Research Article

Journal of Chemistry Letters
journal homepage: www.jchemlett.com
ISSN (online) 2717-1892

QSAR Model for Prediction of some Non-Nucleoside inhibitors of Dengue virus serotype 4 NS5 Using GFA-MLR Approach

Samuel Ndaghiya Adawara[a,*], Gideon Adamu Shallangwa[b], Paul Andrew Mamza[b], Abdulkadir Ibrahim[b]

[a] *Department of Pure and Applied Chemistry, Faculty of Science, University of Maiduguri, Maiduguri, Borno State, Nigeria.*
[b]*Department of Chemistry, Faculty of Physical Sciences, Ahmadu Bello University, Zaria, Kaduna State, Nigeria*

ARTICLE INFO

Article history:
Received 5 March 2020
Received in revised form 25 May 2020
Accepted 5 June 2020
Available online 10 July 2020

Keywords:
B3LYP/631G**
Quantitative structure-activity relationships
Genetic function algorithm
Molecular descriptors

ABSTRACT

B3LYP/631G** basis set of DFT quantum mechanical method was used to optimize the molecular geometry of some non-nucleoside inhibitors of dengue 4 virus. Molecular descriptors were mined from the optimized structure and used along with their experimental inhibitory activity (pIC_{50}) as the database for the study. Genetic function algorithm and multiple linear regressions were used to build a robust quantitative structure-activity relationship model. The statistically satisfactory quality of the model as evidenced by its validation parameters: R^2 = 0.971, R^2_{adj} = 0.961, $cRp^{\wedge 2}$ = 0.809 Q^2 = 0.944 and R^2_{pred} = 0.627. Thus, the model can be used to predict the activity of new chemicals within its applicability domain. The Average Broto-Moreau autocorrelation - lag 1 / weighted by mass, Centered Broto-Moreau autocorrelation - lag 2 / weighted by Sanderson electronegativities, Coefficient sum of the last eigenvector from Barysz matrix / weighted by van der Waals volumes, nhigh lowest polarizability weighted BCUTS and Fraction of sp3 carbons to sp2 carbons are the descriptors that influenced the anti-dengue activity of the studied compounds. The information obtained from the model in this work can be employed to optimize the anti-dengue activity of the compounds.

1.0 Introduction

Dengue infection is a mosquito-borne infection caused by a virus so-called Dengue virus (DNV) a member of the *Flavivirus* mostly found in tropical and sub-tropical regions around the world [1]. The virus is spreads to individuals by infected female Aedes genus, specifically Aedesaegypti or Aedesalbopictus [2-3]. In recent times, dengue infection has been reported in the Caribbean area, South America, and Europe [4].

Annually, at least 40 to 100 million persons are infected by DNV, and over half of the world's population at risk of infecting by this virus [4]. Infections caused by DNV can cause high fever and flu-like symptoms. These infections, in some instances, may also advance into a more acute stage know as dengue hemorrhagic fever and dengue shock syndrome [5-7]. Hence, DNV infections constitute a grave threat globally.

Out of the seven known nonstructural protein (NS) of DNV NS1 to NS5, only NS3 and NS5 have been considered so far as drug targets because they are essential to virus growth and demonstrate enzyme activity, which is desirable in regards to drug screening [8].

With all its fatal consequences, yet there are no effective drugs against dengue viruses [9-11]. This problem is also worsened by the persistent dispersal of these viruses to different geographic expanses as foretold more than a decade ago [12].

The nonexistence of particular medication for the treatment of Dengue fever presages great danger to the global health being of man, particularly in developing countries. The marked anti-dengue potentials of synthetic and medicinal plants have made their in silico structural modification geared towards the design of potent novel drug candidates a sine qua non. This will indeed provide an inroad to the development of the much-expected novel drugs against this virus.

In recent times, Computational methodologies have advanced as an imperative instrument for any drug

* Corresponding author.Tel.: +8067811759; e-mail: agapalawa@gmail.com

discovery program, playing a vital part in lead optimization from hit identification [13-15].

Q-SAR (quantitative structure-activity relationship), which is an informatics-based instrument predicts the activity of a dataset of molecules using regression such as molecules. Q-SAR has been employed extensively to estimate biological activities of important molecules.

In this study, the prominent method, which is called a quantitative structure-activity relationship (QSAR) has been established and used for predicting the biological activity of compounds by employing molecular structures and experimental biological activity data. Through this method, biological properties can be obtained easily without any experimental efforts for the synthesis of novel compounds [16].

2.0 Material and Methods

2.1 Software

The computation work in this study was carried out on an hp computer system, with the processor properties of Intel ® Core i3-5005U CPU Dual @ 2.00 GHz, 8 GB (RAM). The software packages used on the computer system include Spartan 14 Version (1.1.2) by Wavefunction Inc., Material studio software version (8.0), Chemdraw Ultra software Version (12.0.2), PADELDescriptor Version (2.20), Microsoft Excel 2013 version, and Drug Theoretics & Cheminformatics (DTC) laboratory software (MLRplusValidation1.3 and DatasetDIvision 1.2).

2.2 Dataset

The dataset used was reported literature to possess anti-dengue activity [17]. Their activity reported as IC_{50} (µM) was converted to IC_{50} (M) and later to Log $1/IC_{50}$ in other to moderate the skewness in the data [18]. The result is presented in Table 1 as pIC_{50} with the names of the compounds.

Table 1-IUPAC names and anti-dengue activity in logarithm unit (pIC_{50}) for the dataset compounds

No.	Name	Exp. pIC_{50}	Pred. pIC_{50}
1	2,2'-([1,1'-Biphenyl]-3,5-diyl)diacetic acid	3.752	3.538
2*	2,2'-(2'-Chloro-[1,1'-biphenyl]-3,5-diyl)diacetic acid	3.716	3.814
3	2,2'-(3'-Chloro-[1,1'-biphenyl]-3,5-diyl)diacetic acid	3.850	4.022
4	2,2'-(4'-Chloro-[1,1'-biphenyl]-3,5-diyl)diacetic acid	4.207	4.237
5	2,2'-(5-(Furan-2-yl)-1,3-phenylene)diacetic acid	3.675	3.477
6	2,2'-(5-(Thiophen-2-yl)-1,3-phenylene)diacetic acid	4.823	4.838
7	2,2'-(5-(5-Chlorothiophen-2-yl)-1,3-phenylene)diacetic acid	4.585	4.132
8*	2,2'-(5-(5-Methylthiophen-2-yl)-1,3-phenylene)diacetic acid	4.017	2.626
9	2,2'-(5-(5-Bromothiophen-2-yl)-1,3-phenylene)diacetic acid	4.408	4.639
10	2,2'-(5-(5-Cyanothiophen-2-yl)-1,3-phenylene)diacetic acid	3.701	4.113
11	2,2'-(5-(5-(3-Hydroxyprop-1-yn-1-yl)thiophen-2-yl)-1,3 phenylene)diacetic acid	5.769	5.827
12	2,2'-(5-(5-(4-Hydroxybut-1-yn-1-yl)thiophen-2-yl)-1,3 phenylene)diacetic acid	4.537	4.687
13	2-(3-(5-(3-Hydroxyprop-1-yn-1-yl)thiophen-2-yl)-4-methoxyphenyl)acetic acid	5.124	4.964
14	3-(5-(3-Hydroxyprop-1-yn-1-yl)thiophen-2-yl)-4-methoxybenzoic acid	5.638	5.634
15*	5-(5-(3-Hydroxyprop-1-yn-1-yl)thiophen-2-yl)-4-methoxy-2-methylbenzoic acid	6.275	4.443
16	3-(5-(2-Methoxy-5-(1H-tetrazol-5-yl)phenyl)thiophen-2-yl)prop-2-yn-1-ol	5.619	5.826
17	3-(5-(5-(3-Hydroxyprop-1-yn-1-yl)thiophen-2-yl)-4-methoxy-2-methylphenyl)-1,2,4-oxadiazol-5(4H)-one	5.494	5.514
18	3-(5-(5-(1H-Imidazol-2-yl)-2-methoxy-4-methylphenyl)thiophen-2-yl)prop-2-yn-1-ol	4.356	4.440
19*	5-(5-(3-Hydroxyprop-1-yn-1-yl)thiophen-2-yl)-4-methoxy-2-methyl-*N*-(phenylsulfonyl)benzamide	6.468	6.894
20	N-((5-(5-(3-hydroxyprop-1-yn-1-yl)thiophen-2-yl)-4-methoxy-2-methylphenyl)sulfonyl)acetamide	5.602	5.471
21	5-(5-(3-Hydroxyprop-1-yn-1-yl)thiophen-2-yl)-4-methoxy-2-methyl-*N*- (phenylsulfonyl)benzamide	6.769	6.894
22	5-(5-(3-Hydroxyprop-1-yn-1-yl)thiophen-2-yl)-2,4-dimethoxy-*N*-(phenylsulfonyl)benzamide	6.602	6.594
23*	5-(5-(3-Hydroxyprop-1-yn-1-yl)thiophen-2-yl)-2,4-dimethoxy-*N*-((3 methoxyphenyl)sulfonyl)benzamide	6.769	5.416
24	4-Chloro-5-(5-(3-hydroxyprop-1-yn-1-yl)thiophen-2-yl)-2-methoxy-N-((3-methoxyphenyl)sulfonyl)benzamide	6.853	6.646
25	5-(5-(3-Hydroxyprop-1-yn-1-yl)thiophen-2-yl)-4-methoxy-2-methyl-*N*-(quinolin-8-ylsulfonyl)benzamide	7.638	7.509

2.2.1 Molecular structure optimization and descriptor calculation

Molecular structures of the compounds were properly drawn using Chem-Draw Ultra software V12.0.2 and subsequently exported to Spartan for optimization, with the aim of computing their equilibrium geometries[19], DFT B3LYP/6-31G** quantum mechanical technique was used i.e. Becke's (3) exchange functional (B3) [20] combined with Lee-Yang-Parr correlation functional (LYP) [21] by means of 6-31G** basis set [22]. Many have used this method. Optimized molecules from Spartan program were saved as SDF format, and then exported to PaDEL descriptors software. This software computes diverse molecular descriptors [23, 19].

2.2.2 Dataset pretreatment

In this, all descriptor columns having constant numerical values were eliminated. In a pair of descriptors having a correlation coefficient greater than 0.8, one was discarded whose correlation coefficient with the activity value is

less significant. The pretreatment process was carried out to reduce redundancy and the selection of best descriptors [18].

2.2.2 Dataset division

Kennard-Stone algorithm (KS) available in DatasetDIvision 1.2 [24] was used to divide the dataset into training and test set. KS has been reported to produce excellent data division [25].

2.3 Selection of optimal descriptor and multi-co-linearity analysis

This encompasses the selection of descriptors blend between remaining descriptors that can predict the dependence between descriptors and the dependent variable. The genetic algorithm of multiple linear regressions (GA-MLR) methods was used in this study. This method has been used to examine the correspondence between biological activity and physicochemical properties of a set of bioactive compounds. It describes how a Y-variable relates to two or more X-variables. Material Studio version 8.10 was used to generate equation (model) in this research, which is expressed as:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \mathcal{E} \quad (1)$$

Where Y_i, x_i, β_o, β_p, and $\mathcal{E}$ are dependent variables, explanatory variable, y-intercept constant term, slope coefficient for each explanatory variable, and model error term called residual respectively [26].

This method has the importance of generating more than one combination of descriptors that can be used to build a model. It provides the user the control over the equation length and utilizes a lack-of-fit (LOF) function to hinder over-fitting and moderate redundancy in a model [27]. The validity of the model evaluated by statistical methods.

Also, the presence of a high degree of correlation among the descriptors that make up the model selected by GFA was evaluated with the variance inflation factor (V.I.F) value for each descriptor represented by equation 2.

$$(V.I.F)_i = \left(R_{ij}^2\right)^{-1} \quad (2)$$

where R^2_{ij} in equation 2 represents the correlation coefficient of the multiple regression between the

descriptor i and the remaining j descriptors within the model [28].

2.3.1 Q-SAR model and authentication

The descriptors that constitute the best blend selected by the GFA were copied into a separate spreadsheet for individual training and test sets. Then, training and test set data matrices were imported into the MLRplusValidation1.3 software for various internal and external statistical validation as recommended [24].

2.3.2 Models domain of applicability

The level of extrapolation approach based on compounds leverage (hi) values and standardized residual (SDR) obtained from the model was used to define the applicability domain (AD) of the Q-SAR model [29]. Compounds h_i are obtained as the diagonal component of hat matrix **H**:

$$\boldsymbol{H = m\left(m^T m\right)^{-1}.m^T} \quad (3)$$

where **m** is the descriptor matrix and $\mathbf{m^T}$ is the transpose of m, and SDR was obtained using equation 4:

$$SDR = \frac{\hat{y} - y}{\sqrt{\frac{\sum_{i=1}^{n}(\hat{y}-y)^2}{n}}} \quad (4)$$

Where y and $\hat{y}$ are observed and predicted activity value for either of the dataset respectively and n is the number of compounds in the set involved. Model AD was well-defined by the borderline $0 < h_i < h^*$ and $-3 < SDR < 3$. Where h* is the cautionary leverage h* obtained using equation 5:

$$h^* = \frac{3(q+1)}{n} \quad (5)$$

Where q is the number of descriptors in the model and n is the number of compounds that made up the training set. A quick visual assessment of the model AD is a plot of SDR versus h_i known as Williams plot was made [30].

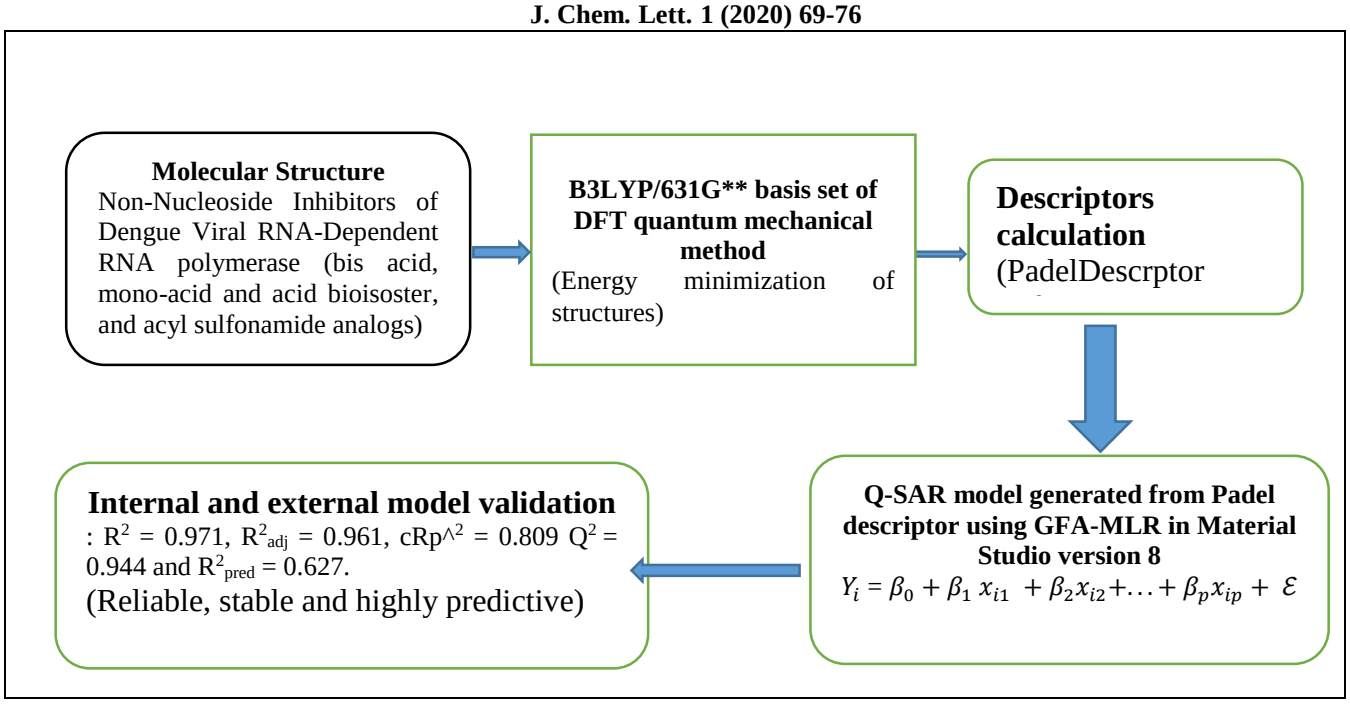

Figure 1. Schematic workflow of the study

3.0 Result and discussion

3.1 Dataset structure

About 20 training set and 5 test set compounds were reported by the dataset division technique used in the study. Figure 1 depicts the workflow for the whole process involved in this work. The test compounds are marked with the letter asterisks superscript in Table 1. Descriptive statistics executed on the two sets revealed that the test set maximum was less than the training set maximum; the test set minimum was greater than the training set minimum as shown in Table 2. Besides, other parameters reported in the table were similar for both sets. This indicated that the data division algorithm method employed in this study successfully obtain the test set data within the activity range of the training set. Dissimilarity analysis depicted in Figure 2 revealed that the test set descriptor spaces were in the range of training set descriptors space.

Table 2. Training and test set data descriptive statistics

Parameters	Training	Test
Mean	5.150	5.449
Standard Error	0.260	0.652
Sample Variance	1.356	2.129
Range	3.962	3.052
Minimum	3.675	3.716
Maximum	7.638	6.769

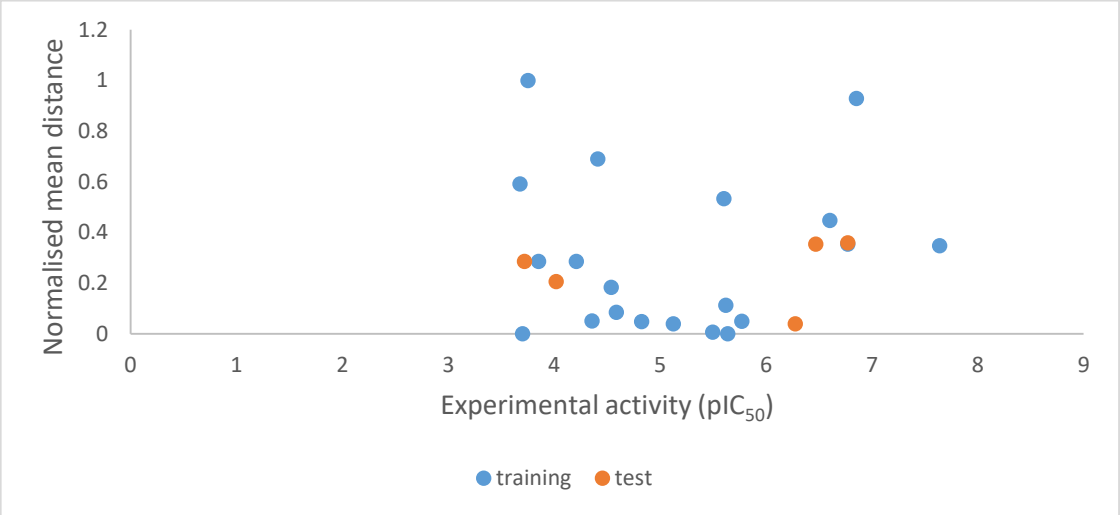

Figure 2. Diversity analysis of dataset compounds

Y = 0.037743709 * **AATS1m** + 0.493680190 * **ATSC2e** - 3.185535703 * **VE1_Dzv** - 5.253142608 * **BCUTp-1l** - 16.257556067 * **HybRatio** + 25.631233244 (6)

The Q-SAR model was built from 20 training set of compounds and 5 test set for external validation and it contained 5 descriptors. The model (Eq. 6) was used to predict the activity values (pIC_{50}) for both training and test presented in Table 1. The plot of SDR against the experimental activity value (Figure 3) showed that the residuals were evenly distributed around the line SDR = 0, indicating the absence of systematic error in the model [27] [31], so this method is statistically acceptable.

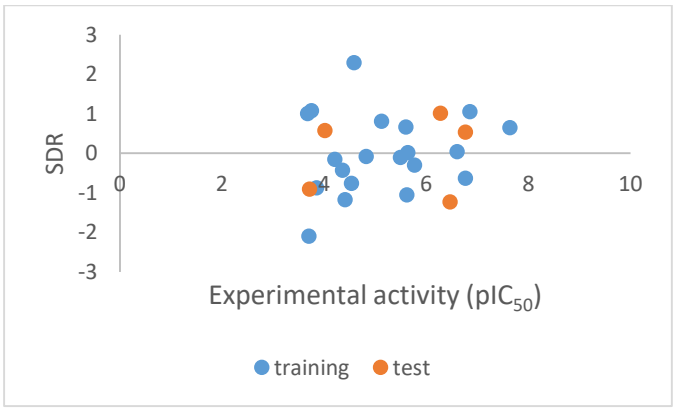

Figure 3. Distribution of residuals to the experimental pIC_{50} values for train and test set

The graph of predicted versus experimental activity (pIC_{50}) of the model (Figure 4) indicated that an undeviating relationship existed between the two variables and the model had good internal prediction ability. The multi-co-linearity investigation result revealed that the VIF values for descriptors in the model were in the range of 1-5 as shown in Table 3, specifying the model was satisfactory and void of the multi-co-linearity owing to coincidental correlation [28]. The correlation matrix for selected descriptors also is reported in the Table 3. As see, there is absence of dependency between the descriptors.

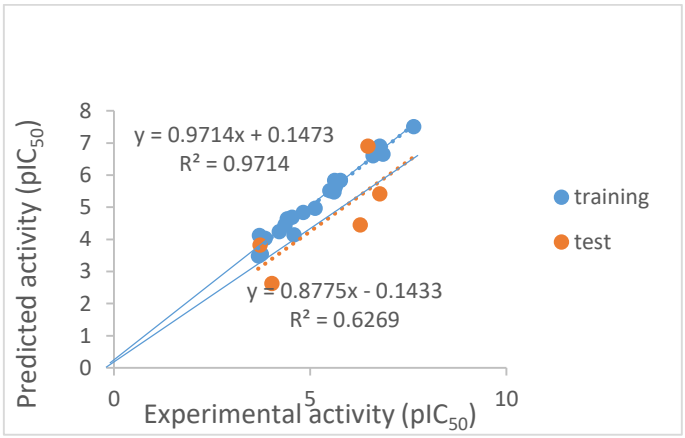

Figure 4. Predicted versus experimental activity value

Table 3. Descriptors correlation matrix and variance inflation factor

	AATS1m	*ATSC2e*	*VE1_Dzv*	*BCUTp-1l*	*HybRatio*	VIF
AATS1m	1					1.201
ATSC2e	-0.249	1				2.385
VE1_Dzv	-0.188	0.359	1			1.899
BCUTp-1l	-0.360	0.707	0.664	1		4.929
HybRatio	0.081	-0.217	-0.502	-0.599	1	1.892

3.3 Model validation parameters

Comprehensive validation statistical parameters computed for the model as well as the recommended threshold values are presented in Table 4. The result revealed that the values for R^2; R^2_{adj}; Q^2; R^2pred; and r^2 are more than 0.6. Hence, the model had a good internal and external predictive capability and it is not obtained by coincidental correlation [24]. The model also agreed with all Golbraikh and Tropsha criteria for a predictive model [32].

Table 4. Model Validation parameters and their statistical satisfactory threshold values

Parameter	Formula	Threshold value	Model score	Remark
	Internal validation			
$\mathbf{R^2}$	$\frac{\left[\sum\left\{(Y-\bar{Y})\times\left(\hat{Y}-\bar{\hat{Y}}\right)\right\}\right]^2}{\sum(Y-\hat{Y})^2\times\sum\left(\hat{Y}-\bar{\hat{Y}}\right)^2}$	$R^2 > 0.6$	0.9714	passed
$\mathbf{R^2_{adj}}$	$\frac{(N-1)\times R^2-p}{N-1-p}$	$R^2_{adj} > 0.6$	0.9612	passed
$\mathbf{Q^2}$	$1-\frac{\sum(Y-\hat{Y}_{loo})^2}{\sum(Y-\bar{Y})^2}$	$Q^2 > 0.6$	0.9439	Passed
	Random model			
$\mathbf{^cR^2_p}$	$R^2\times\left(1-\sqrt{\lvert R^2-\bar{R}^2_r\rvert}\right)$	$^cR^2_p > 0.6$	0.8097	Passed
	External validation			
$\mathbf{R^2_{Pred}}$	$1-\frac{\sum(Y_{ext}-\hat{Y}_{ext})^2}{\sum(Y_{ext}-\bar{Y})^2}$	$R^2_{pred} > 0.6$	0.6269	Passed
$\mathbf{r^2}$	Coefficient of determination for the plot of predicted versus observed for test set	$r^2 > 0.6$	0.6269	Passed
$\mathbf{r_0^2}$	**r^2 at zero intercept**		**0.3896**	**Passed**
$\mathbf{r_0'^2}$	r^2 for the plot of observed versus predicted activity for the test set at zero intercept		0.6301	Passed
$\mathbf{\lvert r_0^2-r_0'^2\rvert}$		$\lvert r_0^2-r_0'^2\rvert < 0.3$	0.2368	Passed
$\boldsymbol{k}$	The slope of the plot of predicted versus observed activity for test set at zero intercept	$0.85 < k < 1.15$	1.1341	Passed
$\mathbf{\frac{r^2-r_0^2}{r^2}}$		$\frac{r^2-r_0^2}{r^2} < 0.1$	0.3785	Passed
$\mathbf{k'}$	Slope of the plot of observed versus predicted activity at zero intercept	$0.85 < k' < 1.15$	0.8526	Passed
$\mathbf{\frac{r^2-r_0'^2}{r^2}}$		$\frac{r^2-r_0'^2}{r^2} < 0.1$	0.0009	Passed

Y is the observed activity value for the training set, $\bar{Y}$, the average of the observed activity for training set $\hat{Y}$, Predicted activity for the training set, $\hat{Y}_{loo}$ leave one out cross-validation predicted activity for training, Y_{ext} observed activity for the test set, and $\hat{Y}_{ext}$ predicted activity for the test set

3.4 Model applicability domain

The cautionary leverage for the model h* was 0.90 as obtained from equation 3. More so, the AD of the model is defined by a square area bounded by 0< h < 0.9 and -3 < SDR < 3 as shown graphically by the models William's plot (Figure 5). It could be seen that all the dataset compounds were within the boundary limit of the AD of the model. Thus, the dataset does not contain any outliers.

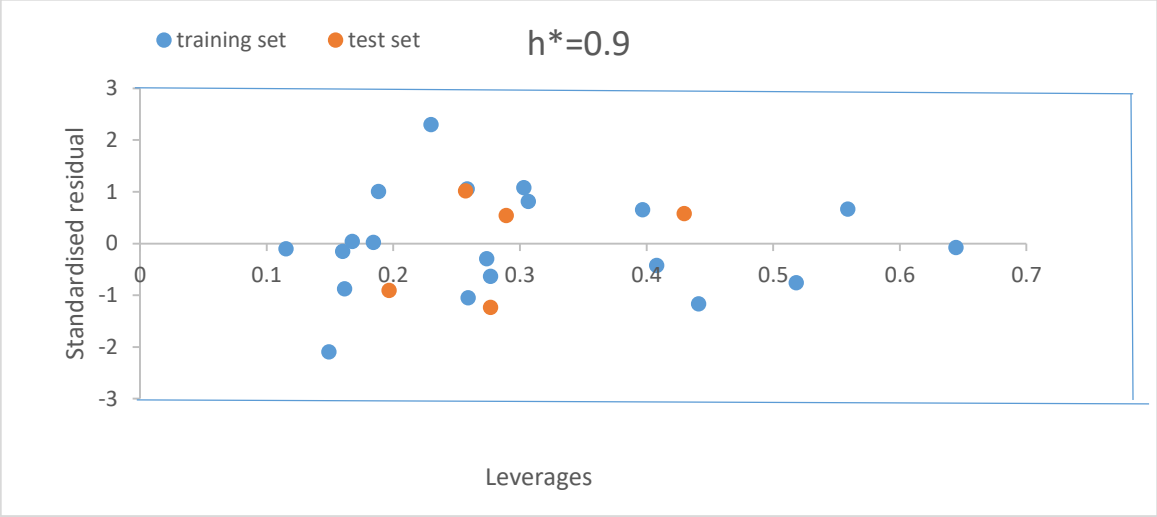

Figure 5. William's plot for the model

3.5 Interpretation of descriptors

Calculated descriptors for each molecule in the dataset are contained in equation 6. Average Broto-Moreau autocorrelation - lag 1 / weighted by mass (AATS1m) is the first descriptor in the model and is positively correlated to the activity of the studied compounds, this entails an increase in its value could improve the biological activity of the compounds. Centered Broto-Moreau autocorrelation - lag 2 / weighted by Sanderson electronegativities (ATSC2e) it is positively correlated to the activity of studied compounds, which also signifies improvement in biological activity with an increase in the value of such descriptor. Also, coefficient sum of the last eigenvector from Barysz matrix / weighted by van der Waals volumes (VE1_Dzv), nhigh lowest polarizability weighted BCUTS (BCUTp-1l) and Fraction of sp3 carbons to sp2 carbons (HybRatio) are the descriptors contained in the model which correlate negatively with the biological activity (pIC_{50}) which signifies increase in biological activity with decrease in the value of the descriptors.

4.0 Conclusion

This work establishes the quantitative structure-activity relationship (Q-SAR) between some non-nucleoside inhibitors and their pIC_{50} against DNV-4 NS5. The result revealed AATS1m; ATSC2e; VE1_Dzv; BCUTp-1l and HybRatio molecular descriptors to influence the anti-dengue activity of the studied compounds. These descriptors showed that increasing the electronegativity of the molecules on the addition of electronegative elements in the molecular system and the decrease in the number of sp3 carbons to sp2 carbon could improve the anti-dengue activity of the studied compounds. The model produced in the study also had a good performance in terms of its validation parameters and can be used to screen compounds for anti-dengue activity.

Acknowledgement

The authors gratefully acknowledged the technical effort of Dr. David Ebukar Arthur, Member of the department of Pure and Applied Chemistry, University of Maiduguri.

References

[1] S. S. Swain, D. Dude, Anti-dengue medicinal plants: a mini review. *Res. Rev. J Pharmacogn Phytochem.*, 1 (2013) 5-9.

[2] M. E. Beatty, A. Stone, W. D. Fitzsimons, J. N. Hanna, S. K. Lam, S. Vong, Best practices in dengue surveillance: a report from the Asia-Pacific and Americas Dengue Prevention Boards. *PLoS. Negl. Trop. Dis*., 4 (2010) 890.

[3] Z. Fatima, M. Idrees, M. A. Bajwa, Z. Tahir, O. Ullah, M.Q. Zia, Serotype and genotype analysis of dengue virus by sequencing followed by phylogenetic analysis using samples from three mini outbreaks-2007-2009 in Pakistan. *BMC Microbiol.*, 11 (2011)200.

[4] E. P. Toepak, U. S. F. Tambunan, In silico design of fragment-based drug targeting host processing α-glucosidase i for dengue fever. *Materials Science and Engineering*, 172 (2017) 01201.

[5] C. Nitsche, S. Holloway, T. Schirmeister and C. D. Klein, Biochemistry and medicinal chemistry of the dengue virus protease. *Chem Rev.*, 114 (2014) 11348–11381.

[6] A. J. Stevens, M. E. Gahan, S. Mahalingam, P. A. Keller, The medicinal chemistry of dengue fever. *J. Med. Chem.*, 52 (2009) 7911–7926.

[7] S. P. Lim, Q. Y. Wang, C. G. Noble, Y. L. Chen, H. Dong, B. Zou, F. Yokokawa, S. Nilar, P. Smith, D.

Beer, J. Lescar, P. Y. Shi, Ten years of dengue drug discovery: Progress and prospects. *Antiviral Res.*, 100 (2013) 500–519.

[8] B. Canard, Antiviral Research and Development against Dengue Virus CDC (2010) The Dengue Update: Dengue Outbreaks Worldwide. *Centre for Disease Control and Prevention*, 2 (2017) (1.1).

[9] N. M. Nguyen, C. N. Tran, L. K. Phung, K. T. Duong, K. T.; Huynh, Hle. A.;J. Farrar, Q. T. Nguyen, H. T. Tran, C. V. Nguyen, L. Merson, L. T. Hoang, M. L. Hibberd, P. P. Aw, A. Wilm, N. Nagarajan, D. T. Nguyen, M. P. Pham, T. T. Nguyen, H. Javanbakht, K. Klumpp, J. Hammond, R. Petric, M. Wolbers, C. T. Nguyen, C. P. A. Simmons, Randomized, double-blind placebo controlled trial of balapiravir, a polymerase inhibitor, in adult dengue patients, *J. Infect. Dis., 207 (*2013) 1442-1450.

[10] Bhatt, S.; Gething, P.W.; Brady, O.J.; Messina, J.P.; Farlow, A.W.; Moyes, C.L.; Drake, J.M.;Blanton, R.E., L. K. Silva, V. G. Morato Genetic ancestry and income are associated with dengue hemorrhagic fever in a highly admixed population, *European Journal of HumanGenetics*, vol. 16, no. 6, (2008) 762–765,

[11] M. G. Guzman, G. Kouri, Dengue haemorrhagic fever integral hypothesis: confirming observations, 1987–2007. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 102 (6) (2008) 522–523

[12] P. M. Zanotto, E. A. Gould, G. F. Gao, P. H. Harvey, E. C. Holmes, Population dynamics of flaviviruses revealed by molecular phylogenies. *Proceedings of the National Academy of Sciences of the United States of America*, 93 (2) (1996) 548– 553.

[13] A. S. Ndaghiya, A. I. Ogadimma, S. Sani, Modelling of some Schiff bases as anti-Salmonella typhi drugs: A QSAR approach. *Journal of Computational Methods in Molecular Design*, *5*(4) (2015) 147-157.

[14] S. N. Adawara, P. Mamza, G. A. Shallangwa, I. Abdulkadir. Anti-Dengue potential, Molecular Docking Study of Some Chemical constituents in the leaves of Isatis tinctoria. *Chemical Review and Letters*, 3(3) (2020) 104-109.

[15] J. C. Madden, M. T. Cronin, Structure-based methods for the prediction of drug metabolism. *Expert opinion on drug metabolism & toxicology*, *2*(4) (2006) 545-557.

[16] Y. C. Martin, *Quantitative Drug Design*, Marcel Dekker, New York, NY, USA, 1978.

[17] F. Yokokawa, S. Nilar, C. G. Noble, S. P. Lim, R. Rao, S. Tania, G. Wang, G. Lee, J. Hunziker, R. Karuna, U. Manjunatha, Discovery of potent non-nucleoside inhibitors of dengue viral RNA-dependent RNA polymerase from a fragment hit using structure-based drug design. *J. Medi. Chem*., 59(8) (2016) 3935-52.

[18] A. Tropsha, Best practices for QSAR model development, validation, and exploitation. *Mol. Inform.*, 29(6–7) (2010) 476–488.

[19] M. Abdullahi, G. A. Shallangwa, A. Uzairu, In silico QSAR and molecular docking simulation of some novel aryl sulfonamide derivatives as inhibitors of H5N1 influenza A virus subtype. *Beni-Suef University Journal of Basic and Applied Sciences*, *9*(1) (2020) 2.

[20] A. D. Becke, Density-functional thermochemistry III. The role of exact exchange. *J. Chem. Phys.*, 98 (1993) 5648–5652.

[21] C. Lee, W. Yang, R. G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev.*, B 37(1988) 785.

[22] A. Schäfer, C. Huber, R. Ahlrichs, Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *J. Chem. Phys.*, 100 (1994) 5829–5835.

[23] C.W. Yap, PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.*, **32**(7) (2011) 1466–1474.

[24] R.W. Kennard, L.A. Stone, Computer aided design of experiments. *Technometrics* 11(1) (1969) 137–148.

[25] P. Ambure, R.B. Aher, A. Gajewicz, T. Puzyn, K. Roy, "NanoBRIDGES" software: Open access tools to performQ-SARand nano-QSAR modeling. *Chemom. Intell. Lab. Syst*., 147 (2015) 1–13.

[26] N. Rob, Tutorial 6: Linear Regression, (2014) 1–14.

[27] D. E. Arthur, A. Uzairu, P. Mamza, S. E. Abechi, G. Shallangwa, Insilco study on the toxicity of anti-cancer compounds tested against MOLT-4 and p388 cell lines using GA-MLR technique. *Beni-Suef Univ. J. Basic Appl. Sci.*, 5 (2016b) 320–333.

[28] A. Beheshti, E. Pourbasheer, M. Nekoei, S. Vahdani, Q-SAR modeling of antimalarial activity of urea derivatives using genetic algorithm–multiple linear regressions. *J. Saudi Chem. Soc*., 20 (2016) 282–290.

[29] T. I. Netzeva, A. P. Worth, T. Aldenberg, R. Benigni, M. T. Cronin, P. Gramatica, J. S. Jaworska, S. Kahn, G. Klopman, C. A. Marchant, Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships, *ATLA*, 33 (2005) 155–173.

[30] S. Dimitrov, G. Dimitrova, T. Pavlov, N. Dimitrova, G. Patlewicz, J. Niemela, O. Mekenyan, A stepwise approach for defining the applicability domain of SAR and Q-SAR models, *J. Chem. Inf. Model.* 45 (2005) 839–849.

[31] S. Alimohammadi, A. Hamidi, P. Pargolghasemi, N. Nourani, and M. S. Hoseininezhad-Namin, QSAR study of antiproliferative drug against A549 by GA-MLR and SW-MLR methods. *Chemical Review and Letters*, *2*(4) (2019)193-198.

[32] A. Golbraikh, A. Tropsha, Beware of q2! J. *Mol. Graph. Model*, **20**(4) (2002) 269–276.