RESEARCH

Open Access

Molecular docking and QSAR theoretical model for prediction of phthalazinone derivatives as new class of potent dengue virus inhibitors

Samuel Ndaghiya Adawara[1*] (iD), Gideon Adamu Shallangwa[2], Paul Andrew Mamza[2] and Abdulkadir Ibrahim[2]

Abstract

Background: Dengue fever is a key public health unease in various tropical and sub-tropical regions. The improvement of existing agents that can inhibit the dengue virus is therefore of utmost importance. In this work, the QSAR study was carried out on 25 molecules of phthalazinone derivatives which have been reported to possess excellent dengue virus inhibitory activity. Density functional computational technique was used in the optimisation of the molecules with the basis set at theory level (B_3LYP, 6-31G*) respectively. The multiple linear regression (MLR) model was built using genetic function approximation (GFA) in the material studio software package. Also, in this study, molecular docking simulation was carried between dengue virus serotype 2 protease (PDB CODE: 6mol) and some selected phthalazinone derivatives (compounds 1, 2, 7, 11, and 21).

Results: The model was robust as evidenced by validation and robustness statistical parameter which include predicted R^2_{pred}, adjusted R^2_{adj}, cross-validated Q^2 and R^2 regression coefficient, etc ($R^2_{pred.} = 0.71922$, $R^2_{adj.} = 0.939699$, $Q^2_{CV} = 0.905909$, $R^2 = 0.955567$) respectively. The molecular docking studies conducted in this study have outlined the binding affinities of the selected compounds (1, 2, 7 11, and 21) which are all in good correlation with their respective pIC_{50} values. The free binding affinities of the selected compounds were found to be ($- 8.7$, $- 8.8$, $- 8.7$, $- 8.3$, and $- 8.9$ kcal/mol) respectively, compound 21 with the binding affinity of $- 8.9$ kcal/mol had the best binding free energy with the protease relative to other compounds under consideration.

Conclusion: The MLR-GFA model study alongside with the molecular docking analysis has essentially provided a valuable and in-depth understanding as well as knowledge for the development of novel chemical compounds with enhanced inhibitory potential against the dengue virus serotype 2 (DNV-2). Hence, the developed model can be applicable in predicting the anti-dengue activity of a new set of chemical compounds that fall within its applicability domain.

Keywords: Dengue, QSAR, GFA-MLR, NS, Phalazinones, Descriptors, Validation, Docking

---

* Correspondence: agapalawa@gmail.com
[1]Department of Pure and Applied Chemistry, Faculty of Science, University of Maiduguri, Maiduguri, Borno State P.M.B. 1069, Nigeria
Full list of author information is available at the end of the article

1 Background

Dengue infection is a mosquito-borne infection caused by a virus called Dengue virus (DNV) a member of the *Flavivirus* found predominantly in tropical and subtropical areas around the world [1]. DNV spreads among humans by the infected *Aedes aegypti* or *Aedes albopictus* specifically female of the Aedes genus [2, 3].

They are classified into four different but closely inter-related serotypes (DNV-1, DNV-2, DNV-3, and DNV-4). Infection with one serotype confers life-long immunity; however, secondary infection by a different serotype can increase the risk of developing severe dengue, because cross-immunity to the other serotypes is only partial and temporary [4].

Dengue virus serotype 2 (DNV-2) is responsible for the major infections and accounts for the largest death rate and hence considered as the virulent strain among other serotypes of the four [5]. The risk of developing dengue shock syndrome (DSS) and dengue hemorrhagic fever (DHF) is associated with infection by multiple serotypes as a result of antibody-dependent enhancement [6].

The dengue virus has about seven non-structural (NS) proteins, NS-1, NS-2A, NS-2B, NS-3, NS-4A, NS-4B, and NS-5. The Flavivirus proteases are evolutionally conserved and exceedingly stable. In NS-3, there is the presence of N-terminal serine protease domain, although inactive but becomes active upon complexation with NS-2B. Also, this protease can assume a structural conformation of either open or closed. In the closed state that is catalytically active, NS-2B is completely tied round NS3 and becomes a component of the active site. In the open and inactive conformation, NS-2B has partially bound to NS-3 and far away from the active site and hence inactive. The highly conserved Flavivirus NS-2B/NS-3 protease is necessary for viral replication and hence a druggable target [7].

In recent times, dengue infection has been reported in the Caribbean region, South America, and Europe [8]. Thus, DNV infection constitutes a serious threat globally. Approximately 40 to 100 million people are infected by DNV annually and more than 50% of the population of the world are at high risk of the infection by this virus [8]. These infections, in some persons, can progress into a more acute stage known as (DHF) and (DSS) [9–11], thus constitute a serious fatal threat in major dengue cases, around 2.5% from 500,000 clinical cases [9].

Despite these fatal consequences of DNV infection likewise the possible imminent outbreaks, there have been no antiviral drugs to prevent or treat DNV infections [12–14]. This problem is also worsened by the long-lasting dispersal of these viruses to diverse geographical regions as foretold more than 20 years ago [15]. The present certified dengue vaccine, Dengvaxia, has upraised alarms about the efficacy and increased danger of severe syndrome for seronegative persons at the phase of clinical trials [16].

Anti-dengue potentials of synthetic and medicinal plants have been described in the literature [17, 18].

Searching of biochemical libraries of these compounds is a real stride in the right track for the design of potent drug candidates against these viruses.

Quantitative structure-activity relationship (QSAR) is essential in drug improvement as it investigates the properties of the drug through its models which characterize mathematical equations correlating the response of chemicals (i.e., biological activity) with their structural and physicochemical information in the form of numerical quantities named descriptors [19]. QSAR studies are directed at developing correspondence models through a response of chemicals and chemical information data in a statistical approach.

For the reliability of QSAR models, they are subjected to various authentication tests to check for the consistency of the developed correlation models. After its development, a QSAR model is usually verified by employing multiple statistical validation strategies giving an estimation of its predictive strength and stability [19, 20].

QSAR analysis is an effective process for improving lead compounds and designing new drugs of the desired property. It is also used in predicting the biological activity of compounds based on the molecular descriptors of compounds recognized in the appropriate mathematical models.

The goal of this investigation is to obtain a model, to forecast the activity of the selected dataset and hopefully able to predict new compounds with improved activities capable of mitigating dengue viral replication.

A better understanding and insight of the structural necessities for the design of effective and specific inhibitors against flaviviral protease would contribute to the development of targeted therapies for infections by these viruses.
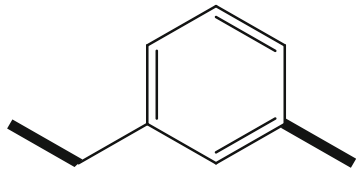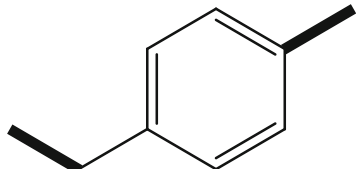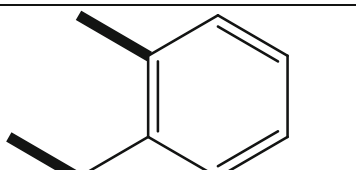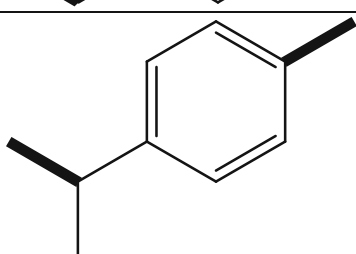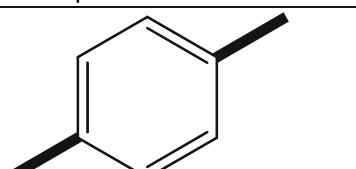
2 Methods

2.1 Data collection

The dataset used in this study was phthalazinone derivatives reported in a published literature to experimentally possess anti-dengue activity [21]. It is recommended that biological activities values such as IC_{50} and EC_{50} used for the building of any given QSAR model should be obtained from the same species using the related procedures [22, 23].

2.2 Molecular geometry optimization

The two-dimensional (2-D) structures of the obtained compounds presented in Table 1, 2, 3 and 4 were drawn using the ChemDraw software [24]. The spatial conformations of the compounds were

Table 1 Molecular structure of phthalazinone derivatives with the substitution of "A" ring and their biological activity (p IC_{50})

SN	A ring	p IC_{50}
1		6.19382
2		6.207608
3		5.548214
4		5.896196
5		5.872895

exported from 2-D structure to three-dimensional (3-D) structure using the Spartan 14 V1.1.4 by Wavefunction programming package. The 3-D structures were geometrically optimized by minimizing energy. In the process, the chemical structures were first of all minimized by a molecular mechanics force field to remove tension energy of the molecules' conformation. Density functional theory (DFT) technique was further employed using the Becke's three-parameter exchange functional (B3) hybrid alongside the Lee, Yang, and Parr correlation functional (LYP), termed as B3LYP hybrid functional, for thorough geometric optimization of the structures. The Spartan files of all the optimized molecules were then saved in SD file format, which is one the readable input format in PaDEL-Descriptor software [25].

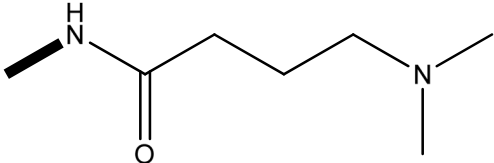### 2.3 Biological activities (pIC$_{50}$)
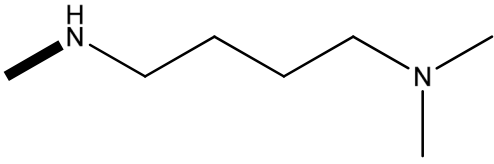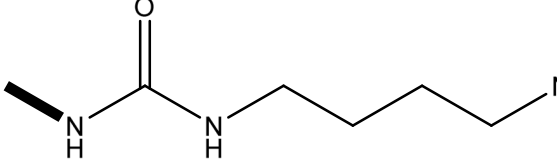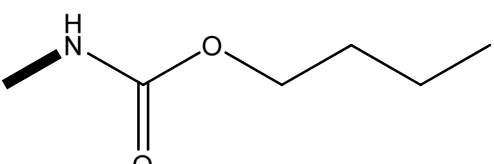
The obtained biological activities of phthalazinone derivatives against cytoplasmic DNV-RNA replication measured in IC_{50} (μM) were converted to the logarithm unit (pIC$_{50}$) using the Eq. (1) to increase the linearity of activity values and approach normal statistical distribution. The observed structures and the biological activities of these compounds are presented in Fig. 1a–d and Tables 1, 2, 3 and 4 respectively.

$$pIC_{50} = -\log(IC_{50}) \qquad (1)$$

2.4 Molecular descriptor generation

Molecular descriptors which are the mathematical values describing the properties of a molecule we

Table 2 Molecular structure of phthalazinone with a substituent on benzyl moiety their biological activity

SN	R	p IC$_{50}$
6	NH$_2$	4.699839
7		6.000
8		5.978811
9		5.361511
10		5.853872

determined. Quantum chemical descriptors calculation for all the 25 molecules of phthalazinone derivatives were calculated using PaDEL-Descriptor software V2.20. A total number of about 1870 molecular descriptors were calculated and combined with those obtained from the 3-D structure by the Spartan program software.

2.5 Splitting of data-set into modeling train and external evaluation test sets
To build the QSAR models, the data set which is the chemical compound was separated into two sets in the ratio of 80:20, the train set and test set respectively. The train set is used for building the QSAR model; it contains 80% of the entire chemical compounds under consideration. While the test set which constitutes the remaining 20% of the total chemical compound data set was not involved in the building of the QSAR model but to ascertain the analytical quality of the built model [26].
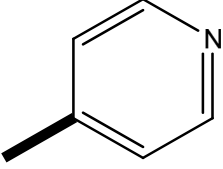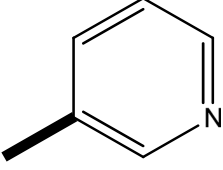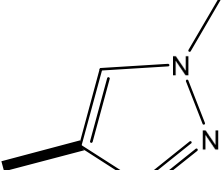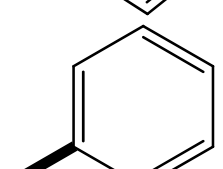
2.6 MLR-GFA model building
Statistical analysis by genetic function approximation (GFA) techniques of the Material Studio software 8.0 version was used to build the models based on multiple linear regression (MLR). The MLR is used to establish a direct relationship between a dependent variable Y (pIC$_{50}$) and independent variable X (molecular descriptors). The model fits well such that sum of the square difference between the experimental and predicted pIC$_{50}$ values is lessen. In regression analysis, a contingent mean of dependent variable (pIC$_{50}$) Y relies on (Descriptors) X. MLR examination utilizes and also lengthens this idea to combine more multiple autonomous variables, and regression equation assumes the form:

$$Y = k1x1 + k2x2 + k3x3 + ...C \qquad (2)$$

where Y is the dependent variable, "k"s are regression coefficients for corresponding "x"s (independent variables), and "C" is intercept or a regression constant.

Table 3 Molecular structure of phthalazinone with substitution of B ring and their biological activity

SN	B ring	pIC$_{50}$
11		4.640354
12		5.205512
13		5.185087
14		5.571865

The GFA calculates a fitness function identified as a lack of fit (L.O.F). This fitness function is not used by the system for indicating equations that are the best model, rather estimate the superiority of the models previously built by the system thus helping in deciding the models to use based on quality. Quality of the model is inversely proportional to the

Table 4 Molecular structure of phthalazinone with substitution of phenyl moiety and their biological activity

SN	R	p IC$_{50}$
15	4-OCH$_3$	5.712198
16	4-CN	6.154902
17	4-CF$_3$	5.832683
18	3-F	5.821023
19	3-OCH$_3$	6.522879
20	3-CN	5.764472
21	3-CF$_3$	6.886057
22	2-F	5.879426
23	2-OCH$_3$	5.853872
24	2-CN	5.258061
25	2-CF$_3$	5.701147

LOF value and it is computed using the mathematical expression:

$$\text{LOF} = \frac{\text{LSE}}{\left(1 - ((c + dp)/M)\right)^2} \qquad (3)$$

In this equation, LSE is the least-squares error of the model, c is the number of descriptors in the model, d is the smoothing parameter (which has a default value of 1.0), p is the sum of all descriptors, and M is the total number of compounds involved in the model building [27].

2.7 Model quality assessment

Predictive capacity and the robustness of the developed model was appraised internally and externally using statistical parameters such as R^2 (square correlation coefficient), Q^2_{CV} (cross-validation coefficient), R^2 pred. (external test set correlation coefficient), cR^2_p (coefficient of determination for Y-randomization), etc. The statistical validation parameters were compared with the minimum value suggested for a generally satisfactory QSAR model [28] presented in Table 5.
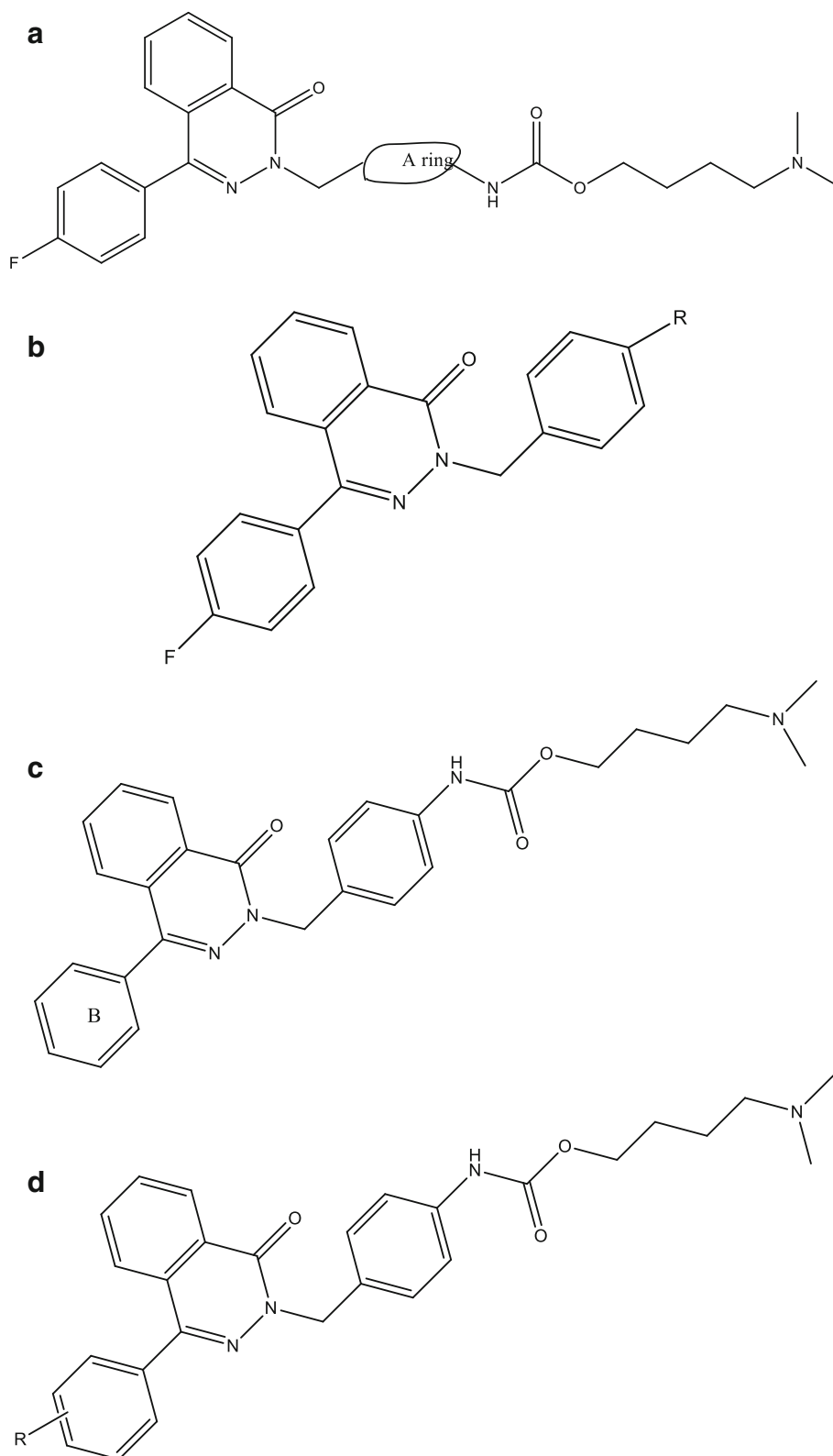
Fig. 1 a Parent structure with substitution on ring A of phthalazinone. **b** Parent structure of phthalazinone with substituents on the benzyl moiety. **c** Parent structure of phthalazinone with substitution of B ring of phthalazin core. **d** Parent structure of phthalazinone with a substituent on phenyl moiety

Table 5 Minimum suggested value of authentication parameters for a generally satisfactory QSAR model

Symbol	Name	Value
R^2	Coefficient of determination	≥ 0.6
$P_{(95\%)}$	Confidence interval at 95% confidence level	< 0.05
Q^2_{CV}	Cross validation coefficient	> 0.5
$R^2 - Q^2_{CV}$	Difference between R^2 and Q^2_{CV}	≤ 0.3
$N_{ext.\ test\ set}$	Minimum number of an external test set	≥ 5
cR^2_p	Coefficient of determination for Y-randomization	> 0.5

2.8 Validation of the QSAR model

The authentication of a QSAR model is mainly accomplished based on the chemical compound used in model development. It comprises activity estimate of the studied compounds and subsequent estimation of some validation parameters for verifying the accuracy of model predictions capacity. To judge the quality and goodness-of-fit of the model, internal validation is an ideal technique. Internal validation, which is regularly used to select a better model among contending models, was done using the data that create the models. The following internal validation parameters were calculated:

the cross-validated squared correlation coefficient (R^2_{CV} or Q^2):

$$R^2_{cv} = Q^2 = 1 - \frac{\sum \left(Y_{obs.} - Y_{pred.}\right)^2}{\sum \left(Y_{obs.} - \overline{Y}_{obs.}\right)^2} \quad (4)$$

$Y_{obs.}$ and $Y_{pred.}$ are the experimental and predicted response values respectively and $\overline{Y}_{obs.}$ is the average of the experimental biological activity value for the train set data. A satisfactory predictive model should have Q^2 value greater than 0.5 [29].

Also, another important parameter R^2, known as the determination coefficient: is square of the correlation coefficient between the observed and predicted response values of the training set compounds. It is the most used parameter and may be computed based on the following expression:

$$R^2 = 1 - \frac{\sum \left(Y_{obs.} - Y_{pred.}\right)^2}{\sum \left(Y_{obs.} - \overline{Y}\right)^2} \quad (5)$$

Given that experimental and predicted response values of biological activity have been designated as $Y_{obs.}$ and $Y_{pred.}$ respectively, while $\overline{Y}$ represents the average response value of the training set. R^2 measures the explanatory power of the model describing the variation in the activity value of molecules used in building the model. A perfect model has an R^2 value of unity (1) and as the value deviates from unity, the

fit quality of the model declines. A good model is expected to have an R^2 value at least equal to the threshold value of greater than or equal to 0.5 [30].

R^2_{adj}, known as explained variance: It is an adjusted form of determination coefficient which accounts for the effect of new explanatory variables in the model, by incorporating a degree of freedom to the model [30]. To reflect the described variance in a better way, R^2_{adj} is the candidate of choice since the inclusion of either relevant or irrelevant independent variables in multiple regression analysis often produces non-decreasing R^2 value [31]. It may be computed with the following expression:

$$R^2_{adj.} = \frac{(N-1) \times R^2 - p}{N - 1 - p} \quad (6)$$

where N gives the number of molecules in the data, R^2 is the determination coefficient, p is the number of descriptors in the model, and N-1-p is the degree of freedom [31].

The most essential consideration is the assessment of the generated model is external validation. Usually, prior to generating a QSAR model, the whole data set is shared into the train and test sets based on different algorithms The test set compounds are not involved in the training of the QSAR model and, hence, are used in external authentication procedure. The most recommended criteria for external validation are evaluated. In this, the biological activities of the test set compounds are predicted for determining the predictive power of the model. The most commonly used parameter for evaluating the predictive performance of the model is a coefficient of squared correlation (R^2pred.) for the test set that is evaluated by the following expression:

$$R^2_{pred.} = 1 - \frac{\sum \left(Y_{obs.(test)} - Y_{pred.(test)}\right)^2}{\sum \left(Y_{obs.(test)} - \overline{Y}_{(train)}\right)^2} \quad (7)$$

where $Y_{obs.\ (test)}$, $Y_{pred.(test)}$, and $_{(train)}$ are observed, predicted, and average values of biological responses for test and train sets, respectively. R^2 value varies from 0 to unity (1), and it is recommended that it should not be less than 0.6 [32, 33].

2.8.1 Statistical Y-scrambling evaluation

In this evaluation, random MLR models are created by haphazardly shuffling the dependent variable while keeping the independent variables untouched. The new QSAR models are expected to have considerably low R^2 and Q^2 values for numerous trials, which confirm that the developed QSAR models are robust. In the process, a very important parameter, cRp^2 is

likewise considered which should exceed a value of 0.5 for scaling through this test as recommended [29].

$$^{C}R_{p}^{2} = R^{2} \times \left(1 - \sqrt{\left| R^{2} - \overline{R}_{r}^{2} \right|}\right) \qquad (8)$$

where R^2 is the square correlation coefficient for the regression analysis of non-randomized model and $\overline{R}_{r}^{2}$ is the average of the square correlation coefficient for the regression analysis of all randomization scores.

2.8.2 Evaluation of the applicability domain of the model

The built QSAR model was also appraised based on the applicability domain (AD) method to prove that the model is robust and reliable to predict the (pIC_{50}) of compounds [34]. The leverage method was involved in defining and describing the applicability domain of models built [28]. Leverage of a given chemical compound, hi, is defined by Eq. 9:

$$hi = Z\left(Z^{T}Z\right)^{-1}.Z^{T} \qquad (9)$$

where Z is the descriptor matrix and Z^{T} is the transpose of Z, and standardized residual (SDR) was obtained as follows:

$$SDR = \frac{\hat{y} - y}{\sqrt{\frac{\sum_{i=1}^{n}(\hat{y} - y)^{2}}{m}}} \qquad (10)$$

where the experimental and predicted activity values for either of the datasets are represented by y and $\hat{y}$ respectively and m is the number of molecules in the set under consideration for each case. Also, model AD is defined by the boundary $0 < h_{i} < h^{*}$ and $-3 < SDR < 3$. Meanwhile, h^{*} indicates cautionary leverage value.

The cautionary leverage (h^{*}) is also the boundary of values for X outliers and is defined as:

$$h^{*} = \frac{3(g + 1)}{n} \qquad (11)$$

where g is the number of descriptors in the model and n is the number of compounds that are comprised of the train set used in building the model. A summary graphical evaluation of the model AD is the plot of SDR versus leverage h_{i} called William's plot was made [35].

2.8.3 Multi-collinearity test

The existence of a high degree of correspondence between the descriptors contained in the best descriptors arrangement reported by GFA was calculated with the variance inflation factor (VIF) value for each descriptor:

$$VIF_{i} = \frac{1}{1 - R_{ij}^{2}} \qquad (12)$$

where R^{2}_{ij} is the correlation coefficient of the multiple regression between the descriptor i and the remaining j descriptors in the model [36].

2.8.4 Mean effect

$$MEj = \frac{\beta_{j} \sum_{i=1}^{i=n} R_{ij}}{\sum_{j}^{m} \cdot \left(\beta_{j} \sum_{i}^{n} R_{ij}\right)} \qquad (13)$$

MFj is defined as the mean effect for the considered molecular descriptor j, while bj is the coefficient of the descriptor j, Rij represents the values for the target descriptors of each molecule, and m is the total number of descriptors in the model. The ME values demonstrate the relative implication of a descriptor, associated with other descriptors in the model. Its sign shows the variation direction in the estimations of the model as an effect of the descriptor values.

2.9 Molecular docking studies

To gain a detailed understanding of the nature of the interaction of compounds with the DNV-2 NS2B-NS3 protease, molecular docking was accomplished with the help of Auto Dock Vina of PyRx v software tool. The binding energy determination and visual analysis of the docked compound were accomplished using AutoDock Vina of PyRx and Discovery Studio visualization software, respectively. The crystal structure of the DNV protease was obtained from the protein data bank (PDB Code 6mol). All the heteroatoms associated with the receptor were removed from the three-dimensional structure of the DNV-2 (NS2B-NS3) receptor (Fig. 2a) and its structure was minimized, protonated, and saved in PDBQT format. Also, the 3D structures of the optimized compounds were converted to PDBQT format with the aid of AutoDock 4.2 software. The protein-ligand interaction was analyzed and visualized with the aid of Discovery studio visualization software [37].

3 Results

3.1 QSAR model quality

Based on the genetic algorithm of the descriptors, a multi-linear regression model was developed containing five (5) descriptors. The selected MLR-(GA) model is represented by Eq. (14)

$$\begin{aligned} pIC_{50} = {} & 0.009574310*ATS6e - 0.397088635*AATSC6m \\ & - 20.187750817*GATS2v - 0.005498349*VR1\_Dzv \\ & + 18.229472834*SpMax3\_Bhv - 52.345926233 \end{aligned}$$
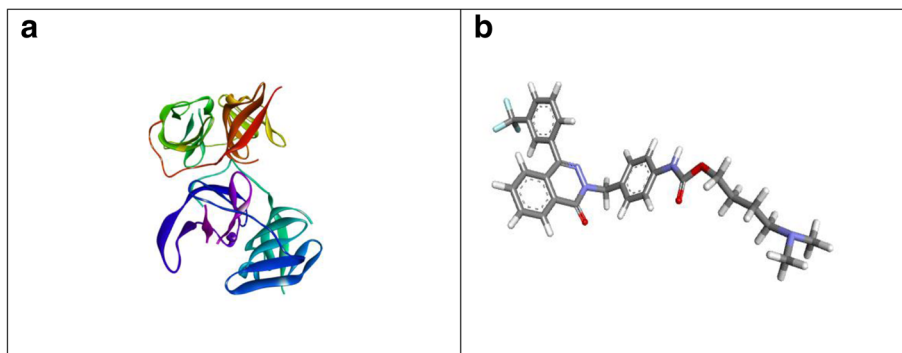
$$(14)$$

Fig. 2 Prepared structure of the target (NS2B-NS3) (**a**) and 3D structure of the prepared ligand (21) (**b**)

$$R^2_{\text{adj.}} = 0.939699, \quad R^2_{\text{pred.}} = 0.71922, cR = 0.749517,$$
$$Q^2_{\text{CV}} = 0.905909, \quad N_{\text{train}} = 20, \quad N_{\text{set}} = 5, ,$$
$$R^2 = 0.955567$$

From the above model (Eq. (14)), it can be deduced that the five (5) most significant descriptors includes: ATS6e, AATSC6m, GATS2v, VR1_Dzv, and SpMax3_Bhv.

The plot of predicted pIC50 against experimental pIC50 values is displayed in Figs. 3 and 4, which shows close agreement between the predicted activity of the test set and that of the train set.

Table 6 gives a detail view of the numerical values of the train and the test sets as well as the respective predicted value which show minimal residual value thereby entailing good predictive strength of the model.

Table 7 provides the statistical internal validation parameters of the model obtained from the material studio program package.

The result of the Y-randomization test is shown in Table 8, indicating a robust model evidenced by its parameters.

The domain within which the models can predict the biological activity (pIC$_{50}$) and the absence of outlier as well as influential compound is depicted by Fig. 5.

Table 9 provides a detailed description of the descriptors in the model as well as their quality in terms of chance correlation and degree of contribution to the model.

3.2 Docking simulation studies

Table 10 summarizes the docking result presenting the binding scores, protease residue with interaction distance as well as interaction type. Figures 2a, b, 6a–d, and 7a,b showed prepared structure of the target (NS2B-NS3) and 3D structure of the prepared ligand 21, 2D interaction type for ligand 1, 2, 7, and 11 with different amino acids in the active site of protease and 2-D interaction type and H-bond molecular interaction between ligand 21 and the target respectively.
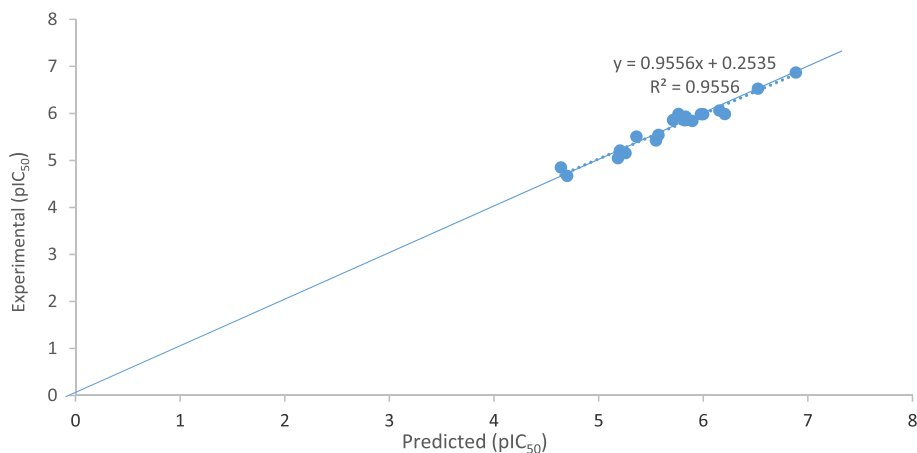

Fig. 3 Illustrating the plot of experimental pIC$_{50}$ and predicted pIC$_{50}$ value of train set chemical compounds of model M1
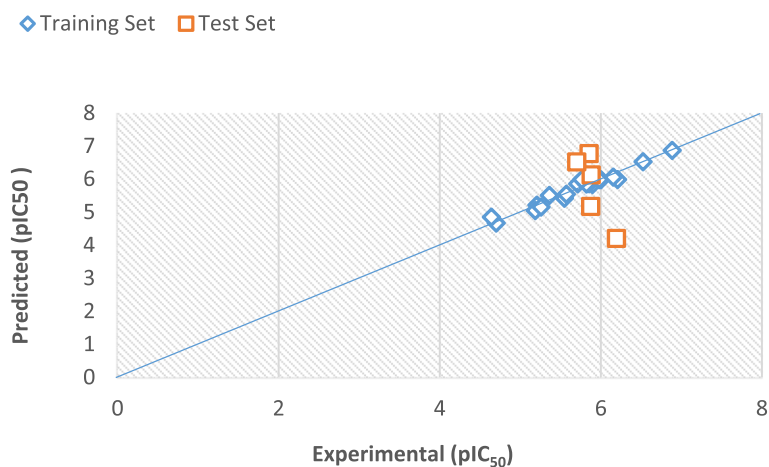
Fig. 4 The plot of experimental against predicted values of biological activities expressed as (pIC$_{50}$) of phthalazinone

Table 6 Experimental, predicted and residual values of phthalazinone derivatives

S/N	Experimental activity	Predicted activity	Residual
1*	6.19382	4.203771	1.990049
2	6.207608	5.984294	0.223314
3	5.548214	5.420270	0.127944
4	5.896196	5.836166	0.060030
5*	5.872895	5.171202	0.701693
6	4.699839	4.667067	0.032772
7	6.000000	5.981356	0.018644
8	5.978811	5.981260	− 0.002449
9	5.361511	5.503827	− 0.142316
10	5.853872	5.861122	− 0.007250
11	4.640354	4.848607	− 0.208252
12	5.205512	5.208835	− 0.003323
13	5.185087	5.044666	0.140421
14	5.571865	5.539619	0.032246
15	5.712198	5.856986	− 0.144788
16	6.154902	6.057303	0.097599
17	5.832683	5.927983	− 0.095300
18	5.821023	5.851128	− 0.030105
19	6.522879	6.522358	0.0005210
20	5.764472	5.987768	− 0.223296
21	6.886057	6.868233	0.017824
22*	5.879426	6.126705	0.247279
23*	5.853872	6.770595	0.916723
24	5.258061	5.152296	0.105765
25*	5.701147	6.521113	0.819966

Asterisk represents the test set

4 Discussion

4.1 GA-MLR model (QSAR)

The GFA model was successfully built from 20 train set compounds of 25 and 5 descriptors were contained in the model. The built model was subsequently used to predict the biological activity values for both the train and test reported in Table 6.

The multiple linear regression of genetic function algorithm (mlr-GFA) was used to produce three models; model (M1) was selected for its statistical significance as the best model with the following statistical parameters values (LOF = 0.088598, R^2_{adj} = 0.939699, R^2_{pred} = 0.71922, cR^2_p = 0.749517, Q^2_{CV} = 0.905909, and R^2 = 0.955567). Nevertheless, the statistical significance of this model is based on the suggested authentication standard as contained in Table 5. Though, based on the model parameters above, the model (M1) has satisfied

Table 7 Internal validation parameters of the genetic function approximation from the material Studio Program Package

S/N	Name	Value
1	Friedman LOF	0.088
2	R-squared	0.956
3	Adjusted R-squared	0.939
4	Cross validated R-squared Q^2_{CV}	0.906
5	Significant Regression	Yes
6	Significance-of-regression F value	60.217
7	Critical SOR F-value (95%)	2.978
8	Replicate points	0
9	Computed experimental error	0.000
10	Lack-of-fit points	14
11	Min expt. error for non-significant LOF (95%)	0.104

Table 8 Y-scrambling test parameters

Model Type	R	R^2	Q^{2LOO}
Original	0.977531	0.955567	0.905909
Random 1	0.639447	0.408892	0.167531
Random 2	0.741613	0.54999	− 0.19155
Random 3	0.830995	0.690552	0.461819
Random 4	0.476259	0.226823	− 0.51646
Random 5	0.383505	0.147076	− 0.81078
Random 6	0.606862	0.368281	− 0.25851
Random 7	0.693163	0.480475	− 0.63633
Random 8	0.658695	0.43388	− 0.06581
Random 9	0.447318	0.200094	− 1.19575
Random 10	0.585719	0.343067	− 0.66571
Random models parameters			
Average R	0.640101		
Average R^2	0.436791		
Average $Q^{2(LOO)}$	− 0.25506		
cRp^2	0.749517		

all the requirements for a satisfactory QSAR model. Having the above-stated validation values for satisfactory model values is an indication that the generated model has a good predictive capability. Five descriptors remained designated to construct the linear model, which was able to predict the corresponding pIC$_{50}$ values of all the selected compounds using the MLR-GFA statistical method.

The predicted pIC$_{50}$ values for the training and test sets were plotted against the experimental pIC$_{50}$ values as shown in Fig. 4. It is also noticeably from Fig. 3 that the calculated values for the pIC$_{50}$ were in a pact with

those of the experimental value, which entails the absence of error as observed in the model. A good correlation between experimental pIC$_{50}$ compared to the estimated pIC$_{50}$ of the compounds in the train set molecules was observed as demonstrated by Fig. 3, evidenced by the good correlation value (R^2 = 0.955) in Table 7 which is in agreement with the required validation threshold as suggested in Table 5 is an indication of the robustness of the built model [26, 34].

4.2 Y-randomization test
The outcome of the Y-scrambling test is depicted in Table 8 in which the values of R^2 and Q^2 are within the standard recommend statistical values. Also, the recommended value of greater than 0.5 was obtained for cR^2p which shows that the model has good predictive capacity.

4.3 Applicability domain
The applicability domain evaluation process as shown in Fig. 5 displays William's plot of the dataset, for which standardized residuals for both the train and test dataset were plotted against their respective leverage values identified no outlier for the compounds as all the data points were inside the limit of ± 3 domain. However, one outlier (compound 1) was observed to have exceeded the precautionary leverage of (h^* = 0.9). Furthermore, a close assessment revealed that it was not an actual outlier as disregarding this compound did not result in any perfection of the model statistical parameters and predictive strength and as such, it was retained. Furthermore, the other reason for not eliminating this compound from the test set was to avoid the use of excessively slight sort of endpoint values.
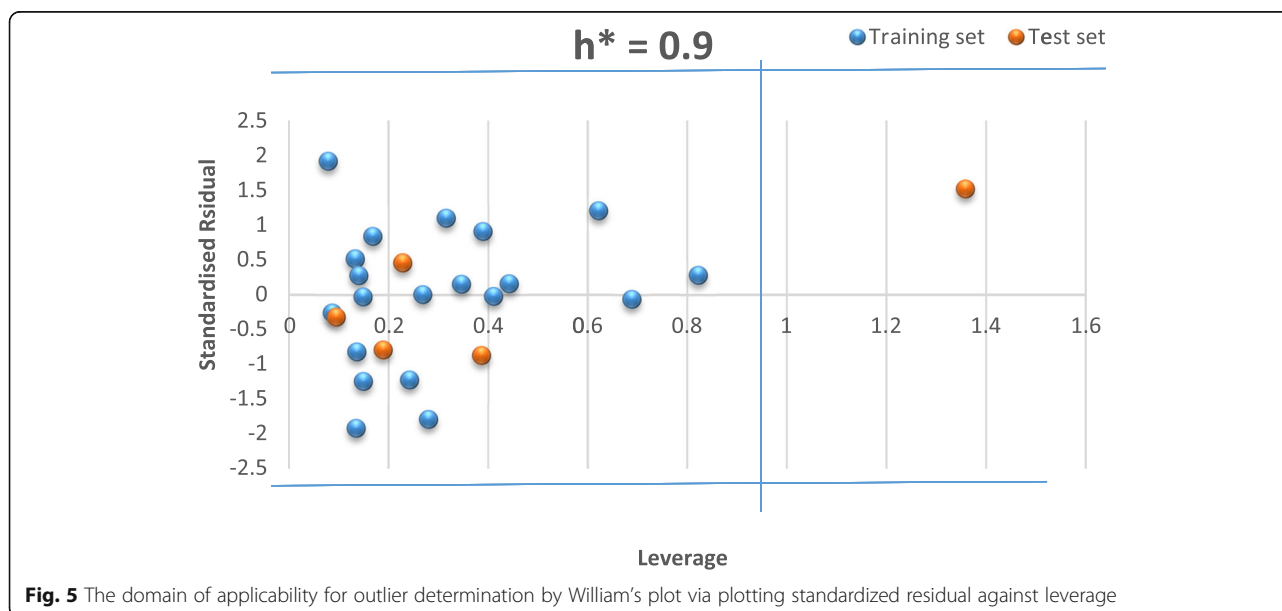

Fig. 5 The domain of applicability for outlier determination by William's plot via plotting standardized residual against leverage

Table 9 List of descriptors and their respective classes and statistical parameters

S/N	Descriptors	Description	Descriptor class	VIF	ME
1	ATS6e	Broto-Moreau autocorrelation—lag 6/weighted by Sanderson electronegativities.	2D	7.282	0.186
2	AATSC6m	Centered Broto-Moreau autocorrelation—lag 6/weighted by mass	2D	4.322	0.023
3	GATS2v	Geary autocorrelation—lag 2/weighted by van der Waals volumes.	2D	3.489	−0.313
4	VR1_Dzv	Randic-like eigenvector-based index from Barysz matrix/weighted by van der Waals volumes.	2D	5.468	−0.060
5	SpMax3_Bhv	Largest absolute eigenvalue of Burden modified matrix—n 3/weighted by relative van der Waals volumes.	2D	2.499	1.165

4.4 Variance inflation factor

Table 9 shows the list of descriptors, descriptions, classes, and other related statistical parameters (VIF) that possess a relevant influence on selected relevant descriptors. For all the five descriptors, the numerical values of the VIF were all less than 10 indicating that the specifications of the model were coronal, and the model's consistency is of great significant [19, 38–40].

4.5 Descriptors interpretation and mean effect

To have an insight into relevant factors responsible for the biological activity of the compound, there is a need for interpretation of the descriptors in the model and their respective individual relevant contribution in the model. The molecular descriptors in model M1 are ATS6e, AATSC6m, GATS2v, VR1_Dzv, and SpMax3_Bhv which had the following individual mean effect values (0.186, 0.023, − 0.313, − 0.060, and 1.165) respectively as obtained from Table 9.

ATS6e is the Broto-Moreau autocorrelation—lag 6/weighted by Sanderson electronegativities. It measures the strength of the relationship between relative electronegativity of two atoms in a molecule which are separated by 6 bonds; it has a positive correlation coefficient. Increment in its numerical value favors the increase in anti-dengue activity of the compounds. Also, these observations suggest that electronegativity of atoms that made up the compound had a substantial effect on the activity (pIC_{50}).

Also, AATSC6m is the centered Broto-Moreau autocorrelation—lag 6/weighted by mass. It measures the strength of the relationship between relative atomic mass of the atom pairs in a molecule separated by 6 bonds; it has a positive correlation coefficient. Therefore, increment in its numerical value would lead to increment in pIC_{50} as well. The BrotoMoreau autocorrelation descriptors (ATS) are given by

$$ATSdw = \sum_{i=1}^{n} . \sum_{j=1}^{n} . \delta_{ij} . w_{id} . w_{jd} \tag{15}$$

where n is the atom number, δ_{ij} is the Kronecker delta function (if dij = d, zero otherwise, then $\delta_{ij}=1$), d is the considered topological distance (the lag in the autocorrelation parameters), and wi and wj are the normalized atomic properties for atoms i and j respectively. The normalized van der Waals volume, atomic mass, and electronegativity can be appropriated for the atomic property.

GATS2v which is the Geary autocorrelation—lag 2/weighted by van der Waals volumes, which has a negative mean effect is suggested to contribute negatively to anti-dengue activity. It is evaluated or determined in the same way as the ATS but with the introduction of Geary coefficient; it measures the strength of the relationship between van der Waals volumes of two atoms in a molecule that are eight bond apart.

The Geary autocorrelation descriptors are given by

$$GATSdw \quad \frac{\frac{1}{2\Delta}\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_{ij}(w_i - w_j)^2}{\frac{1}{A-1}\sum_{i=1}^{n}(w_i - \hat{W}_j)^2} \tag{16}$$

where $\hat{W}$ represents the average coefficient of the considered property for the molecule and Δ is the number of vertex pairs from a distance equal to d.

Furthermore, VR1_Dzv is the Randic-like eigenvector-based index from the Barysz matrix/weighted by van der Waals volumes. It is negatively correlated to the anti-dengue activity meaning that decrease in its value enhances the activity of the compounds. They are based on the coefficients eigenvector associated with the largest negative eigenvalue of the distance matrix of a molecule.

However, SpMax3_Bhv is the largest absolute eigenvalue of Burden modified matrix—n 3/weighted by relative van der Waals volumes. It is obtained from modified connectivity matrix whose diagonal element is replaced by relative van der Waals volume of the atoms in the molecules. It also quantifies the topology of the chemical structure on the basis of connectivity of atoms present in the structure. This descriptor contributed the largest in determining the inhibitory activity which suggests its significance in the model as evidenced by its positive mean effect value.

Table 10 Interactions between compounds with a therapeutic target (DNV-2-NS-2B/NS-3 PDB 6M01)

Compound ID	Amino acid residue	Interaction distance (Å)	Type	(ΔG) Binding energy (Kcal/mol)
1 Vs. 6mo1	B:GLN64	2.21658	Conventional Hydrogen Bond	− 8.7
	A:LEU1018	2.33101	Conventional Hydrogen Bond	
	B:GLU66	3.30183	Halogen (Fluorine)	
	B:ALA1108	3.49375	Halogen (Fluorine)	
	B:GLU62	3.54375	Pi-Anion	
	B:ALA1108	2.70466	Pi-Donor Hydrogen Bond	
	B:ALA57	4.87127	Alkyl	
	B:LEU1098	5.29588	Alkyl	
	B:PRO1106	5.00654	Alkyl	
	B:ALA1108	5.03772	Alkyl	
	B:ALA65	5.43738	Pi-Alkyl	
	B:ARG1107	4.83196	Pi-Alkyl	
	B:ALA1108	4.58092	Pi-Alkyl	
2 Vs. 6mo1	A:TYR1023	2.80588	Conventional Hydrogen Bond	− 8.8
	B:GLN64	2.18891	Conventional Hydrogen Bond	
	B:GLU66	2.54142	Conventional Hydrogen Bond; Halogen	
	A:LEU1018	2.33372	Conventional Hydrogen Bond	
	B:ALA65	3.06887	Carbon Hydrogen Bond; Halogen	
	B:GLN64	3.29816	Halogen (Fluorine)	
	B:ALA1108	3.38615	Halogen (Fluorine)	
	B:ALA1108	2.58947	Pi-Donor Hydrogen Bond	
	B:ALA57	4.93795	Alkyl	
	B:LEU1098	4.7028	Alkyl	
	B:PRO1106	5.11146	Alkyl	
	B:ALA1108	4.61181	Alkyl	
	B:PRO1106	4.88773	Pi-Alkyl	
	B:ALA65	5.39978	Pi-Alkyl	
	B:ARG1107	4.82916	Pi-Alkyl	
	B:ALA1108	4.45163	Pi-Alkyl	
7 Vs. 6mo1	B:GLN64	2.31071	Conventional Hydrogen Bond	− 8.7
	B:GLU66	2.33594	Conventional Hydrogen Bond; Halogen	
	B:ALA65	3.05757	Carbon Hydrogen Bond; Halogen	
	A:LEU1018	3.593	Carbon Hydrogen Bond	
	B:GLN64	3.45412	Halogen (Fluorine)	
	B:ALA1108	3.18305	Halogen (Fluorine)	
	B:ALA1108	2.66277	Pi-Donor Hydrogen Bond	
	B:ALA57	4.78348	Alkyl	
	B:LEU1098	5.00203	Alkyl	
	B:PRO1106	4.65372	Alkyl	
	B:ALA1108	4.90607	Alkyl	
	B:PRO1106	5.46319	Pi-Alkyl	
	B:ALA65	5.33194	Pi-Alkyl	
	B:ARG1107	4.86108	Pi-Alkyl	
	B:ALA1108	4.50071	Pi-Alkyl	

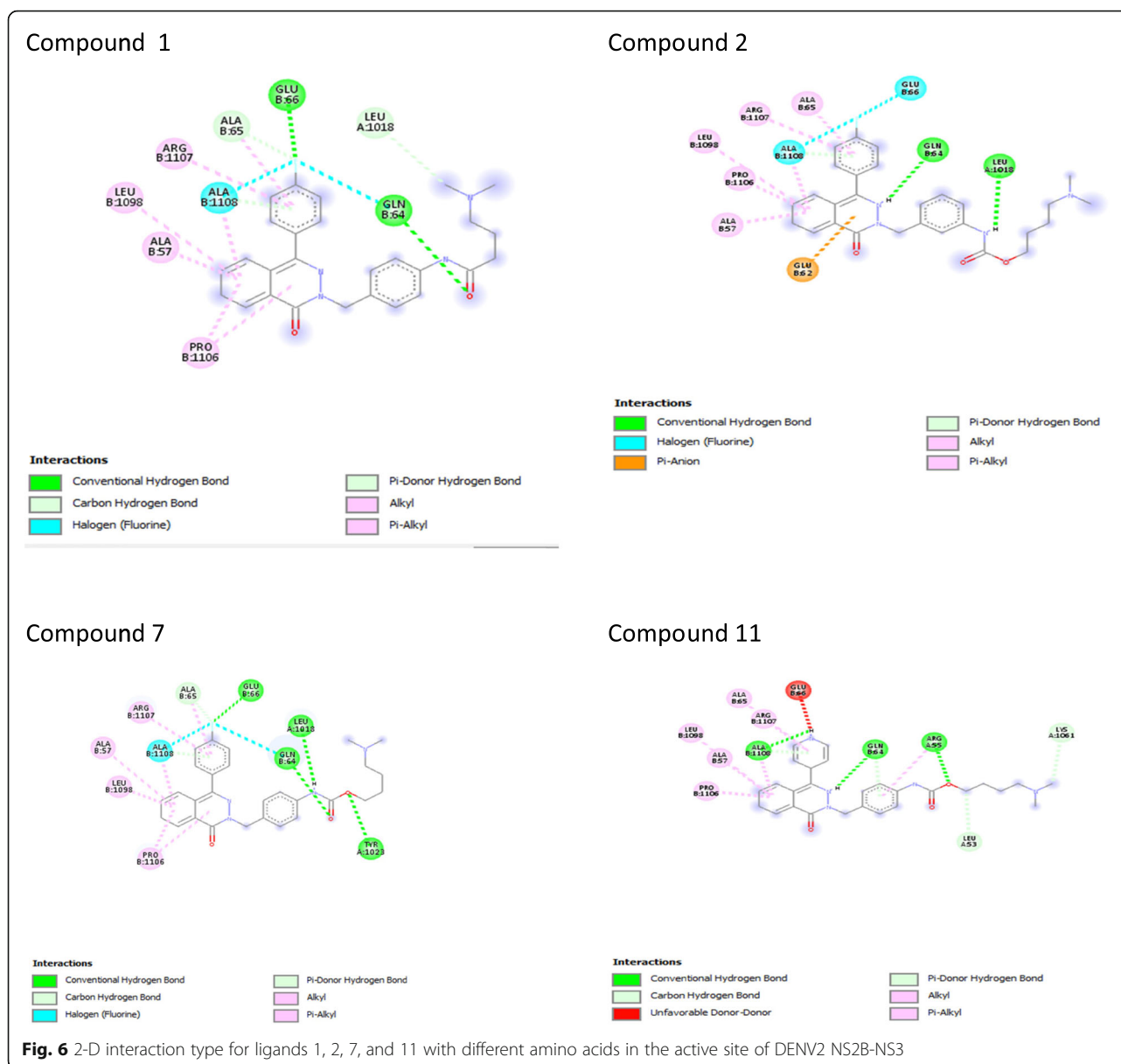**Table 10** Interactions between compounds with a therapeutic target (DNV-2-NS-2B/NS-3 PDB 6M01) *(Continued)*

Compound ID	Amino acid residue	Interaction distance (Å)	Type	(ΔG) Binding energy (Kcal/mol)
11 Vs. 6mo1	A:ARG55	2.1633	Conventional Hydrogen Bond	− 8.3
	B:GLN64	2.97358	Conventional Hydrogen Bond	
	B:ALA1108	2.81224	Conventional Hydrogen Bond	
	A:LEU53	3.48199	Carbon Hydrogen Bond	
	A:LYS1061	3.57148	Carbon Hydrogen Bond	
	B:GLN64	2.98809	Pi-Donor Hydrogen Bond	
	B:ALA1108	2.53115	Pi-Donor Hydrogen Bond	
	B:ALA57	4.80428	Alkyl	
	B:LEU1098	5.21523	Alkyl	
	B:PRO1106	4.54569	Alkyl	
	B:ALA1108	5.03321	Alkyl	
	A:ARG55	5.15853	Pi-Alkyl	
	B:ALA65	5.15773	Pi-Alkyl	
	B:ARG1107	4.75336	Pi-Alkyl	
	B:ALA1108	4.37424	Pi-Alkyl	
21 Vs. 6mo1	A:ARG55	2.63892	Conventional Hydrogen Bond; Halogen	− 8.9
	A:TYR1023	2.88122	Conventional Hydrogen Bond	
	B:GLN64	2.24277	Conventional Hydrogen Bond	
	A:LEU101	2.48849	Conventional Hydrogen Bond	
	A:ASP58	3.27412	Halogen (Fluorine)	
	B:GLU66	3.57383	Halogen (Fluorine)	
	B:GLU66	2.8329	Halogen (Fluorine)	
	B:ALA1108	2.72639	Pi-Donor Hydrogen Bond	
	B:ALA57	4.68629	Alkyl	
	B:LEU1098	4.67262	Alkyl	
	B:PRO1106	4.94631	Alkyl	
	B:ALA1108	4.82817	Alkyl	
	B:ARG1107	4.02028	Alkyl	
	B:PRO1106	4.87684	Pi-Alkyl	
	B:ALA65	5.23425	Pi-Alkyl	
	B:ARG1107	5.3005	Pi-Alkyl	
	B:ALA1108	4.30225	Pi-Alkyl	

The descriptors with positive mean effect value is an indication that an increase in the value of such descriptor will lead to an increase in the DENV-2 inhibitory activity (pIC_{50}) while a negative value indicates negative influence and as such, decrease in such value will enhance the activity (pIC_{50}) also.

4.6 The docking studies

In this study, the molecular docking studies of the phthalazinone with the NS2B-NS3 protease (PDB CODE: 6MO1) was investigated using AutoDock Vina of PyRx and Discovery Studio visualization software for energy grid calculations and visual analysis of the docking

pose respectively for a detailed understanding of the nature of the described interaction of inhibitors (compounds 1, 2, 7, 11, and 21) with the DNV-2 protease. The docking studies showed that these compounds docked well with the target and the binding affinity (− 8.7, − 8.8, − 8.7, − 8.3, and − 8.9 kcal/mol) of the 5 ligands under consideration with the target are all in close agreement with their respective pIC_{50} values (6.193, 6.21, 6.00, 4.64, and 6.886) for the targets respectively. Amino acids, numerical data of interaction distances, and binding free energies (ΔG) between the compounds, and NS2B-NS3 protease are shown in Table 10. However, the result showed that compound 21 had the best

Fig. 6 2-D interaction type for ligands 1, 2, 7, and 11 with different amino acids in the active site of DENV2 NS2B-NS3

binding interaction than the remaining 4. Compounds 1, 2, 7, and 21 were considered for the docking studies due to their high experimental activity (pIC_{50}) values while compound 11 was included to establish a basis for the variation in the observed biological activity with the ones with the best biological activity based on binding affinity due to its low pIC_{50}. Furthermore, it can be seen that in all the compounds docked, their binding energy corresponds with their inhibitory activity which shows that these compounds have great potentials. It can be seen in Table 10 that the compounds 1, 2, 7, 11, and 21 form conventional hydrogen bond with the following residue (GLN64, LEU1018: 2.22 Å, 2.33 Å), (TYR1023, GLN64, GLU66, LEU1018: 2.81 Å, 2.19 Å, 2.54 Å, 2.33 Å), (GLN64, GLU66: 2.31 Å, 2.34 Å), (ARG55, GLN64,

ALA1108: 2.16 Å, 2.97 Å, 2.81 Å), and (ARG55, TYR1023, GLN64, LEU101: 2.64 Å, 2.88 Å, 2.24 Å, 2.49 Å) respectively. From Table 10 also, it is likely to verify among the compounds that the increase in the amount of halogen bond interactions and the type would result in the lowering of binding free energy, which indicates a higher degree of the spontaneity of the interactions, which is also evidenced by the absence of halogen in compound 11 despite having two conventional hydrogen bond as (compound 1), the observed high binding affinity in compounds 1, 2, 7, and 21 could be attributed to fluorine. Also, compounds 2, 7, and 21 form key conventional hydrogen bond with the carbonyl of the carbamate core and the fluorine of the phenyl ring on the phthalazin core in which they both act as a hydrogen

Fig. 7 2-D interaction type (**a**) and H-bond molecular interaction (**b**) between Ligand 21 and the target
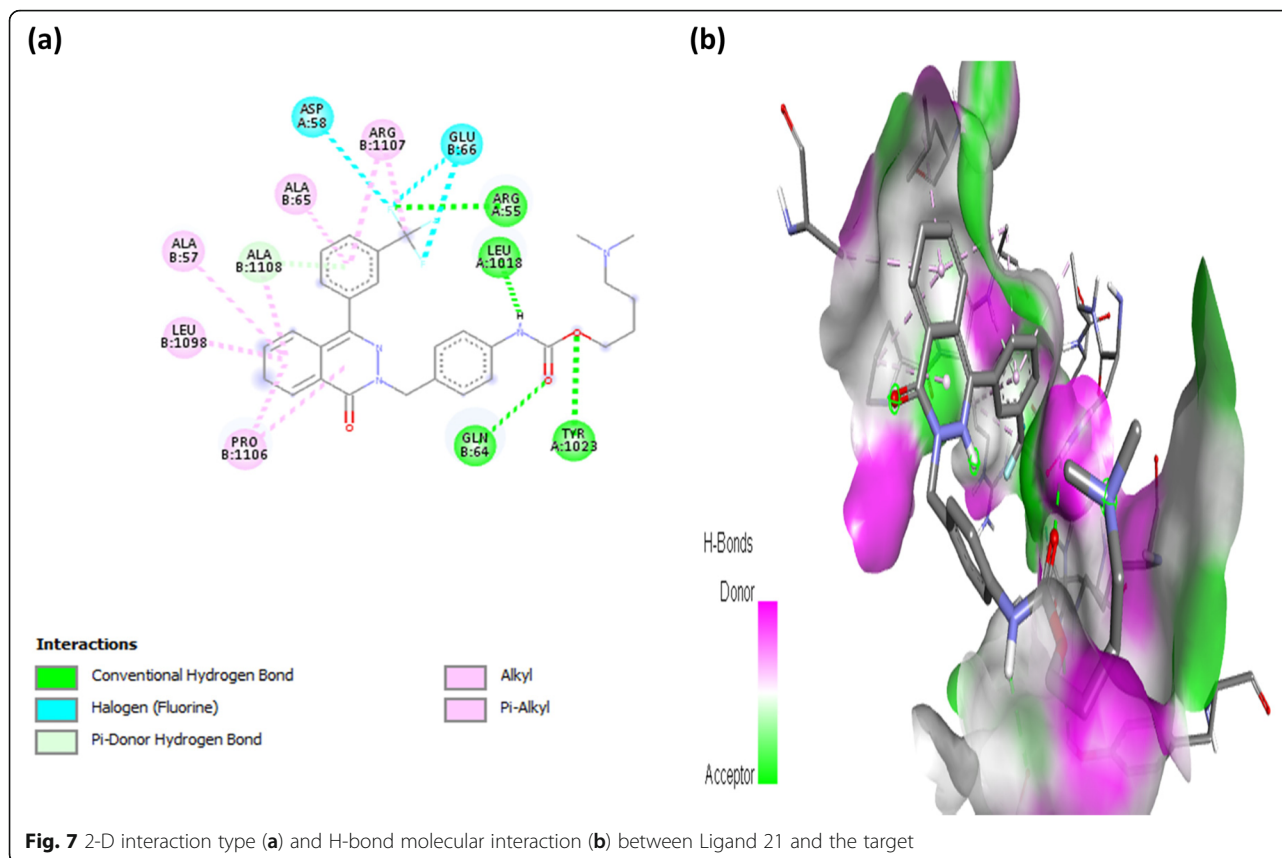
acceptor, and this entail relevance of such functional group. Figure 7 depicts the hydrogen bond interaction formed between the compound 21 and the target. The hydrogen bond interaction formed between ligand 21 with the highest binding affinity of − 8.9 kcal/mol suggests that the observed biological activity is not obtained by chance since it forms the most stable complex and as such, it can be utilized as a model compound for improving anti-dengue activity of the phthalazinone derivatives.

5 Conclusion

The present study targeted to produce a highly predictive MLR-GFA model capable of revealing the structural requirements for the experimental pIC_{50} of phthalazinone derivatives against dengue virus; the results from the acceptably validated model showed that the pIC_{50} of the studied molecules against dengue virus is determined by the descriptors ATS6e, AATSC6m, GATS2v, VR1_Dzv, and SpMax3_Bhv. The molecular docking simulation study reveals that among the studied compounds, compound 21 had the best binding energy (− 8.9 kcal/mol) and the binding energy of all the studied compounds correspond with their dengue virus inhibitory activity (pIC_{50}) and the most common interaction formed with the amino acid in all the studied compound

with the receptor are hydrogen and hydrophobic interactions; the presence of fluorine has a significant effect as observed in those with better activity (pIC_{50}). The information provided by the QSAR model may simplify further design of novel and highly potent dengue virus inhibitors. The studies also revealed that the compounds docked well with the targets suggesting that the ligands are efficacious in the treatment of DNV-2 infection.

Authors' contributions
SNA designed and wrote the manuscript; GAS, PAM, and AI supervised and carried out the statistical analysis. All authors read and approved the manuscript.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Pure and Applied Chemistry, Faculty of Science, University of Maiduguri, Maiduguri, Borno State P.M.B. 1069, Nigeria. [2]Department of Chemistry, Faculty of Physical Sciences, Ahmadu Bello University, Zaria, Kaduna State P.M.B. 1044, Nigeria.

References
1. Swain SS, Dudey D (2013) Anti-dengue medicinal plants: a mini-review. Res Rev J Pharmacogn Phytochem 1(2):5–9
2. Beatty ME, Stone A, Fitzsimons WD, Hanna JN, Lam SK, Vong S (2010) Best practices in dengue surveillance: a report from the Asia-Pacific and Americas Dengue Prevention Boards. PLoS Negl Trop Dis 4(11):890
3. Fatima Z, Idrees M, Bajwa MA, Tahir Z, Ullah O, Zia MQ (2011) 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(1):200
4. Guzman MG, Alvarez M, Halstead SB (2013) Secondary infection as a risk factor for dengue hemorrhagic fever/dengue shock syndrome: a historical perspective and role of antibody-dependent enhancement of infection. Arch Virol 158:1445–1459
5. Balasubramanian A, Teramoto T, Kulkarni AA, Bhattacharjee AK, Padmanabhan R (2017) Antiviral activities of selected antimalarials against dengue virus type 2 and Zika virus. Antiviral Research 137:141–150
6. Stephenson JR (2005) Understanding dengue pathogenesis: implications for vaccine design. Bull World Health Organization 83:308–314
7. Yao Y, Huo T, Lin Y, Nie S, Wu F, Hua Y, Wu J, Kneubehl AR, Vogt MB, Rico-Hesse R, Song Y (2019) Discovery, X-ray crystallography and antiviral activity of allosteric inhibitors of flavivirus NS2B-NS3 protease. J Am Chem Soc 141(17):6832–6836
8. Toepak EP, Tambunan USF (2017) In silico design of fragment-based drug targeting host processing α-glucosidase i for dengue fever. Mater Sci Eng. 172:01201
9. Nitsche C, Holloway S, Schirmeister T, Klein CD (2014) Biochemistry and medicinal chemistry of the dengue virus protease. Chem Rev 114(22): 11348–11381
10. Stevens AJ, Gahan ME, Mahalingam S, Keller PA (2009) The medicinal chemistry of dengue fever. J Med Chem 52(24):7911–7926
11. Lim SP, Wang QY, Noble CG, Chen YL, Dong H, Zou B, Yokokawa F, Nilar S, Smith P, Beer D, Lescar J, Shi PY (2013) Ten years of dengue drug discovery: Progress and prospects. Antiviral Res 100(2):500–519
12. Nguyen TTH, Lee S, Wang H-K, Chen H-Y, Wu Y-T, Lin SC, Kim D-W, Kim D (2013) In Vitro Evaluation of Novel Inhibitors against the NS2B and NS3 Protease of Dengue Fever Virus Type 4. Molecules 18:15600–15612
13. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Blanton RE, Silva LK, VG VGM (2008) Genetic ancestry and income are associated with dengue hemorrhagic fever in a highly admixed population. Eur J Hum Gen 16(6):762–765
14. Guzman A, Isturiz RE (2010) Update on the global spread of dengue. Int J Antimicrob Agents 36:40–42
15. De PM, Zanotto A, Gould EA, Gao GF, Harvey PH, Holmes EC (1996) Population dynamics of flaviviruses revealed by molecular phylogenies. Proc Natl Acad Sci USA 93(2):548–553
16. Vannice KS, Wilder-Smith A, Barrett ADT, Carrijo K, Cavaleri M, de Silva A, Durbin AP, Endy T, Harris E, Innis BL, Katzelnick LC, Smith PG, Sun W, Thomas SJ, Hombach J (2018) Clinical development and regulatory points for consideration for second-generation live-attenuated dengue vaccines. Vaccine 36:3411–3417
17. Joubert J, Foxen EB, Malan SF (2018) Microwave optimized synthesis of N-(adamantan-1- yl)-4-[(adamantan-1-yl)-sulfamoyl] benzamide and its derivatives for anti-dengue virus activity. Molecules 23(7):1678
18. Kadir SLA, Yakoob H, Zulkifli RM (2013) Potential anti-dengue medicinal plants: a review. J Nat Med 67:677–689
19. Roy K, Kar S, Das RN (2015) A primer on QSAR/QSPR modeling: fundamental concepts. Springer
20. Dastmalchi S, Hamzeh-Mivehroud M, Sokouti B (2018) Quantitative structure-activity relationship: a practical approach, vol 53. CRC Press
21. Lu D, Liu J, Zhang Y, Liu F, Zeng L, Peng R, Zuo J (2018) Discovery and optimization of phthalazinone derivatives as a new class of potent dengue virus inhibitors. Eur J Med Chem 145:328–337
22. Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, Consonni V (2014) QSAR modeling: where have you been? Where are you going to? J Med Chem 57(12):4977–5010
23. Dearden JC, Cronin MT, Kaiser KL (2009) How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). SAR QSAR Environ Res 20(3-4):241–266
24. Li Z, Wan H, Shi Y, Ouyang P (2004) Personal experience with four kinds of chemical structure drawing software: a review on ChemDraw, ChemWindow, ISIS/Draw, and ChemSketch. J Chem Info Comput Sci 44(5): 1886–1890
25. Hehre WJ, Huang WW (1995) Chemistry with Computation: An introduction to SPARTAN. Wavefunction, Inc
26. Tropsha A (2010) Best practices for QSAR model development, validation, and exploitation. Mol Inform 29(6–7):476–488
27. Friedman JH (1991) Multivariate adaptive regression splines. The annals of statistics.:1–67
28. Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, Agrawal RK (2011) Validation of QSAR models-strategies and importance. Int J Drug Des Discov. 3:511–519
29. Golbraikh A, Tropsha A (2002) Beware of q2! Journal of molecular graphics and modelling. 20(4):269–276
30. Larose S, Frédéric G, Michel B (2002) Attachment, social support, and loneliness in young adulthood: A test of two models. Personal Soc Psychol Bull. 28(5):684–693
31. Brandon K, Aline O (2015) Comprehensive R archive network (CRAN)
32. Roy PP, Leonard JT, Roy K (2008) Exploring the impact of the size of training sets for the development of predictive QSAR models. Chemometr Intell Lab Syst 90(1):31–42
33. Roy K, Ambure P (2016) The double cross-validation software tool for MLR QSAR model development. Chemometr Intelligent Lab Syst 159:108–126
34. Tropsha A, Gramatica P, Gombar VK (2003) Validation is the absolute essential for successful application and interpretation of QSPR models. QSAR Comb Sci 22:69–77
35. Dimitrov S, Dimitrova N, Parkerton T, Comber M, Bonnell M, Mekenyan O (2005) Base-line model for identifying the bioaccumulation potential of chemicals. SAR QSAR Environ Res 16(6):531–554
36. Weintrop D, Beheshti E, Horn M, Orton K, Jona K, Trouille L, Wilensky U (2016) Defining computational thinking for mathematics and science classrooms. J Sci Educ Technol 25(1):127–147
37. Olasupo SB, Uzairu A, Sagagis BS (2017) Density functional theory (B3LYP/6-31G*) study of the toxicity of polychlorinated dibenzofurans. Int J Comput Theor. Chem 5:4–24
38. Ameji JP, Uzairu A, Samuel H, Oluwaseye A, Samuel AN, Chinweuba OC (2015) Insilico prediction of octanol-air partition coefficient of some persistent organic pollutants through QSPR modelling. J Comput Methods Mol Design 5:46–60
39. Shapiro S, Guggenheim B (2008) Inhibition of oral bacteria by phenolic compounds. Part 1. QSAR analysis using molecular connectivity. Quantitative Structure Activity Relationships 17(04):327–337
40. Jaiswal M, Khadikar PV, Scozzafava A, Supuran CT (2004) Carbonic anhydrase inhibitors: the first QSAR study on inhibition of tumor-associated isoenzyme IX with aromatic and heterocyclic sulfonamides. Bioorganic Med Chem Letters 14(12):3283–3290

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.