jcec

The Journal of Engineering and Exact Sciences – JCEC, Vol. 04 N. 02 (2018)
journal homepage: https://jcec.ufv.br
doi: https://doi.org/10.18540/jcecvl4iss2pp0265-0275
OPEN ACCESS – ISSN: 2527-1075

IN-SILICO STUDIES OF SOME INDOLE DERIVATIVES AS AN ANTI-HEPATITIS C DRUG

A.S. BELLO[1*], A. UZAIRU[1], G.A. SHALLANGWA[1] AND A. IBRAHIM[1]

[1] Department of chemistry Ahmdu Bello University, P.M.B. 1044, Zaria, Kaduna state Nigeria.

*Corresponding author. Ahmdu Bello University, Department of chemistry, Zaria, Kaduna, Nigeria, Phone: +234 8039194040.
e-mail addressl: sadiqdanburam@gmail.com.

ARTICLE INFO

Article history:
Received 2018-04-12
Accepted 2018-05-27
Available online 2018-06-30

keywords

QSAR
Molecular Docking
Indole
NS5B polymerase
HCV
Binding Energy

ABSTRACT

A combined three-dimensional quantitative structure-activity relationship (QSAR) modeling and molecular docking studies were carried out on the 64 indole derivatives and was accomplished to profoundly understand the structure-activity correlation of indole-based inhibitors of the HCV NS5B polymerase against HCV. Genetic function approximation (GFA) of Material studio software version 8 was used to perform the QSAR study while Autodock vina version 4.0 of Pyrx software was used for molecular docking studies of the selected indole derivatives. The optimum model builds exhibited statistically significant results: squared correlation coefficient (R^2) of 0.760, adjusted squared correlation coefficient (R^2 adj) value of 0.708, Leave one out (LOO) cross-validation coefficient value of 0.634 and the external validation (R^2 pred) of 0.621. Molecular docking study of the indole derivative with 1G8Q as the protein target revealed that the best binding affinity with the docking scores of -9.4 kcal/mol formed hydrophobic interaction and H-bonding with amino acid residues of HCV NS5B polymerase. The QSAR model generated and molecular docking results proposed that the model had a good level of stability, strength, and predictability at internal and external validation, and the physicochemical parameters are to be analyzed when designing new indole derivatives agent with better activity against the 1G8Q target site.

1. INTRODUCTION

Hepatitis C virus (HCV) was identified in 1989 by Michael Houghton and his colleagues(Choo et al., 1989). HCV is a member of Flaviviridae family and a positive-sense single-stranded RNA virus with a single open frame of ~9600 nucleosides. The viral genome encodes a polyprotein containing more than 3000 amino acids, and the polyprotein is classified into two categories: (1) structural proteins the nucleocapsid core protein (C) and two glycoproteins E1 and E2; (2) non-structural proteins (NS) NS2, NS3, NS4A, NS4B, NS5A, and NS5B, because of their primary role in the replication of HCV virus. The HCV NS5B polymerase is an RNA dependent RNA polymerase that is necessary for the replicating viral RNA of HCV (Sofia, Chang, Furman, Mosley, & Ross, 2012) (Moradpour, Penin, & Rice, 2007) and (Vrontaki, Melagraki, Mavromoustakos, & Afantitis, 2015). Hepatitis C Virus (HCV) a significant human pathogen of global public health, important as one of the major pathogens that cause chronic hepatitis, cirrhosis and hepatocellular carcinoma (HCC) (Shepard, Finelli, & Alter, 2005) and (Alter, 2007). 2.8% of the world population (about 180 million individuals according to the database of World Health Organization) has infected with HCV and 3-4 million new infections each year. (Mohd Hanafiah, Groeger, Flaxman, & Wiersma, 2013) (Haudecoeur, Peuchmaur, Ahmed-Belkacem, Pawlotsky, & Boumendjel, 2013) and (Lavanchy, 2009).

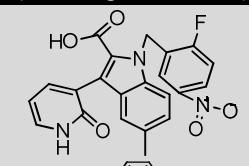
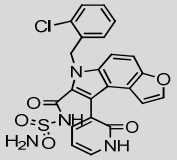
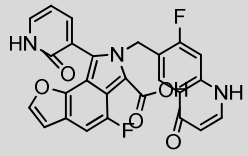
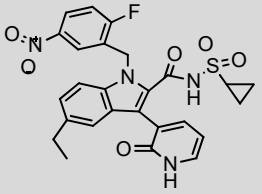
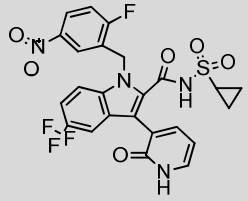
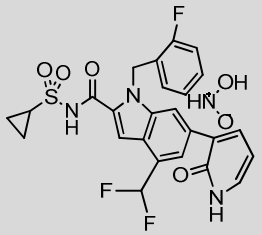
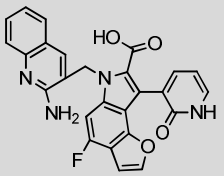
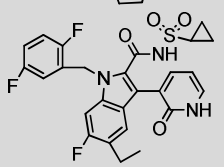
Slow progress and mild symptoms, these features make it a hidden epidemic and most infections progress a chronic state that lasts for decades (Shepard et al., 2005). HCV symptoms include muscle aches, tenderness in the upper abdomen, yellow tinge to the skin and eyes, dark urine (jaundice), and light-colored bowel movements. At present, the anti-HCV vaccine is unavailable (Fauvelle et al., 2013) and (Law, Landi, Magee, Tyrrell, & Houghton, 2013) and the standard of care (SOC) includes a combination of a protease inhibitor with pegylated α-interferon (PEG-IFN-α) and the oral nucleoside antiviral agent ribavirin (RVB) (Lü & XUE, 2011). Therefore, it is very important to produce new anti-HCV drugs with encouraging activity and less toxicity. The drug design has widely been used in the discovery and development of drugs due to its slow and time-consuming advantages, cost reduction, high efficiency in silico screening and prediction of competitor drugs with improvements in computer technologies and simulation programs (Mohammad & Zohreh, 2013). The quantitative relationship between activity and activity QSAR in the simplest terms is a way of constructing mathematical models trying to establish a statistically the moral relationship between structure and function using the chemical technique. The QSAR method is capable of estimating the properties of new chemical compounds without having to synthesize and test (Barril & Morley, 2005). Molecular docking is one of the most widely used techniques in structure-based drug design SBDD due to its ability to predict with a substantial degree of accuracy, the confirmation of small-molecule ligands within the appropriate target binding site (Meng, Zhang, Mezei, & Cui, 2011). The aim of this research was to develop various QSAR models using Genetic Function Algorithm (GFA) method for predicting the activities of some selected indole derivatives and to predict the strength of interactions between indole derivatives (inhibitors) and NS5B polymerase protein (PDB code 1G8Q), an enzyme that is responsible for Hepatitis C.

2. MATERIAL AND METHOD

2.1 Datasets used

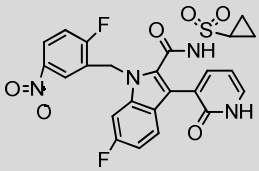
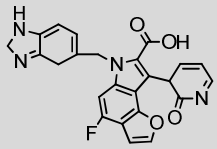
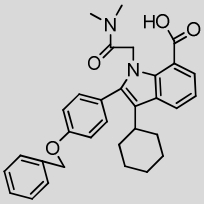
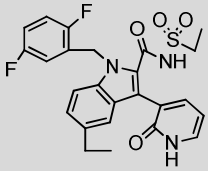
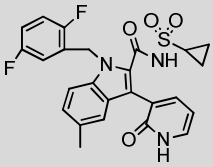
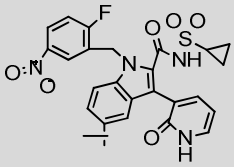
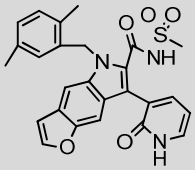
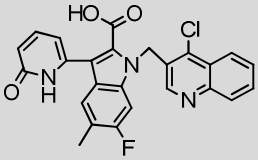
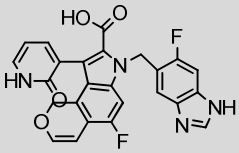
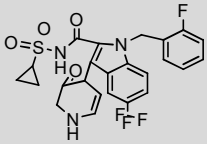
Sixty-four (64) Molecules of indole derivatives were selected from the literature and used for the present study (Wei et al., 2016). The activities of the indole molecules measured as IC50 (nM) were expressed as the logarithmic scale. The pIC50 (pIC50 =log1/IC50) was used as dependent variable thus linearly linking the data with the independent variable/ descriptors. Table 1 shows the observed structures and the biological activities of indole compounds

Table 1 Structures and activities of indole-based inhibitors of the HCV NS5B polymerase.

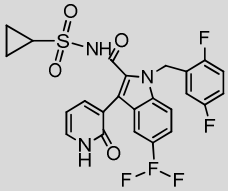
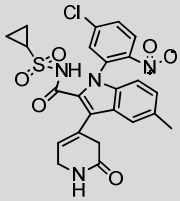
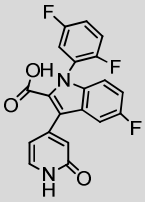
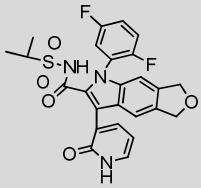
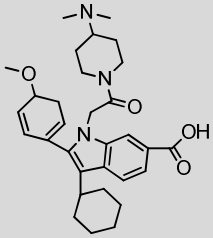
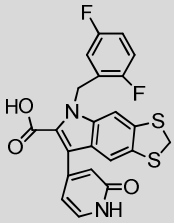
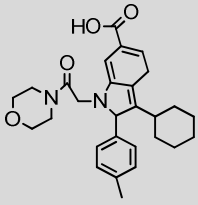
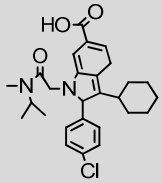
S/N	Structure (a= training set, b= test set)	pIC_{50} (nM)	Pred. pIC_{50}
1b		0.47	0.894319
2b		0.47	0.89955
3b		0.47	0.689778
4b		0.60	0.625196
5b		0.60	0.623552
6a		0.60	0.61163800
7a		0.60	0.84144900
8a		0.69	0.62180500

JCEC - ISSN 2527-1075.

Continued Table 1

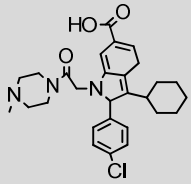
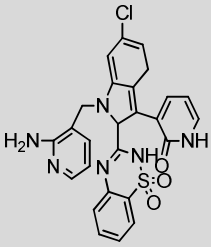
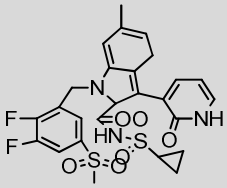
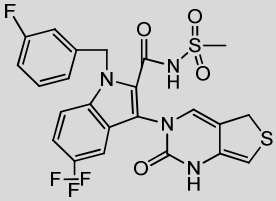
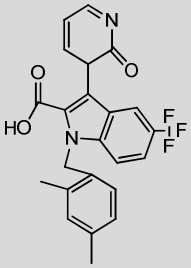
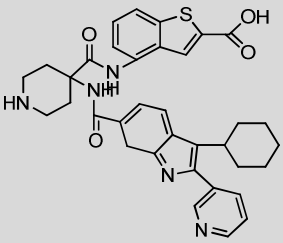
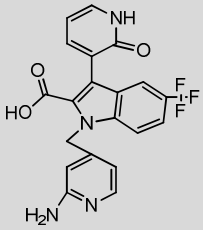
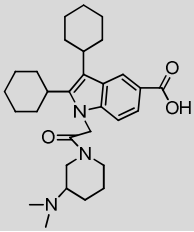
S/N	Structure (a= training set, b= test set)	pIC_{50} (nM)	Pred. pIC_{50}
9a		0.69	0.61163800
10a		0.69	0.92878400
11a		0.77	1.08877500
12a		0.77	0.79022500
13a		0.77	0.75628500
14a		0.77	0.67161300
15a		0.77	1.04552500
16a		0.77	0.91831900
17a		0.77	0.91831900
18a		0.84	1.25556900
19a		0.84	0.85970000

Continued Table 1

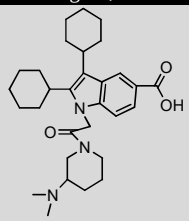
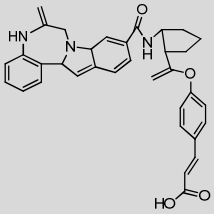
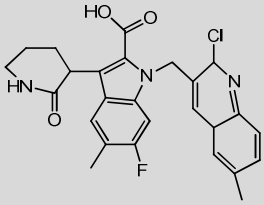
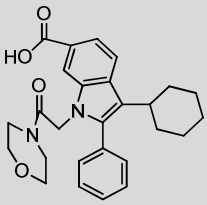
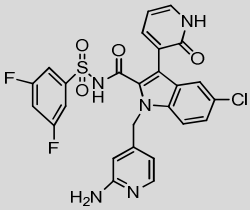
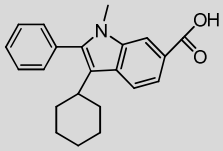
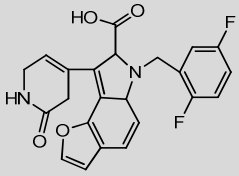
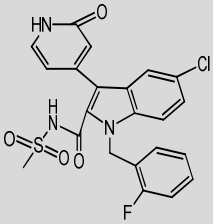
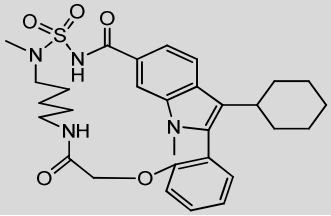
S/N	Structure (a= training set, b= test set)	pIC_{50} (nM)	Pred. pIC_{50}
20a		0.84	0.64399000
21a		0.84	0.95642100
22a		0.90	0.98399300
23a		0.90	0.73702600
24a		0.95	1.15918900
25a		0.95	1.20297600
26a		1.00	0.89981500
27a		1.04	1.34769200
28a		1.04	1.30875700

JCEC - ISSN 2527-1075.

Continued Table 1

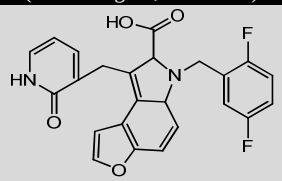
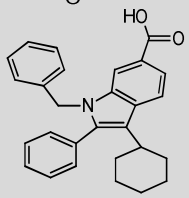
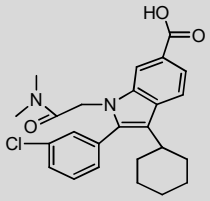
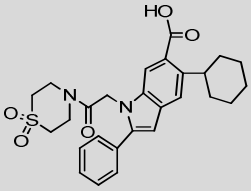
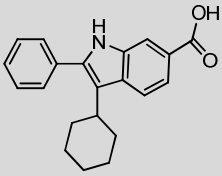
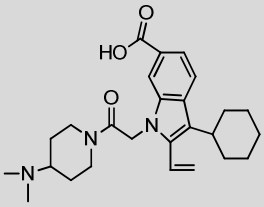
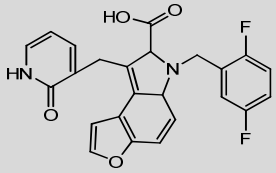
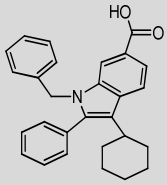
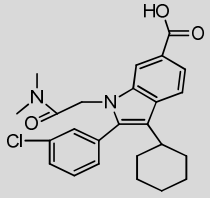
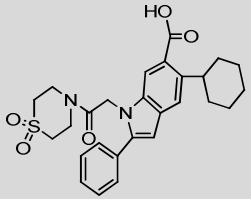
S/N	Structure (a= training set, b= test set)	pIC_{50} (nM)	Pred. pIC_{50}
29b		1.07	1.450111
30b		1.07	1.033415
31b		1.14	0.741121
32b		1.17	1.037696
33b		1.20	0.983785
34a		1.23	1.13189800
35a		1.23	1.01762500
36a		1.25	1.30983600

Continued Table 1

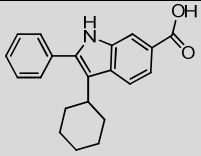
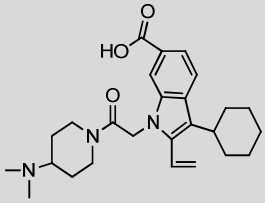
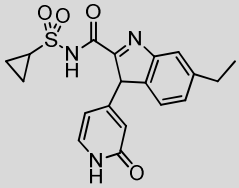
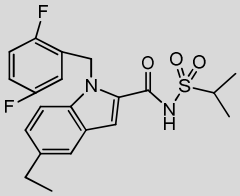
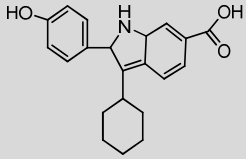
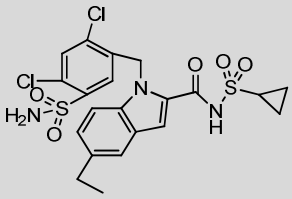
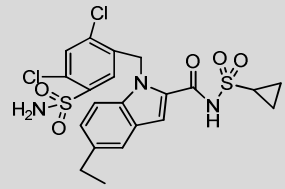
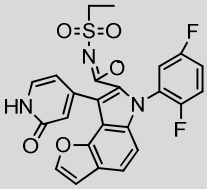
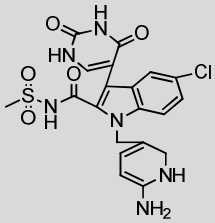
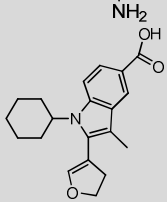
S/N	Structure (a= training set, b= test set)	pIC_{50} (nM)	Pred. pIC_{50}
37a		1.27	1.11075000
38b		1.27	1.439499
39a		1.31	1.22951600
40a		1.41	1.19276700
42a		1.41	1.09410700
43a		1.44	0.87885000
44a		1.49	1.26526300
45a		1.51	1.15058900
46a		1.51	1.11961100

JCEC - ISSN 2527-1075.

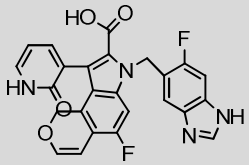
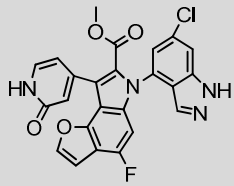
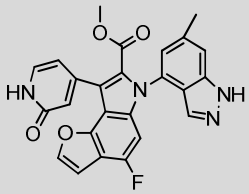
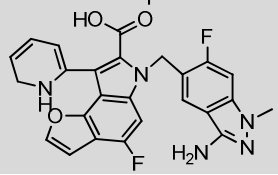
Continued Table 1

S/N	Structure (a= training set, b= test set)	pIC_{50} (nM)	Pred. pIC_{50}
47a		1.53	1.63567600
48a		1.56	1.14033900
49a		1.60	1.26422400
50a		1.63	1.20623300
51a		1.68	1.37459200
52a		1.69	1.44762900
47a		1.53	1.63567600
48a		1.56	1.14033900
49a		1.60	1.26422400
50a		1.63	1.20623300

Continued Table 1

S/N	Structure (a= training set, b= test set)	pIC_{50} (nM)	Pred. pIC_{50}
51a		1.68	1.37459200
52a		1.69	1.44762900
53a		0.30	1.12874300
54a		0.60	0.90982800
55a		1.68	0.72235700
56a		0.77	1.73616300
57a		0.77	0.79031400
58a		0.77	0.70726600
59a		1.43	0.83885300
60b		1.64	1.322014

Continued Table 1

S/N	Structure (a= training set, b= test set)	pIC_{50} (nM)	Pred. pIC_{50}
61b		0.77	1.300057
62b		0.47	0.663852
63b		0.47	0.604856
64b		0.95	0.669321

2.2 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, Uzairu, & Uba, 2017) was conveyed to PaDEL-Descriptor version 2.18 toolkits (Yap, 2011) where the calculation of 1D, 2D, and 3D descriptors took place.

2.3 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 advanced model. (Kennard & 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 that were 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, identifying genetically essential functions, 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 follows:

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

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.

2.4 Quality assurance of the model.

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

2.5 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.

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

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 (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. R^2 adj is defined as follows:

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

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 the following:

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

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 & Kyani, 2004).

2.6 Applicability domain.

Applicability Domain (AD) is the chemical descriptor space incorporated by a 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:

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

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

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

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. Fig. 3, shows that two of the training set and six of the test set fall inside the domain of the model (the warning leverage limit is 0.4), hence they are accepted as Y influential.

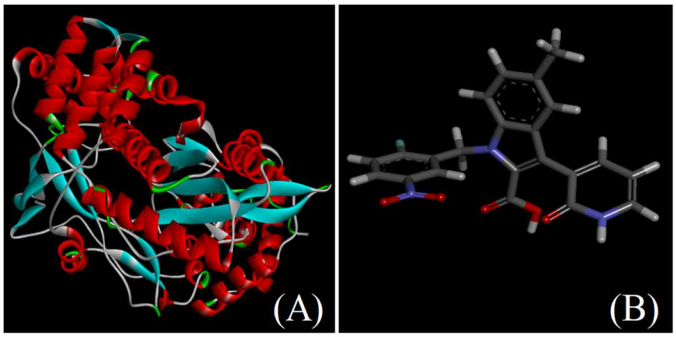


Figure 3 - (A) Prepared Structure of (1G8Q) protein; (B) Prepared structure of ligand (indole derivatives).

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

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

2.7 Molecular Docking studies.

Molecular docking is one of the most frequently used methods in drug design because of its ability to predict the conformation of small-molecule ligands within the appropriate target binding site. The molecular docking studies of active anti-hepatitis C compounds were performed by AutoDock Vina and PyRx virtual screening software using the reference of the template substrate. Running on HP core i3, Microsoft operation windows 10 professional version 2010 computer system, with Intel ® Core™ i3 Dual CPU 5157U @2.50 GHz 2.50GHz, 8GB of RAM. The score function, dock function (S, kcal/mol) developed by Autodock program was used for evaluation of the binding affinity of the indole derivatives (ligands) with the receptor (1G8Q).

2.8 Preparation of Ligands and Receptor for Docking.

The preparation of ligands are as follows; (i) conversions of 2D to 3D, (ii) correcting structures, (iii) validation and optimizing the structures. All these tasks were performed using Spartan'14 version 1.1.2. The crystal structure of NS5B polymerase (receptor) with the PDB code of (1G8Q) was download from Protein Databank website (PDB). The preparation of the crystal structure of the receptor was performed using Autodock version 4.2 software (Veerasamy et al., 2011).

2.9 Docking using Autodock version 4.0 of Pyrx software.

The molecular docking of ligands (indole derivatives) with the receptor (NS5B polymerase) was performed using Autodock version 4.0 of pyrx software (Trott & Olson, 2010). Docking is a virtual screening of a database of compounds and predicting the efficiently binding ligand(s) based on different scoring functions. The ligand library has been generated by collecting all the 64 indole derivatives in an Autodock version 4.0 (Autodock vina) folder of pyrx software (Trott & Olson, 2010). The library setup helps to make a simple comparison between ligands by performing simultaneous docking of multiple ligands against the receptor. The network batch docking was also performed. The result of each docked molecule shown in terms of the final minimum score (Dock score interaction/ docking energy of receptor-ligand).

3. RESULT AND DISCUSSION

All the five developed QSAR models were identified and 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. 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.

$$pIC_{50} = 0.031920049\,\mathbf{ASTm1} - 0.045332344\,\mathbf{ASTm4} + 0.355723777\,\mathbf{MLFER_A} - 3.535873143\,\mathbf{RotBFrac} + 0.004913636\,\mathbf{VABC} + 1.070655$$

N = 64, $R^2_{ext.}$ = 0.621098, R^2 = 0.76039600, R^2_{adj} = 0.70806200, Q^2_{cv} = 0.63417700, LOF = 0.26279000, Min expt. Error for non-significant LOF (95%) = 0.212326700.

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

S/N	Symbols	Name of descriptors	Class
1	ASTm1	ATs autocorrelation descriptors weighted by scale atomic mass.	2D
2	ASTm4	ATs autocorrelation descriptors weighted by scale atomic mass.	2D
3	MLFER_A	Overall or summation solute hydrogen bond acidity.	2D
4	RotBFrac	The fraction of rotatable bonds, excluding terminal bond.	2D
5	VABC	Van der Waals volume calculated	2D

The highly calculated Q^2 LOO value (0.760) for pIC50 indicates a good internal validation of the model. The external sample validation for R^2_{ext} (0.621) was also performed, and the test set containing 25% of the data set was used to validate the

JCEC - ISSN 2527-1075.

external form which is higher than the standard value of 0.5 for the model.

From figure 1, the developed model is stable and the residuals on both sides of zero are randomly propagated.

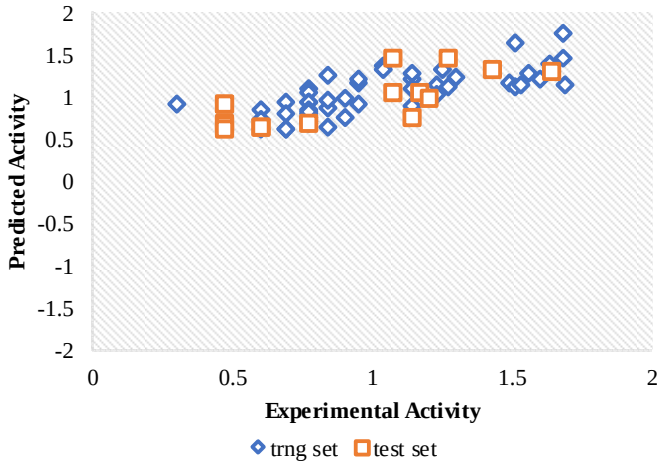


Figure 1 - The plot of the Experimental and predictive activity of both training and test set of the best model (1).

Figure 2 of Williams' plot shows the leverages for each compound in the dataset, which were drawn against its standard residual, resulting in the discovery of 8 influentials. Figure 2 also shows that two sets of training compounds with (pIC50 of 1.14 and 1.68) and six test set compounds (PIC50 of 0.69, 0.47, 0.47, 0.9, 1.14, and 0.77) were out of the applicability domain of the model. All of these compounds have their leverage values higher than the warning leverage value (h* = 0.4), and their high leverage value is responsible for influencing the performance of the model.

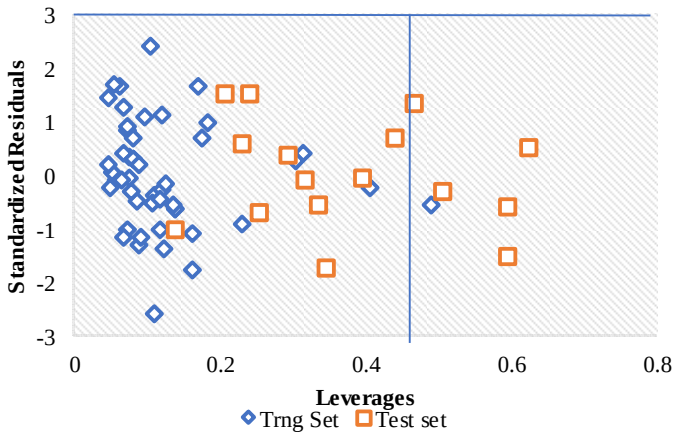


Figure 2 - The plot of the standardized residuals versus the leverage value of both the training set and test set of model 1 (h* = 0.4)

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 parameter for the models using genetic function approximation.

Validation parameters.	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
Friedman LOF	0.26279000	0.26732800	0.26825300	0.26855700	0.2691520
R-squared	0.76039600	0.75280300	0.75125700	0.75074900	0.7497530
Adjusted R-squared	0.70806200	0.69956500	0.69783500	0.69726600	0.6961520
Cross-validated R-squared (Q^2_{cv})	0.63417700	0.61480600	0.59284700	0.62490700	0.5970510
Significant Regression	Yes	Yes	Yes	Yes	Yes
Significance-of-regression F-value	10.70809800	10.38366800	10.31894100	10.29777200	10.2564170
Critical SOR F-value (95%)	2.46448800	2.46448800	2.46448800	2.46448800	2.4644880
Replicate points	1	1	1	1	1
Computed experimental error.	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
Lack-of-fit point.	41	41	41	41	41
Min expt. Error for non-significant LOF (95%)	0.21326700	0.21510100	0.21547200	0.21559400	0.2158330

Table 5 Pearson's correlation matrix for the selected

	ATSm1	ATSm4	MLFER_A	RotBFrac	VABC
ATSm1	1				
ATSm4	0.863254	1			
MLFER_A	-0.22554	-0.27328	1		
RotBFrac	0.402991	0.244536	0.07342	1	
VABC	0.617905	0.665186	0.268955	0.530202	1

3.1 Interpretation of descriptors in model 1.

ASTm1 and **ASTm4** have defined as 2D correlated descriptors ATs autocorrelation descriptor, weighted by scale atomic mass, **MLFER_A** is 2D MLFER Descriptors and is defined as overall or summation solute hydrogen bond acidity. **RotBFrac** is also 2D PaDEl rotatable bonds count Descriptors and is defined as the fraction of rotatable bonds, excluding terminal bonds. **VABC** is another 2D VABC Descriptors and is defined as van der Waals volume calculated. From the model, we can conclude that the increase in **ASTm1, MLFER_A, and VABC** and decrease in **ASTm4** and **RotBFrac** will increase the anti-hepatitis C NS5B activity (pIC_{50}) of these indole derivatives.

3.2 Molecular docking studies.

Molecular docking studies between the target protein (1G8Q) and the indole derivatives (ligands) were performed. All the compounds were found to strongly inhibit by completely occupying the active sites in the target protein (1G8Q). All inhibitors showed low energy values (high docking scores) than the bond energies. For target protein, binding energy values range from -6.3 to -9.4 kcal/mol. The number 58a compound

with the binding energies of -9.4 kcal/mol showed the best binding energies than other interconnections of the co-ligands.

3.3 Binding mode of inhibitors.

Table 6 and 7 showed least and the best docking scores, hydrogen bond length (in angstrom) and the reactive residues involved in the laying of docking inhibitors (ligands) at the active side of 1G8Q. Fig. 4 gives the best three result of docking studies. The Ligand number 58a (a compound with the best binding score of -9.4 kcal/mol) shows that LYS216, PHE213 VAL146, PHE126 residues target are involved in Electrostatic and hydrophobic interactions; they also form hydrogen bonds with GLN129 (2.23 A˚). Compound 38b made two hydrogen bond interactions of GLN125 (2.317 A˚) and GLN132 (2.32 A˚) with four residues ALA130, VAL146, LYS216, and PHE213. Compound 46a also creates a hydrogen bond with LYS116 (3.22 A˚) and hydrophobic interactions with ALA230, ALA243, VAL246, LYS116, LYS116, and PHE226.

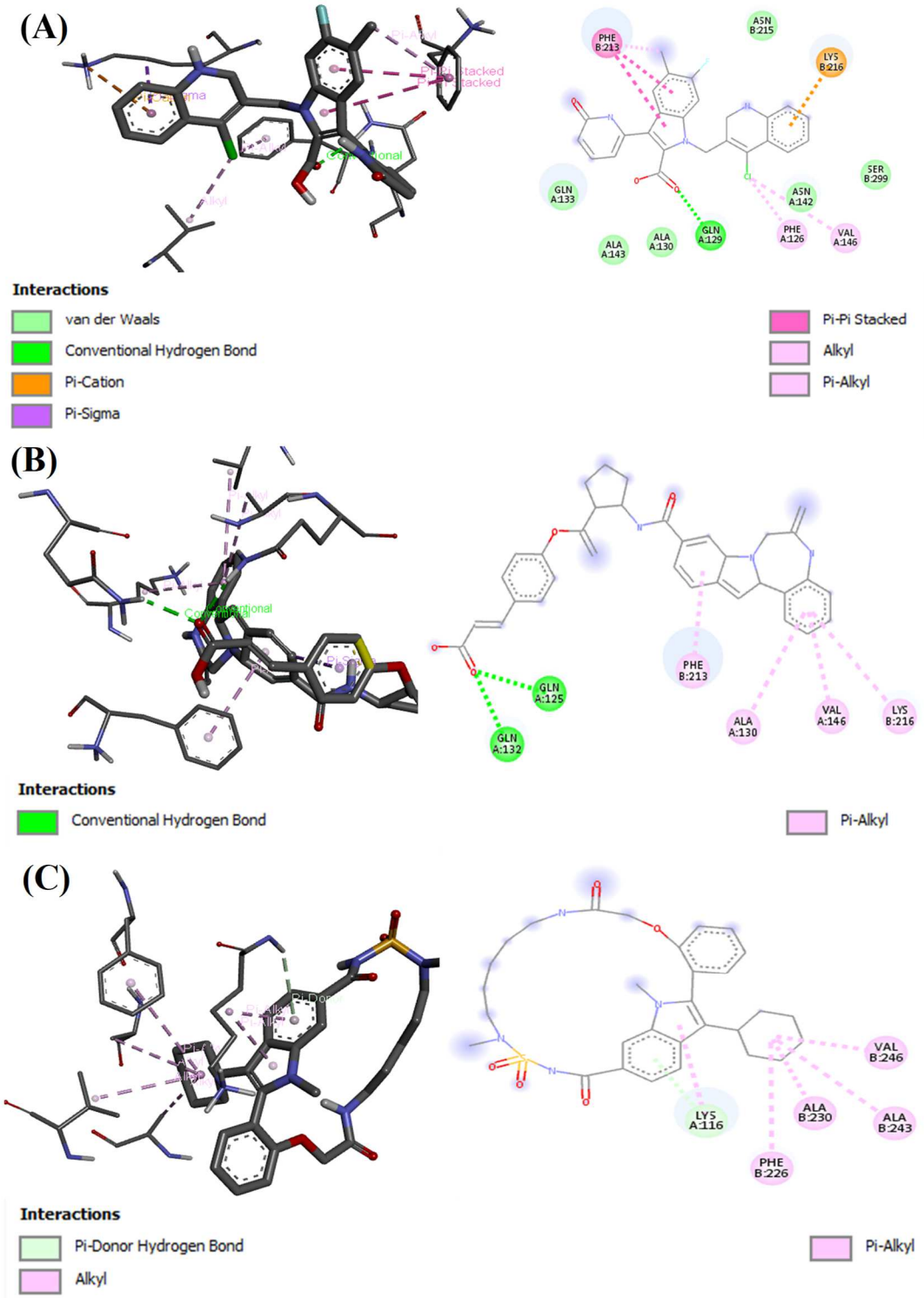


Figure 4 - 3D and 2D structure of the docked - Ligands Complex. (A) Interactions between 1G8Q and Ligand 58a. (B) Interactions between 1G8Q and Ligand 38b. (C) Interactions between 1G8Q and Ligand 46a.

Table 6 Binding Affinity, Hydrogen bond interaction and hydrophobic interaction formed between ligands with least binding energy and the active site of the 1G8Q receptor.

Ligands	Binding energy (Kcal/mol)	Residual interaction	Hydrogen bond	Hydrogen bond distance.
24	-6.3	ASP228, ASP228, LYS221, LYS221 LYS224, LEU262	LYS221	3.33198
14	-6.4	ALA140	ALA143	2.58093
52	-6.4	LEU185, ILE181 ILE181, LEU185	ASN184, ASN184	2.55278, 3.35309
6	-6.6	GLN225, GLN232, ASP228, GLN225, and LEU262	ASP228, THR261, LYS221 and GLN229	2.06288, 2.99654 2.81134 and 2.79842

Table 7 Binding Affinity, Hydrogen bond interaction and hydrophobic interaction formed between ligands with best binding energy and the active site of the 1G8Q receptor.

Ligands	Binding energy (Kcal/mol)	Residual interaction	Hydrogen bond	Hydrogen bond distance.
17	-8.5	PHE126, PHE213 VAL146, LYS216 LYS216, LYS216 LYS216	GLN129, ASP217 LYS216, ASP217 and PHE298	2.17364, 2.15338 2.53616, 2.75181 3.59918
46	-8.8	ALA230,ALA243 VAL246, LYS116 LYS116, PHE226	LYS116	3.22038
38	-8.9	ALA130, VAL146 LYS216, PHE213	GLN125 GLN132	2.31716 2.32755
58	-9.4	LYS216, LYS216 PHE213, PHE213 VAL146, PHE126 PHE213	GLN129	2.23902

4. CONCLUSION

In this research QSAR model was generated with descriptors (ASTm1, RotBFrac, ASTm4 MLFER_A and VABC) which were correlated with biological activities of indole derivatives and have the R^2 of 0.760 and Q^2 of 0.634, the validation parameters showed a good predictive ability of the model. The external predictive power (R^2 = 0.621) was satisfactory. Molecular docking analysis revealed that Compound (58a) with the best binding affinity and docking score of -9.4 kcal/mol against the protein (1G8Q) which have H-bond formed at GLN129 (2.23 A˚) and hydrophobic/residual interaction of LYS216, LYS216, PHE213, PHE213, VAL146, PHE126 and PHE213 with the protein of the target. From the result obtain from this research work we found out that the docking analysis and the binding scores generated were found to be better than the one reported by (Balavignesh, Srinivasan, Ramesh Babu, & Saravanan, 2013). Thus, these findings provide useful guidance and support for the development of the indole-based inhibitors acting as potential inhibitors for hepatitis C virus NS5B polymerase.

A C K N O W L E D G M E N T S

We acknowledge the authors of many of the literature consulted in the course of this research, which Anonymous helped to improve the quality of the manuscript. We also acknowledge the contribution of Yusuf Bello Jada, Ahmadu Bello. Finally, we also want to acknowledge Usman Abdulfatah's valuable comments and proposals.

R E F E R E N C E S

ABDULFATAI, U.; UZAIRU, A.; & UBA, S. 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, 2017.

ALTER, M. J. Epidemiology of hepatitis c virus infection. *World journal of gastroenterology: wjg*, *13*(17), 2436, 2007.

BALAVIGNESH, V.; SRINIVASAN, E.; RAMESH BABU, N.; & SARAVANAN, N.. 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), 2013.

BARRIL, X.; & MORLEY, S. D. Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. *Journal of medicinal chemistry*, *48*(13), 4432-4443, 2005.

CHOO, Q.-L.; KUO, G.; WEINER, A. J.; OVERBY, L. R.; BRADLEY, D. W.; & HOUGHTON, M.

Isolation of a cdna clone derived from a blood-borne non-a, non-b viral hepatitis genome. *Science, 244*(4902), 359-362, 1989.

FAUVELLE, C.; LEPILLER, Q.; FELMLEE, D. J.; FOFANA, I.; HABERSETZER, F.; STOLL-KELLER, F.; FAFI-KREMER, S. Hepatitis c virus vaccines–progress and perspectives. *Microbial pathogenesis, 58*, 66-72, 2013.

HAUDECOEUR, R.; PEUCHMAUR, M.; AHMED-BELKACEM, A.; PAWLOTSKY, J. M.; & BOUMENDJEL, A. 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, 2013.

JALALI-HERAVI, M.; & KYANI, A. 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, 2004.

KENNARD, R. W.; & STONE, L. A. Computer aided design of experiments. *Technometrics, 11*(1), 137-148, 1969.

KHALED, K. Modeling corrosion inhibition of iron in acid medium by genetic function approximation method: a qsar model. *Corrosion science, 53*(11), 3457-3465, 2011.

LAVANCHY, D. The global burden of hepatitis c. *Liver international, 29*(s1), 74-81, 2009.

LAW, L. M. J.; LANDI, A.; MAGEE, W. C.; TYRRELL, D. L.; & HOUGHTON, M. Progress towards a hepatitis c virus vaccine. *Emerging microbes & infections, 2*(11), e79, 2013.

LÜ, W.; & XUE, Y. Prediction of hepatitis c virus non-structural proteins 5b polymerase inhibitors using machine learning methods. *Acta physico-chimica sinica, 27*(6), 1407-1416, 2011.

MENG, X.-Y.; ZHANG, H.-X.; MEZEI, M.; & CUI, M. Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design, 7*(2), 146-157, 2011.

MOHAMMAD, H.; & ZOHREH, A. In-silico prediction of rgs4 inhibitory activity of sometiadiazolidinone. *Int j med pharm, ,1* 2013.

MOHD HANAFIAH, K.; GROEGER, J.; FLAXMAN, A. D.; & WIERSMA, S. T. Global epidemiology of hepatitis c virus infection: new estimates of age-specific antibody to hcv seroprevalence. *Hepatology, 57*(4), 1333-1342, 2013.

MORADPOUR, D.; PENIN, F.; & RICE, C. M. Replication of hepatitis c virus. *Nature reviews microbiology, 5*(6), 453, 2007.

SHEPARD, C. W.; FINELLI, L.; & ALTER, M. J. Global epidemiology of hepatitis c virus infection. *The lancet infectious diseases, 5*(9), 558-567, 2005.

SOFIA, M. J.; CHANG, W.; FURMAN, P. A.; MOSLEY, R. T.; & ROSS, B. S. Nucleoside, nucleotide, and non-nucleoside inhibitors of hepatitis c virus ns5b rna-dependent rna-polymerase. *Journal of medicinal chemistry, 55*(6), 2481-2531, 2012.

TROTT, O.; & OLSON, A. J. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry, 31*(2), 455-461, 2010.

VEERASAMY, R.; RAJAK, H., JAIN, A.; SIVADASAN, S.; VARGHESE, C. P.; & AGRAWAL, R. K. Validation of qsar models-strategies and importance. *International journal of drug design & discovery, 3*, 511-519, 2011.

VRONTAKI, E.; MELAGRAKI, G.; MAVROMOUSTAKOS, T.; & AFANTITIS, A. Exploiting chembl database to identify indole analogs as hcv replication inhibitors. *Methods, 71*, 4-13, 2015.

WEI, Y.; LI, J.; QING, J.; HUANG, M.; WU, M.; GAO, F.; HUANG, W. Discovery of novel hepatitis c virus ns5b polymerase inhibitors by combining random forest, multiple e-pharmacophore modeling and docking. *Plos one, 11*(2), e0148181, 2016.

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