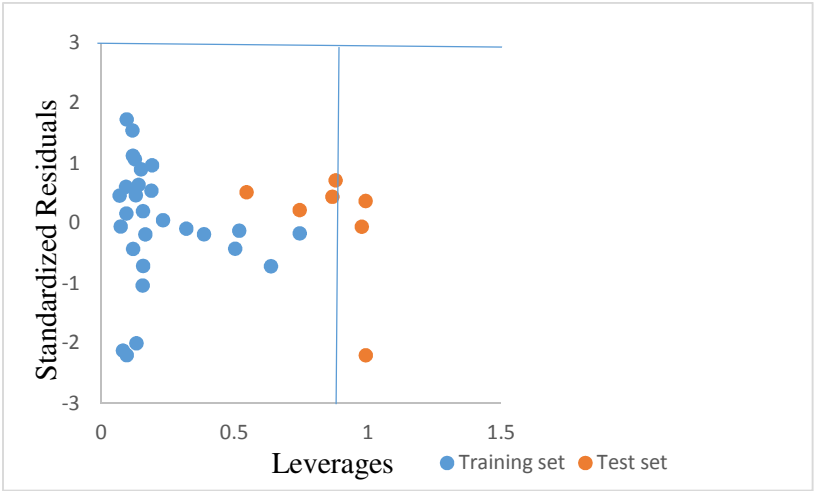
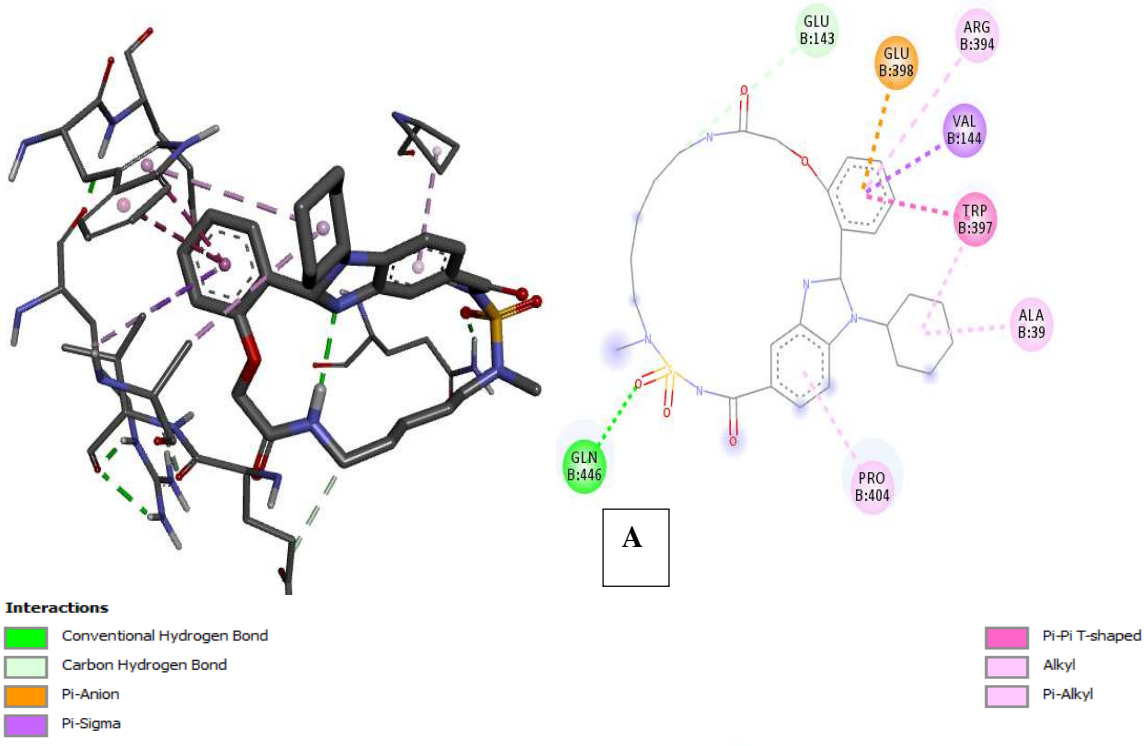


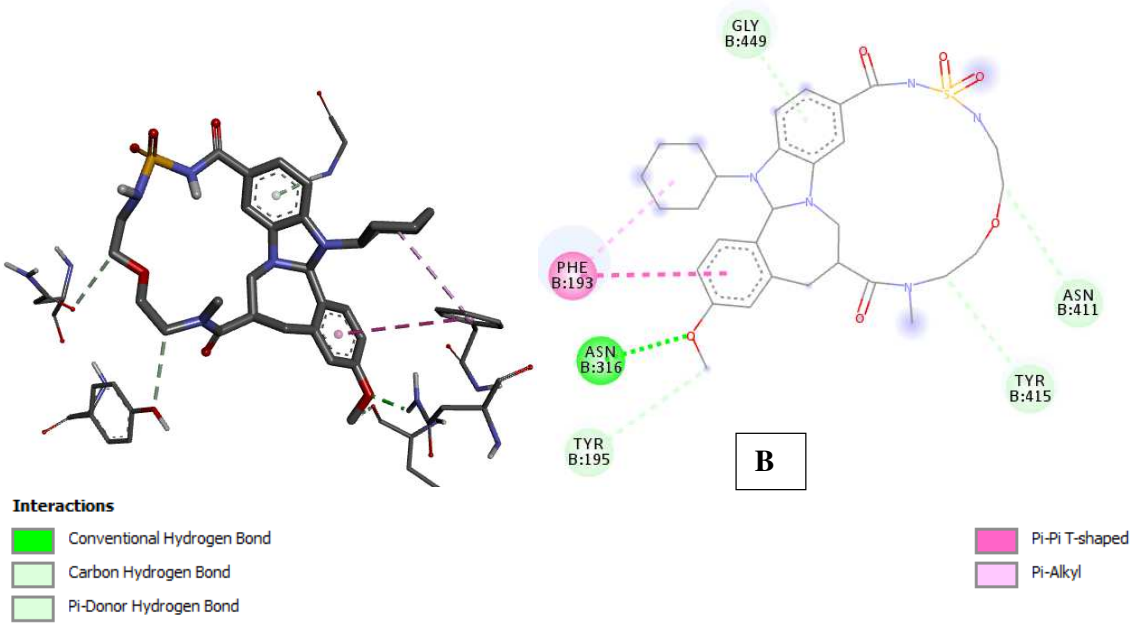
Fig 2. The plot of the Experimental and predictive activity of both training and test set of the best model (1).



h* = 0.7

Figure. 3 Williams plot, the plot of the standardized residuals versus the leverage value of both the training set and test set of model 1.


GLU B:143
GLU B:398
ARG B:394
VAL B:144
TRP B:397
ALA B:39
GLN B:446
PRO B:404
A
Interactions
Conventional Hydrogen Bond
Carbon Hydrogen Bond
Pi-Anion
Pi-Sigma
Pi-Pi T-shaped
Alkyl
Pi-Alkyl


GLY B:449
PHE B:193
ASN B:316
TYR B:195
ASN B:411
TYR B:415
B
Interactions
Conventional Hydrogen Bond
Carbon Hydrogen Bond
Pi-Donor Hydrogen Bond
Pi-Pi T-shaped
Pi-Alkyl

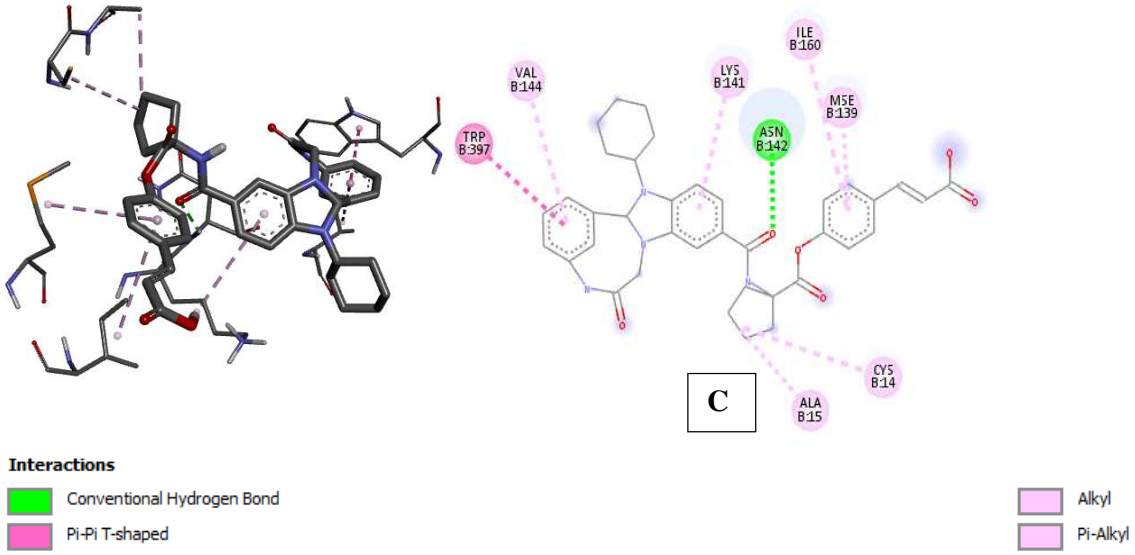


Figure. 4 3D and 2D structure of the docked - Ligands Complex. (A) Interactions between 1CSJ and Ligand 31x. (B) Interactions between 1CSJ and Ligand 32y. (C) Interactions between 1CSJ and Ligand 27y.

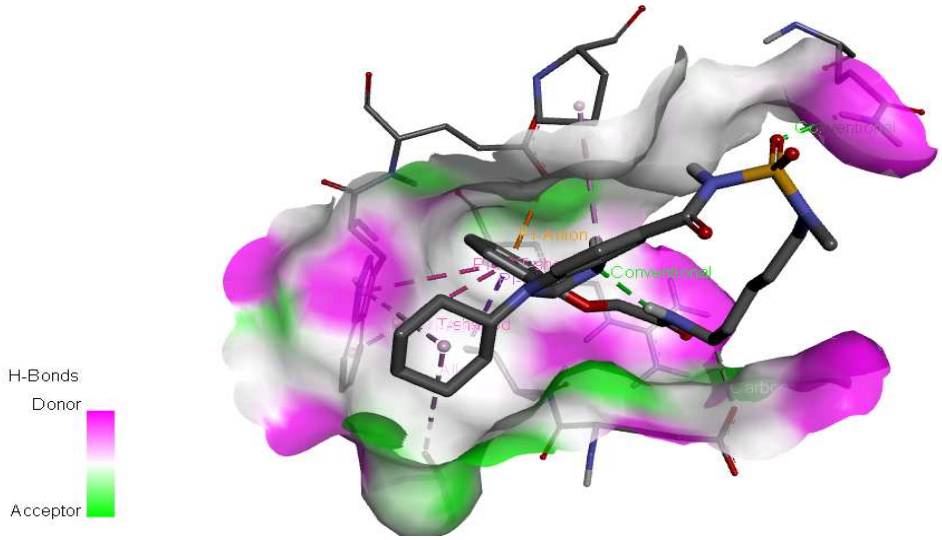


Figure 5: H-bond interaction between the ligand 31 and 1CSJ receptor.

RESULTS AND DISCUSSION

Five QSAR models were developed out of which the best model (model 1) was identified and reported due to the statistical significance and prominent validation parameters. The data set of 35 compounds were divided into a training set of 28 compounds which were used to build the model and a test set of 7 compounds which were used to validate the built model base on Kennard-Stone algorithm technique. The name and symbols of the descriptors used in the QSAR model are shown in Table 3. Table 4gives the result of the validation parameter using Genetic Function Algorithm (GFA) that confirm the stability and robustness of the model which were all in agreement with the minimum recommended value of validation Parameters for a generally acceptable QSAR model presented in Table 2. Based on the statistics generated, Model 1 was selected as the best QSAR model and reported as: pIC50 = 0.000443840 * Energy (aq) – 0.164409745 * Min loclonpot + 1.860290563 * ATSc4 + 7.343032148 * globaltopochargeindex – 0.063495786 * WTPT-4 – 4.58519, N = 35, $R^2_{ext.}$ = 0.572967, R^2 = 0.7114450, Q^2_{cv} = 0.58109700,

The high calculated value of R^2 (0.711) for pIC_{50} indicates a good internal validation of the model. The external validation of the model was also carried out R^2 pred. (0.572), the test set containing 20% of the data set were used for external validation to validate the model which is higher than minimum recommended value for the evaluation of the quantitative QSAR model. From figure 2, the developed model is stable and the residuals on both sides of zero are randomly propagated. The R^2 value of 0.7114 for the training set and R^2_{ext} value of 0.5629 for test set reported in this study was in agreement with Genetic Function Approbation (GFA) derived R^2 value reported in Table 2. This confirms the robustness and reliability of the model.

The Williams' plot shows the leverage value for the entire compounds in the dataset, which were plotted against its standardized residual value resulting in the discovery of influentials compound in the models. The results show that all the compounds were within the square area of $\pm$ 3 of standardized cross-validated residual produced by the model which shows no outlier compound. From figure 3, three compounds of test set were found to be influential since their leverage value are greater than the warning leverage (h* = 0.70). This was attributed to the difference in its molecular structure compared to other compounds in the dataset.

Interpretation of descriptors in model 1

From the model build, it can be concluded that the increase in **Min loclonpot** and **ATSc4** and a decrease in **Energy (aq), globaltopochargeindex** and **WTPT-4** will increase the anti-hepatitis C NS5B activity (pIC_{50}) of these imidazole derivatives.

The Result of molecular docking studies of imidazole derivatives

Molecular docking studies were carried out between the ligands (imidazole derivatives) and its targets (1CSJ). Table 7 shows the docking result of the four ligands with the best binding affinities which indicates that it has correlated with their pIC_{50}. The binding energy values of the target protein ranges from -7.2 to -10.7 kcal/mol. From the docking study, it is observed that ligand number 31x with the highest binding energy of -10.7 kcal/mole (fig 4A) was surrounded by two hydrogen bonding of GLN446 (2.03685 A°) and GLU143 (3.6066 A°) of the target. Hydrophobic interaction is a bond formed between the ligand and the binding pocket of the target site (receptor). Ligand 31x formed a hydrophobic bond of GLU398, VAL144, TRP397, ALA39, PRO404, ARG394, and TRP397 of the target site. Ligand 32y also bounded by two hydrophobic interaction of PHE193 and PHE193 of the target. While hydrogen bonding of ASN316, TYR415, ASN411, TYR195, and GLY449 (2.11111, 3.53307, 3.41174, 3.60541, 2.82589 A°) of the target. Moreover, we realized that the binding scores generated were found to be better than one proposed by Balavignesh *et al.* (2013).

CONCLUSION

QSAR model was generated with a descriptor (Energy (aq), Min loclonPot, ATSc4, global Topo Charge Index, WTPT-4) which were highly correlated with biological activities of imidazole derivatives. These descriptors produced a robust model to predict the anti-hepatitis C activities of these compounds. The significant correlation coefficient of determination R^2value of 0.7114 and Q^2value of 0.5810 indicated a good predictive ability of this model. The external predictive power (R^2= 0.5728) was satisfied with the agreement of the recommended value of validation parameters for a generally acceptable QSAR model. The Molecular docking studies showed compound 31x with the best binding affinity of -10.7 kcal/mol correlate with it pIC_{50} activity which formed H-bond GLN446 (2.03685 A°) and GLU143 (3.6066 A°) and hydrophobic pocket of GLU398, VAL144, TRP397, ALA39, PRO404, ARG394, and TRP397 with amino acid of the target. The high binding affinity of -10.7 kcal/mole implies that imidazole derivatives (inhibitors) will bind tightly to the NS5 polymerase and inhibit the enzyme through amino acid residues. The result of QSAR model alongside with molecular docking study provides a good approach for Pharmaceutical and medicinal researchers to design new anti-hepatitis C agent against NS5B polymerase receptor.

Contributions by corresponding author and supervisory team

All authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, statistical parameters, writing, or revision of the manuscript. Furthermore, each author certifies that this manuscript has not been and will not be submitted to or published in any other publication.

Acknowledgments
We acknowledge the authors of various literature consulted in the course of this research who anonymously helped at improving the quality of the manuscript. We also greatly acknowledge the contribution of Yusuf Bello Jada, Ahmadu Bello and Rukayyatu Muhd Chubado. And lastly we also want to acknowledge Muhammad Tukur Ibrahim, Usman Abdulfatai and David Ebuka Arthur for their valuable comments and suggestions.

Conflict of Interest
No conflict of interest.

Funding
The authors received no direct funding for this research.

Compliance with Ethics Requirements
This article does not contain any study (ies) with human or animal.

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.

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).

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.

Hadziyannis, S. J., Sette, H., Morgan, T. R., Balan, V., Diago, M., Marcellin, P., Ramadori, G., Bodenheimer, H., Bernstein, D., and Rizzetto, M. (2004). Peginterferon-α2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. *Annals of internal medicine, 140*(5), 346-355.

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.

Lavanchy, D. (2009). The global burden of hepatitis C. *Liver International, 29*(s1), 74-81.

Li, Z., Wan, H., Shi, Y., and Ouyang, P. (2004). Personal experience with four kinds of chemical structure drawing software: review on ChemDraw, ChemWindow, ISIS/Draw, and ChemSketch. *Journal of Chemical Information and Computer Sciences, 44*(5), 1886-1890.

Marcellin, P., Asselah, T., and Boyer, N. (2002). Fibrosis and disease progression in hepatitis C. *Hepatology, 36*(S1).

Mas, V. R., Maluf, D. G., Archer, K. J., Yanek, K., Bornstein, K., and Fisher, R. A. (2009). Proteomic analysis of HCV cirrhosis and HCV-induced HCC: identifying biomarkers for monitoring HCV-cirrhotic patients awaiting liver transplantation. *Transplantation, 87*(1), 143.

Meng, X.-Y., Zhang, H.-X., Mezei, M., and Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design, 7*(2), 146-157.

Mohammad, H., and Zohreh, A. (2013). In-silico prediction of rgs4 inhibitory activity of sometiadiazolidinone. *Int J Med Pharm, 1*.

Roy, K., Kar, S., and Das, R. N. (2015). *Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment*: Academic press.

Shepard, C. W., Finelli, L., and Alter, M. J. (2005). Global epidemiology of hepatitis C virus infection. *The Lancet infectious diseases, 5*(9), 558-567.

Tropsha, A., Gramatica, P., and Gombar, V. K. (2003). The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *Molecular Informatics, 22*(1), 69-77.

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.

Yap, C. W. (2011). PaDEL- descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of computational chemistry, 32*(7), 1466-1474.

Zhao, F., Liu, N., Zhan, P., Jiang, X., and Liu, X. (2015). Discovery of HCV NS5B thumb site I inhibitors: core-refining from benzimidazole to indole scaffold. *European journal of medicinal chemistry, 94*, 218-228.