

RESEARCH

Open Access



QSAR and molecular docking studies of novel 2,5-disubstituted-1,3,4-thiadiazole derivatives containing 5-phenyl-2-furan as fungicides against *Phytophthora infestans*

Yusuf Isyaku , Adamu Uzairu and Sani Uba

Abstract

Background: The 1,3,4-thiadiazoles are among the structural moieties that were found to be of utmost importance in the fields of pharmacy and agrochemicals because of their widespread biological activity that includes anti-tumor, antibacterial, anti-inflammatory, antihypertensive, anti-tuberculosis, anticonvulsant, and antimicrobial, among others.

Results: QSAR and molecular docking studies were carried out on thirty-two (32) derivatives of 2,5-disubstituted-1,3,4-thiadiazoles for their antifungal activities toward *Phytophthora infestans*. Using the “graphical user interface” of Spartan14 software, the structure of the compounds of the dataset is drawn and then optimized at DFT/B3LYP/6-31G* quantum mechanical method of the software. Molecular descriptors of the optimized compounds were calculated and later on divided into the training set and test sets (at a ratio of 3:1). The training set was used for model generation and the test set was for external validation of the generated model. Four models were generated by the employment of genetic function approximation (GFA) in which the optimal model (4) turned out to have the following statistical parameters: $R^2 = 0.798318$, $R^2_{adj} = 0.750864$, cross-validation $R^2(Q^2_{cv}) = 0.662654$, and external validation $R^2_{pred} = 0.624008$. On the molecular docking study of thiadiazole compounds with the target protein of *Phytophthora infestans* effector site (PDB ID: 2NAR), compound 13 shows the highest binding affinity with -9.3 kcal/mol docking score and composes hydrophobic as well as H-bond interactions with the target protein (2NAR).

Conclusion: The result of the QSAR study signifies the stability and robustness of the built model by considering the validation parameters and this gave an idea of template/ligand-based design while the molecular docking study revealed the binding interaction between the ligand and the protein site which gave an insight toward an “optimization method” of the structure-based design for the discovery of more potent compounds with better activity against *Phytophthora infestans* using the approach of computer-aided drug design (CADD) in plant pathology.

Keywords: QSAR, Docking, Fungicides, *P. infestans*

* Correspondence: isyakuyusuf01@gmail.com
Department of Chemistry, Ahamadu Bello University, Zaria, Nigeria

1 Background

Phytophthora infestans (also called potato blight) may be the most destructive of all plant pathogens that excessively damage potato/Irish potato leading to famines and immigration in the nineteenth century [13, 14]. Some of the signs and symptoms of this disease can be seen as a white color in potato. *P. infestans* generate some sporanges on the stem and leaves of potato [15]. The sporanges always displayed at the lower superficies of the leaves. However, as in the case of tuber blight, the white hypha usually appears at the superficies of the tuber [11]. In normal circumstances, *P. infestans* perfect its life on potato or tomato leaves at approximately 5 days [22]. The sporanges formed at the surface of foliage thereby dispersing through plants at over 10 °C (50 °F) temperature and humidity of above 75–80% in two or more days. Sometimes the spores are washed away by the rain which gets into the soil and infect the early-stage tubers; and also, these spores can make long distances into the air which can easily get into another host. The premature levels of the blight may have disappeared. Some of the symptoms involve dark blotches displayed at the extreme end of the leaf and on the plant's stem. A grey/dark patch developed on the affected tuber which covered the skin and rapidly decomposed it to an unpleasant odor. And apparently, healthy tubers may later become rotten while stored. According to the FOA report, the most thrusting/attacking problem in the third world apart from poverty must be food shortages. Farmers in Africa are encountering distinct limitations in food production as well as cash crops. Some of those limitations include damages from diseases and pests like fungi. In the search for food and the fight for human survival, the Irish potato has a significant role to play in food supply and, therefore, has been an instrument in addressing the issues of food insecurity, due to its performance in a given area and in a given time. This potato blight cause excessive economic loose, the annual economic loose caused by *P. infestans* in the developing countries begins to approach a \$3-billion mark [5]. Due to its rapid adaptation to the various management skills (such as genetic resistivity), control of this plant pathogen is really challenging [10]. And this makes the synthesis of novel compounds that will inhibit the dangerous *P. infestans* to be among the most considerable in the field of agrochemicals. Some of these researches include computational studies.

The 1,3,4-thiadiazoles derivatives are among the structural moieties that were found to be of utmost importance in the fields of pharmacy and agrochemicals for their widespread biological activity such as anti-tumor [28], antibacterial [25], anti-inflammatory [19], antihypertensive [30], antituberculosis [23], anticonvulsant [18], and antimicrobial [2], among others. Furthermore,

reports identify that compounds containing furan are intensively bioactive. Several researches on the derivatives of furan such as “pyrazole and triazole [6], diacylhydrazine derivatives [7]” containing 5-phenyl-2-furan moiety were carried out in which there appeared to have extensive biological activities including fungicidal and insecticidal activities, among others.

The Quantitative structure–activity relationship (QSAR) study aims to develop correlation models considering the activity of compounds and other chemical information in a statistical approach [16, 27] which will lead us to the design of new compounds. While molecular docking study is “a way of predicting the favorable orientation of one molecule to another when reacted to produce a stable complex”, it will also lead us to the design of more potent compounds.

Our aim in this research work is to predict highly active compounds by the employment of Genetic function approximation (GFA) and perform a molecular docking study between the 1,3,4-thiadiazole compounds and the 2NAR protein of *P. infestans* to predict their stable molecular orientation.

2 Methods

2.1 Dataset

Thirty-two derivatives of 2,5-disubstituted-1,3,4-thiadiazole derivatives containing 5-phenyl-2-furan used in this work were taken from the literature [8]. The activity of the compounds were reported in EC₅₀ (g/L) values, which were converted to pEC₅₀ (pEC₅₀ = -log₁₀/EC₅₀). Presented in Fig. 1 and Table 1 are the molecular structures and their corresponding activities found in the dataset.

2.2 Molecular structure optimization

The structures of the compounds were optimized at the “Density function theory (DFT)” level, “Becke's three-parameter Lee-Yang-Parr hybrid functional (B3LYP)” version together with the “6-31G*” basis set of Spartan14 [4]. In this process, all the molecular structures were drawn in the graphical user interface of Spartan14 software. The energies of the drawn molecules were minimized using Molecular Mechanics Force Field (MMFF) calculation [3].

2.3 Molecular descriptor calculations

Molecular descriptors are the properties of the molecule in numerical/mathematical values. PaDEL descriptor software was used to further calculate additional energy of those low-energy conformers, where a total of 1875 descriptors were calculated.

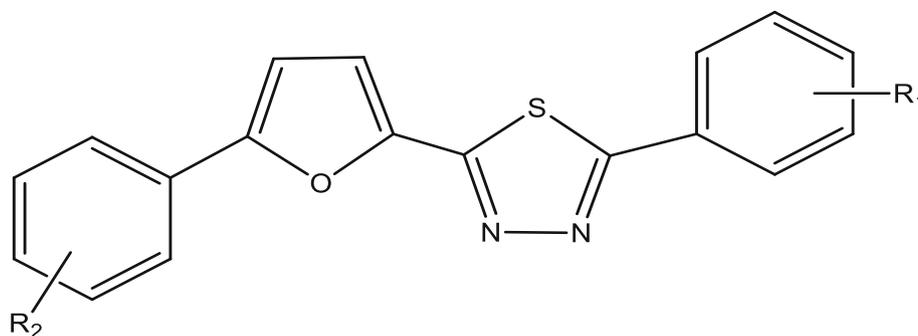


Fig. 1 Parent structure of the dataset compounds

Table 1 Compounds and their pEC₅₀ values

Compound	R ₁	R ₂	pEC ₅₀
1	4-OCH ₃	4-NO ₂	1.344392
2	4-Br	2-Cl	1.786751
3	4-Cl	4-OCH ₃	1.71433
4	2-OCH ₃	4-Cl	1.893207
5	3-Cl	4-Cl	1.910624
6	4-OCH ₃	4-Br	1.962369
7	4-Cl	2-Cl	2.007748
8	2-Cl	2-Cl	1.895423
9	4-OEt	2-Cl	1.840106
10	H	2-Cl	0.869232
11	4-Cl	2,4-di-F	1.501059
12	2-Cl	2,4-di-F	1.926857
13	4-CH ₃	2,4-di-F	1.873902
14	4-OCH ₃	2,4-di-F	1.841359
15	3-CH ₃	4-F	1.506505
16	2-Cl	3-F	1.655138
17	4-CH ₃	4-F	2.051153
18	H	4-F	0.755875
19	4-Cl	H	0.612784
20	2-Cl	H	1.826723
21	3-CH ₃	4-OCH ₃	1.752048
22	4-Cl	4-CH ₃	2.047275
23	4-OCH ₃	2-F	1.870404
24	3-CH ₃	2-F	1.774517
25	H	2-F	0.924279
26	4-CH ₃	4-Cl	1.70927
27	4-OCH ₃	2,6-di-F	1.751279
28	4-Cl	2,6-di-F	1.49693
29	2-Cl	2-NO ₂	1.745855
30	4-CH ₃	2-NO ₂	1.824126
31	4-OCH ₃	3-NO ₂	1.257679
32	H	2,6-di-F	1.74351

2.4 Dataset splitting

Using Kennard–Stone algorithm technics, the dataset of 32 compounds was split into two: the training set and the test set (70% of the training and 30% to the test set) which is found in DatasetDivision GUI 1.2 software. In this technic, the training and the test set were used for model development and its validation (externally) [12].

2.5 Model building

The training set in the dataset was used for model generation through the employment of the GFA method available in the material studio. The regression analysis occurs by considering the inhibition concentration (pEC₅₀) as a dependent variable while the chosen descriptors served as independent variables.

2.6 Internal validation

Internal validation of 22 compounds of the training set took place in the software (Material studio) used for building the model. The validation parameters are as follows:

2.6.1 Cross-validation

This parameter was used to determine the ability of the QSAR model in predicting the activities of newly designed compounds. This indicates the stability of the built model.

$$Q_{cv}^2 = 1 - \frac{\sum (Y_{pred} - Y_{exp})^2}{\sum (Y_{exp} - \bar{Y})^2} \quad (i)$$

where Y_{exp} is the “observed/experimental activity”, Y_{pred} is the “predicted activity”, and \bar{Y} is the “mean value of the observed activity”.

2.6.2 Friedman’s lack of fit (LOF)

The parameter describes the measure of the fitness of the model and it is given by equation. ii below:

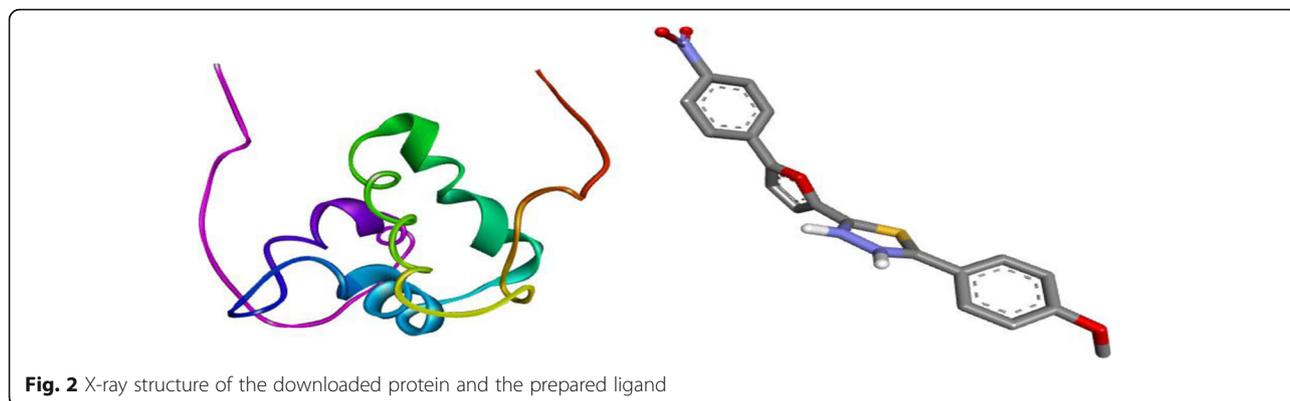


Fig. 2 X-ray structure of the downloaded protein and the prepared ligand

$$\text{LOF} = \frac{\text{SEE}}{\left(1 - \frac{C+dP}{M}\right)^2} \quad (\text{ii})$$

where SEE is the standard error,

$$\text{SEE} = \sqrt{\frac{(Y_{\text{exp}} - Y_{\text{pred}})^2}{N - P - 1}} \quad (\text{iii})$$

C is the “number of terms in the model”, d is the “user-defined smoothing parameter”, P is “the total number of descriptors in the model”, and M is “the number of molecules in the training set”.

The regression model is given by the straight line graphs’ equation, “ $(Y = mx + c)$ ”,

$$Y = D_1x_1 + D_2x_2 + D_3x_3 \dots + D_nx_n + c \quad (\text{iv})$$

where Y is the predicted activity (pEC_{50}), D is the corresponding coefficients, x is the independent variable, and c is the regression constant [17].

2.6.3 The correlation coefficient (R^2)

This is another parameter used to assess the model. The closer the value of R^2 to 1.0, the better the model generated. R^2 is expressed as:

Table 2 Validation parameters of the model 4

Friedman LOF	0.12717000
R^2	0.79831800
Adjusted R^2	0.75086400
Cross validated R^2	0.66265400
Significant regression	Yes
Significance of regression F -value	16.82278600
Critical SOR F -value (95%)	3.01408800
Replicate points	0
Computed experimental error	0.00000000
Lack-of-fit points	17
Min. expt. error for non-significant LOF (95%)	0.13201000

$$R^2 = 1 - \frac{\sum (Y_{\text{exp}} - Y_{\text{pred}})^2}{\sum (Y_{\text{exp}} - \bar{Y}_{\text{train}})^2} \quad (\text{v})$$

The value of R^2 changes instantly with an increase in descriptors; therefore, the reliability of R^2 in measuring the stability of a given model is very minimal. Thus, R^2 has to be adjusted in order to have a fit and strong model. The following equation define adjusted R^2 as [1]:

$$R^2_{\text{adj}} = (1 - R^2) \frac{(n-1)}{n-P-1} = \frac{(n-1)(R^2 - P)}{n-P+1} \quad (\text{vi})$$

where P is the number of independent variables possessed by the model and n is the number of training sets’ compounds [21].

2.7 External validation

The model generated was further validated with the test set of the dataset in order to measure its level of competence in predicting the activity of new compounds. This was done by evaluating the values of the square of the calculated R^2 of the test set values. The closer the R^2 is to 1.0, the better the robustness, fitness, and the prediction capacity of the model as well. Though sometimes R^2 value does not matter if the model fails other statistical analyses such as variance inflation factor (VIF) and mean effect, among others. The coefficient of determination R^2_{pred} is given by the following equation:

Table 3 Minimum recommended values of validated parameters for generally acceptable QSAR

Symbol	Name	Value
R^2	Coefficient of determination	≥ 0.6
$P_{95\%}$	Confidence interval at 95% confidence level	< 0.05
Q^2	Cross validation coefficient	< 0.5
R^2_{ext}	Coefficient of determination for external test set	≥ 0.5
VIF	Variance inflation factor	$1 \leq \text{VIF} \leq 10$

Table 4 External validation

Name	pEC50	AMR	SCH-7	Wlambda3.mass	Wnu1.polar	Ypred
9	1.840106	116.2273	0.693123	1.709892	0.939062	1.949358
16	1.655138	105.4178	0.71753	1.980691	0.907522	1.65767
17	2.051153	104.9836	0.726359	23.74621	0.928937	2.166915
19	0.612784	105.29	0.653839	0	0	1.00147
25	0.924279	100.7016	0.64501	0	0	0.786564
24	1.774517	104.9836	0.71753	0.004722	0.904366	1.597157
30	1.824126	112.2107	0.68231	0	0	1.399054
29	1.745855	112.6449	0.673481	0	0	1.372555
31	1.257679	114.5722	0.666731	0	0	1.412139
32	1.74351	100.8294	0.680102	0.023018	0.917819	1.264777

$$R^2 = 1 - \frac{\sum (Y_{\text{pred}_{\text{test}}} - Y_{\text{exp}_{\text{test}}})^2}{\sum (Y_{\text{exp}_{\text{test}}} - \bar{Y}_{\text{train}})^2} \quad (\text{vii})$$

where $Y_{\text{pred}_{\text{test}}}$ and $Y_{\text{exp}_{\text{test}}}$ are the values of predicted and experimental activities for the test set and \bar{Y}_{train} is the average activity for the training sets' values [3].

2.8 Statistical analysis of the descriptors

2.8.1 Variance inflation factor (VIF)

VIF is defined as the measure of multicollinearity amongst the independent variables (i.e., descriptors). It quantifies the extent of correlation between one predictor and the other predictors in a model.

$$\text{VIF} = \frac{1}{(1-R^2)} \quad (\text{viii})$$

where R^2 gives multiple correlation coefficient between the variables within the model. If the VIF is equal to 1, it means there is no intercorrelation in each variable, and if it ranges from 1 to 5, then it is said to be suitable and

acceptable. But if the VIF turns out to be greater than 10, this indicates the instability of the model and need to be reexamined ([20, 26].

2.8.2 Mean effect (ME)

The average effect (mean effect) correlates the effect or influence of given molecular descriptors to the activities of the compounds that made up the model. The descriptor signs show the direction of their deviation toward the activity of the compounds. That is to say, an increase or decrease in the value of the descriptors will improve the activity of the compounds. The mean effect is defined by the following:

$$\text{Mean effect} = \frac{B_j \sum_i^n D_j}{\sum_j^m (B_j \sum_i^n D_j)} \quad (\text{ix})$$

where B_j and D_j are the j -descriptor coefficient in the model and the values of each descriptor in training set, while m and n stand for the number of molecular descriptors as well as the number of molecules in a training set. To evaluate the significance of the model, the mean effect of each descriptor was calculated [9].

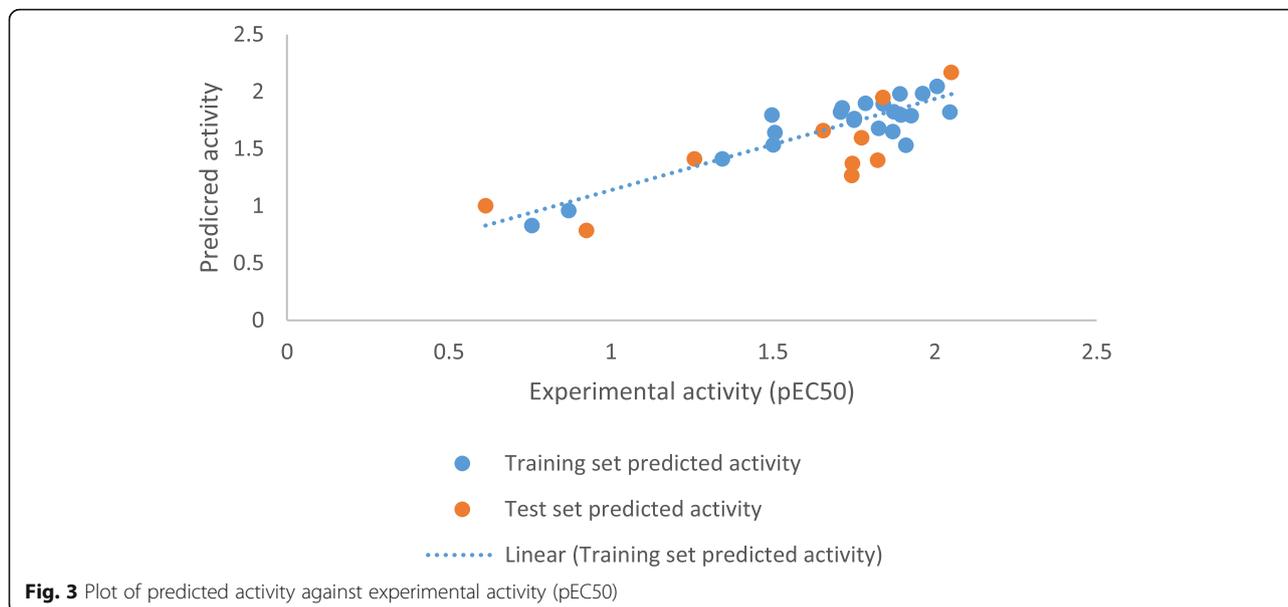
Table 5 Calculations of predicted R^2

$(Y_{\text{pred}} - Y_{\text{obs}})$	$(Y_{\text{pred}} - Y_{\text{obs}})^2$	$Y_{\text{meanTrain}}$	$Y_{\text{obs}} - Y_{\text{meanTr}}$	$(Y_{\text{pred}} - Y_{\text{meanTrain}})^2$
0.109252	0.011936	1.692889	0.147217	0.021673
0.002532	6.41E-06	1.692889	- 0.03775	0.001425
0.115762	0.013401	1.692889	0.358263	0.128353
0.388686	0.151077	1.692889	- 1.08011	1.166627
-0.13771	0.018965	1.692889	- 0.76861	0.590761
-0.17736	0.031456	1.692889	0.081628	0.006663
-0.42507	0.180686	1.692889	0.131237	0.017223
-0.3733	0.139353	1.692889	0.052966	0.002805
0.15446	0.023858	1.692889	-0.43521	0.189408
-0.47873	0.229185	1.692889	0.050621	0.002562

$$\sum (Y_{\text{pred}} - Y_{\text{obs}})^2 = 0.799924$$

$$\sum (Y_{\text{pred}} - Y_{\text{MeanTrain}})^2 = 2.127501$$

$$R^2_{\text{pred}} = (1 - 0.799924/2.127501) = 0.624008$$



2.8.3 Applicability domain

To confirm the reliability of the model and to examine the outliers as well as the influential compounds, it is very important to evaluate the applicability domain of the built model. Its aim at predicting the uncertainty of a compound depends on its similarities to the compounds used in building the model and also the distance between the training and test set of the compounds. This can be achieved by employing William's plot which was plotted using standardized residuals versus the leverages. The leverages for a particular chemical compound is given as follows:

$$h_i = Z_i(Z^T \cdot Z)^{-1} Z_i^T \quad (\text{x})$$

where h_i is the leverage for a particular compound and Z_i is the matrix i of training set. Z is the $n \times k$ descriptor-matrix for a training set compound. Z^T is the transpose of the Z matrix. The warning leverage (h^*) that is the boundary for normal values of Z outliers is given by;

$$h^* = 3 \frac{(p+1)}{n} \quad (\text{xi})$$

Where n is the number of molecules in the training set whereas p gives the amount of descriptors presence in the built model [17].

2.9 Molecular docking studies

With the aid of Autodock Vina of Pyrex software and Discovery Studio, a molecular docking study was performed between 2,5-disubstituted-1,3,4-thiadiazole derivatives and *P. infestans* effector target site to examine the interaction between the binding pocket of the effector and the compounds (i.e., the ligands). A highly

resolute crystal structure of *P. infestans* was downloaded successfully from the protein databank (PDB Code: 2NAR). The downloaded substrate was carefully prepared using Discovery Studio which was later transported to the Pyrex for the docking calculation. With the aid of Spartan14 version 1.1.4, the optimized compounds of 2,5-disubstituted-1,3,4-thiadiazole derivatives (the ligands) were converted to PDB files [24]. The prepared structure of *P. infestans* effector site and prepared ligands were docked using Autodock Vina 4.2 [29]. Discovery Studio Visualizer was also used to visualize the docking results (Fig. 2).

3 Results

3.1 Model building and validation

Below is the equation of the best-chosen model (4).

$$\begin{aligned} \text{pEC}_{50} = Y = & 0.037511826 \cdot \text{AMR} \\ & + 4.846246933 \cdot \text{SCH-7} \\ & + 0.021854712 \cdot \text{WG.unity} \\ & + 0.3299691 \cdot \text{Wnu2.eneg} - 6.116815304 \end{aligned} \quad (\text{xii})$$

The validation parameters are shown in Tables 2 and 3 below.

3.1.1 Model 1

In the first model, $\text{pEC}_{50} = 0.244535617 \times \text{BCUTp-11} - 22.874691031 \times \text{SCH-6} + 0.213428935 \times \text{WA.mass} - 0.025525444 \times \text{Wgamma3.volume} + 10.325883792$, $R^2_{\text{teat}} = 0.395084$, $R^2_{\text{train}} = 0.824826$, $R^2_{\text{adj}} = 0.783609$, $R^2_{\text{cv}} = 0.56979$, $N_{\text{test}} = 10$, $N_{\text{train}} = 22$, $\text{LOF} =$

Table 6 Experimental activity, predictive activity, and residual values of the dataset compounds

Name	pEC ₅₀	Predicted (pEC ₅₀)	Residual values
1	1.344392	1.412139	- 0.067750
2	1.786751	1.896776	- 0.110030
3	1.71433	1.857390	- 0.143060
4	1.893207	1.979555	- 0.086350
5	1.910624	1.529835	0.380790
6	1.962369	1.984080	- 0.021710
7	2.007748	2.044664	- 0.036920
8	1.895423	1.795907	0.099515
9	1.840106	1.949358	1.949358
10	0.869232	0.958705	- 0.089470
11	1.501059	1.532573	- 0.031510
12	1.926857	1.788722	0.138135
13	1.873902	1.822533	0.051369
14	1.841359	1.891726	- 0.050370
15	1.506505	1.640308	- 0.133800
16	1.655138	1.657670	1.657670
17	2.051153	2.166915	2.166915
18	0.755875	0.829351	- 0.073480
19	0.612784	1.001470	1.001470
20	1.826723	1.679846	0.146876
21	1.752048	1.762747	- 0.010700
22	2.047275	1.819474	0.227801
23	1.870404	1.649906	0.220498
24	1.774517	1.597157	1.597157
25	0.924279	0.786564	0.786564
26	1.709270	1.820588	- 0.111320
27	1.751279	1.750914	0.000365
28	1.496930	1.795823	- 0.298890
29	1.745855	1.372555	1.372555
30	1.824126	1.399054	1.399054
31	1.257679	1.412139	1.412139
32	1.743510	1.264777	1.264777

Table 7 Pearson's correlation

	AMR	SCH-7	WG.unity	Wnu2.eneg	VIF
AMR	1				1.032098
SCH-7	- 0.04901	1			1.15397
WG.unity	0.127617	- 0.26704	1		1.172952
Wnu2.eneg	0.067316	0.316544	- 0.31218	1	1.202662

0.12303, and m in experimental error for non-significant LOF (95%) = 0.12303.

3.1.2 Model 2

In the second model, $pEC_{50} = 0.297814107 \times nCl + 0.168441873 \times nBondsS3 + 0.001197233 \times PPSA-1 - 0.024107696 \times Wgamma3.volume + 0.268877261$, $R^2_{teat} = 0.206664$, $R^2_{train} = 0.807874$, $R^2_{adj} = 0.762668$, $R^2_{cv} = 0.558932$, $N_{test} = 10$, $N_{train} = 22$, LOF = 0.12885, and min experimental error for non-significant LOF (95%) = 0.128845.

3.1.3 Model 3

In the third model, $pEC_{50} = 0.139831691 \times nHeavyAtom + 0.314911162 \times nCl + 0.001443139 \times PPSA-1 - 0.024455939 \times Wgamma3.volume - 2.325579534$, $R^2_{teat} = 0.3681753$, $R^2_{train} = 0.800593$, $R^2_{adj} = 0.753674$, $R^2_{cv} = 0.516043$, $N_{test} = 10$, $N_{train} = 22$, LOF = 0.13126, and min experimental error for non-significant LOF (95%) = 0.131264.

Table 4 and Table 5 presented the external validation and calculation of predicted R^2 of the chosen model.

3.2 Statistical analyses of the descriptors

The following are the different analyses: Pearson's correlation, standard regression coefficients, standardized predicted activity against experimental activity, standardized residual against experimental activity (pEC50), and William's plot.

3.3 The results of the docking study

The results can be seen in the receptor–ligand interaction, H-bond interactions, and hydrophobic and electrostatic interactions.

4 Discussion

4.1 QSAR model

The best QSAR model was generated using the GFA method. Four descriptors were used in building the model where four different models were generated and model 4 found to be the best following the statistical parameters. All the values obtained matches the minimum value for evaluating the QSAR model. These values signify that there is a high correlation between the predicted and experimental activity (pEC50, Fig. 3). Internal and external validations, as well as the other statistical analysis, made the model 4 to stand fit, reliable, and highly predictive.

From Tables 2 and 4, the R^2 values of 0.79831800 (internal) and 0.624008 (external) indicate a strong relationship between the experimental and predicted activities. Additionally, the inhibition activities of the compounds increase by the addition of all the descriptors in the best chosen model.

Table 8 The standard regression coefficients “ b_j ”, the values of mean effect (MF), and confidence interval (p values)

Descriptors	Standard regression coefficient (b_j)	Mean effect (MF)	p value (confidence interval)
AMR	0.037512	0.521147	0.002152
SCH-7	4.846247	0.441298	0.000518
WG.unity	0.021855	0.010013	0.000368
Wnu2.eneg	0.329969	0.027542	0.001551

4.2 Interpretation of descriptors

The 2D molecular descriptors *AMR* and *SCH-7* defined as “Molar refractivity” and “Simple chain, order 7” are the first and second highest contributors toward the generation of the selected model with a positive mean effect of 0.52115 and 0.4413. Thus, the addition of those descriptors will significantly enhance the antifungal activity of the compound. 3D descriptors *WG.unity* and *Wnu2.eneg* with the mean effect of 0.01001 and 0.02754 have a low effect on the model therefore their increase will have no much significant on the activity of the compound. They are defined as “Non-directional WHIM, weighted by unit weights” and “Directional WHIM, weighted by Mulliken atomic electronegativities”.

Model 4 was examined as the optimal model considering the descriptors from test set compounds of the dataset.

The experimental activity, predictive activity, and residual values of the compounds are given in Table 6. The residual value is defined as the difference between the experimental and predicted activities. The lower residual values between the experimental and predicted activities indicate the high predictive power of the model.

4.3 Statistical analysis of descriptors

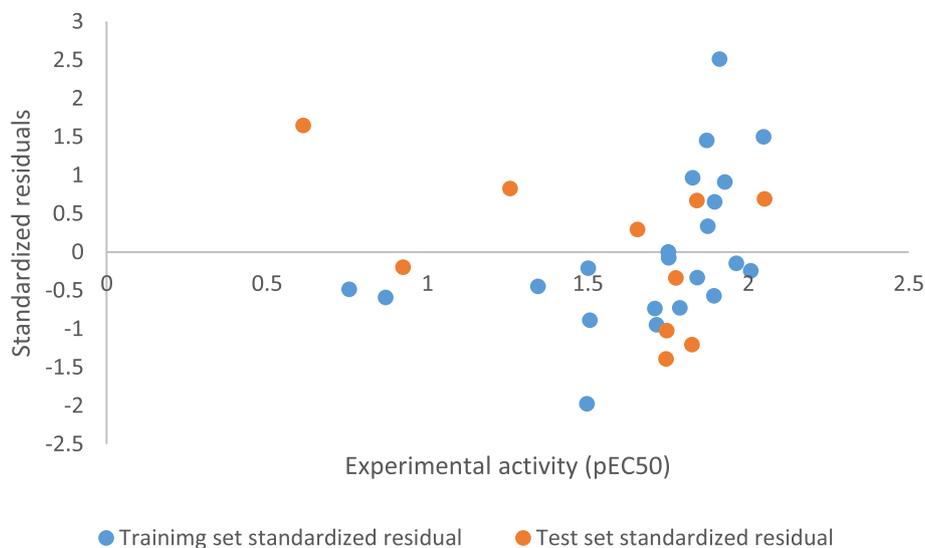
Pearson's correlation (Table 7) was performed between the descriptors of the chosen model in order to evaluate the

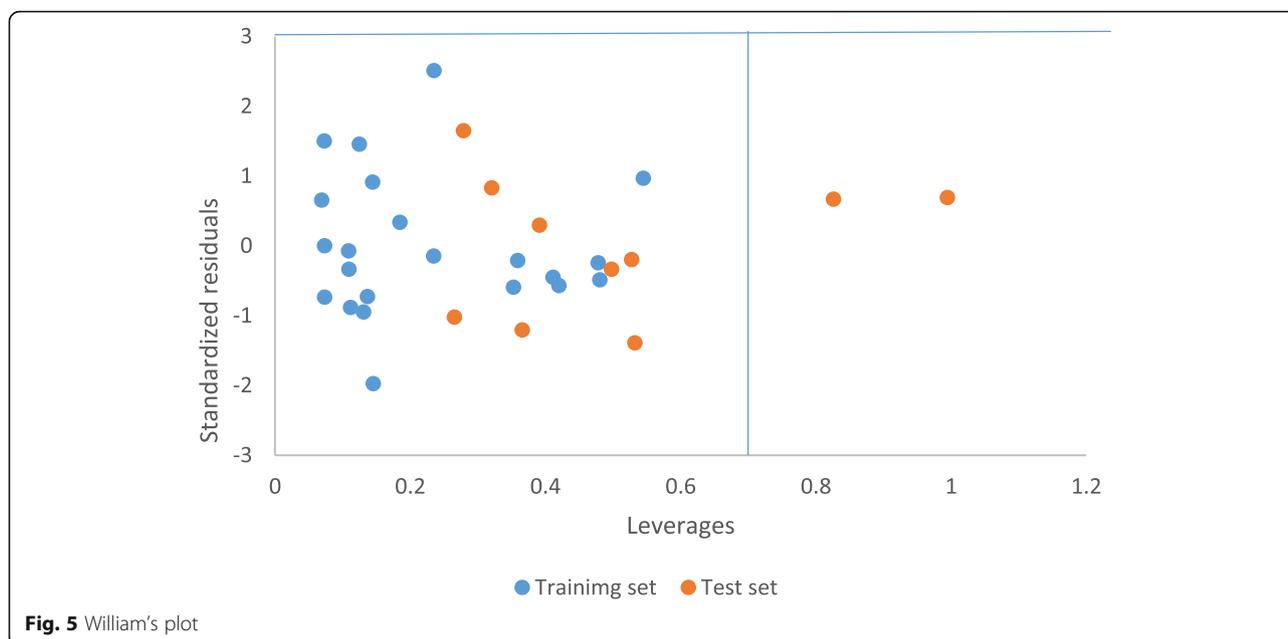
relationship between each of the descriptors. The result of the correlation showed no intercorrelation among the descriptors with a correlation coefficient of less than 0.5, which signified that the descriptors used in the model were good enough. The VIF values are within the range of 1 to 5 which indicated that the descriptors and model are suitable and acceptable.

Table 8 showed the standard regression coefficients “ b_j ”, the values of mean effect (MF), and confidence interval (p values). These give vital information on the effect and contribution of the descriptors toward the built model. The individual capability and inducing power of the selected descriptors toward the activity of the compounds depend on their values, signs, and their mean effects as well. The p values of the four descriptors (at a confidence limit of 95%) that made up the model are all less than 0.05; this implies that there is a significant relationship among the descriptors (as contrary to the null hypothesis) and the inhibitory concentration of the compounds.

Figure 4 which presented a graph of observed activity versus standardized residual shows a random dispersion at the baseline where the standardized residual is zero. Therefore, no systematic error occurred in the built model.

The graph of standardized residuals versus leverages (for all the training set and test set) termed the William's

**Fig. 4** Plot of standardized residual against experimental activity (pEC50)



plot shown in Fig. 5. The domain of applicability is established within a box at ± 3.0 limit for the residuals and a leverage threshold h^* ($h^* = 0.68$). This William's plot functions to figure out the outliers as well as the influencing compounds in the model. Our results revealed that two compounds of the test set (with pEC_{50} of 1.84011 and 2.05115) were outside the applicability domain which signified that the compound may be structurally different from other compounds in the dataset. Thus, the compound was outside the warning leverage h^* which was found calculated as 0.68.

4.4 The docking study

Molecular docking was run between the protein of *P. infestans* effector target site (PDB ID: 2NAR; >95%

purity) and the ligands to investigate/examine the mode of interaction of the ligands with the macromolecular target site of the protein. The interaction of all the 32 compounds with the receptor active site was carried out in which the receptor–ligand interactions with lower energy, i.e., those with better docking scores, were recorded in Table 9. The table consists of the ligands with their binding affinity, the H-bonds, H-bond distances, as well as their hydrophobic and electrostatic interactions. The binding affinity for all the compounds is between the range of -8.2 to -9.3 kcal/mol. Compound 13 possessed the highest binding score with -9.3 kcal/mol and showed an interaction mode with H-bonds (GLU88 with H-bond distance of 2.78089Å and GLN67Å with H-bond distance of 2.91512Å),

Table 9 The binding energy, H-bonds, H-bond distances, hydrophobic and electrostatic interactions of receptor, and the ligands with the highest docking scores

Ligands	Binding energy (kcal/mol)	H-bonds	H-bond distances	Hydrophobic interactions	Electrostatic interactions
11	-9.1	GLU88, GLN67	2.86658, 2.77632	TYR87, TYR71, LEU52, TYR87, ALA69	ASP56, GLU88
13	-9.3	GLU88, GLN67	2.78089, 2.91512	TYR87, TYR71, LEU52, TYR87, ALA69	ASP56, GLU88
14	-8.9	GLN67, GLN67	2.58747, 2.377	TYR87, TYR71, TYR71, ALA69, LEU52	ASP56, GLU88, GLU88
15	-8.9	GLN67, GLN67	2.74137, 2.57359	TYR87, TYR71, ALA69, LEU52	ASP56, GLU88
17	-8.9	GLN67, GLU88, GLN67	3.0818, 3.03569, 2.74824	TYR87, TYR71, LEU52, TYR87 ALA69	ASP56, GLU88
24	-8.9	GLN67, GLY68, GLN67	2.52375, 2.5966, 2.29024	TYR87, ALA69 LEU52	ASP56, GLU88 GLU88
26	-9.0	ASN72	2.22541	TYR87, TYR71, LEU52, TYR87, ALA69, LEU52	ASP56, GLU88
30	-9.0	ASN72	2.49659	TYR87, TYR71, LEU52, TYR87, ALA69, LEU52	ASP56, GLU88 GLU88

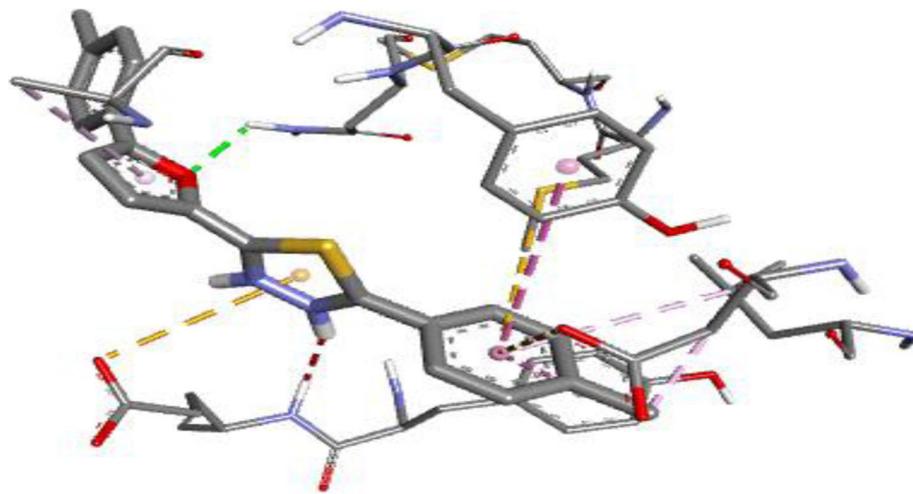


Fig. 6 Receptor–ligand interaction

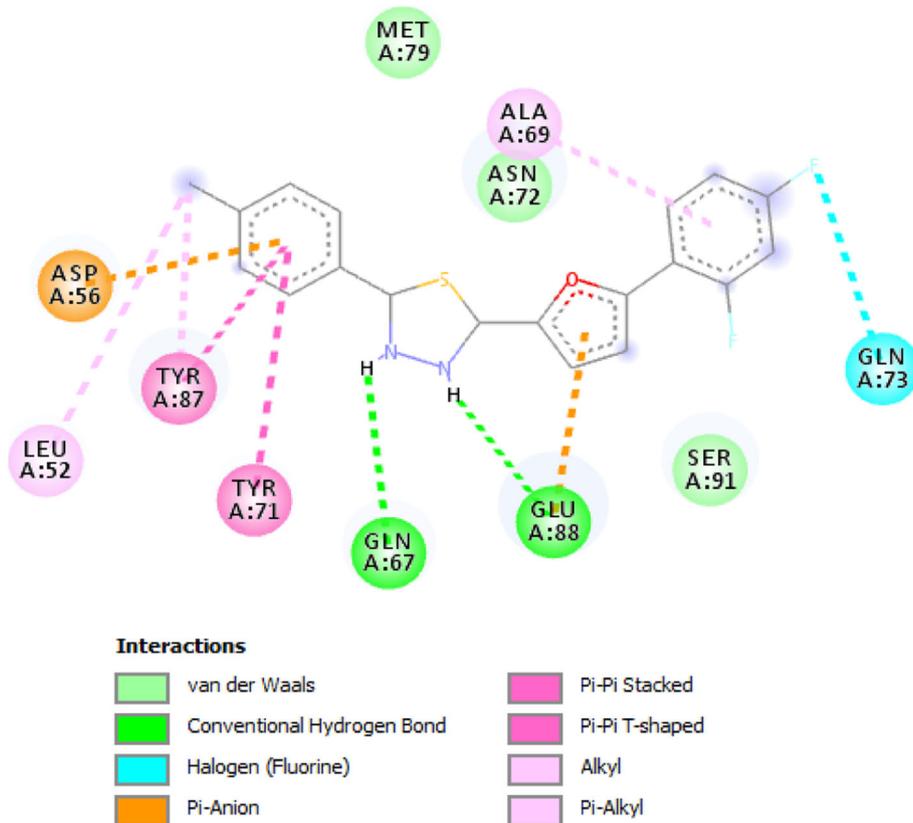


Fig. 7 2D structure showing H-bond interactions between receptor and compound 13

hydrophobic interaction mode of TYR87 (4.7572Å), TYR71, LEU52, TYR87 (4.88051Å), and ALA69 residues.

Figure 6 showed a receptor–ligand interaction while Fig. 7 is the 2D structure which shows that H-bond interaction exists between the receptor and the compound 13 which has a better binding affinity and showed a better interaction with the macromolecular target site of the residue when compared with other compounds as well.

5 Conclusion

This research involves a QSAR and molecular docking studies on 32 compounds of 2,5-disubstituted-1,3,4-thiadiazole derivatives against *P. infestans* effector site. After using DFT to optimize the compounds, GFA was used to generate the built model. Among the four generated models, the fourth model was found to be the optimal, having appreciable statistical parameters with $R^2 = 0.798318$, $R^2_{adj} = 0.750864$, cross-validation R^2 (Q^2_{cv}) = 0.662654, and external validation $R^2_{pred} = 0.624008$. Descriptors AMR and SCH-7 were the first and second highest contributors toward the generation of the selected model, and thus, their increase will increase the activity of the compound while WG.unity and Wnu2.eneg have a low effect on the model, therefore, their increase will have no much significance on the activity of the compound against *P. infestans*.

According to the docking scores, almost all the ligands (compounds) showed high binding affinity/strong inhibition activity against *P. infestans* effector site. However, ligands 11, 13, 14, 15, 17, 24, 26, and 30 showed higher binding affinity ranging from – 8.9 to – 9.3 kcal/mol. With ligand 13 having the highest binding energy of – 9.3 kcal/mol. This compound [13] was able to strongly dock at the binding pocket of the *P. infestans* effector site (2NAR) producing an H-bond as well as hydrophobic interaction with the target site.

The generated QSAR model provides a worthy idea on ligand-based design whereas the molecular docking analysis suggested an approach toward the structure-based design of novel and more potent compounds against *P. infestans*.

Abbreviations

B3LYP: Becke's three-parameter read Yang-Parr hybrid; DFT: Density function theory; GFA: Genetic function approximation; PDB: Protein data bank; *P. infestans*: *Phytophthora infestans*; QSAR: Quantitative structure–activity relationship

Acknowledgements

The authors acknowledge the technical effort of the Department of Chemistry, Ahmadu Bello University, Zaria, Nigeria, and Muhammad Tukur Ibrahim for his useful advice toward the successful completion of the research.

Availability of data and material

Not applicable

Authors' contributions

YI contributed throughout the research work. AU gave directives and technical advices. SU partook in the technical activities. All authors have read and approved the manuscript.

Funding

The authors declare that no funding has been received.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Received: 18 October 2019 Accepted: 22 January 2020

Published online: 14 April 2020

References

- Adeniji SE, Uba S, Uzairu A (2018) QSAR modeling and molecular docking analysis of some active compounds against Mycobacterium tuberculosis receptor (Mtb CYP121). *J Pathog* 2018(1018694):24
- Almajan GL, Barbuceanu SF, Saramet I, Draghici C (2010) New 6-amino-[1, 2, 4] triazolo [3, 4-b] [1, 3, 4] thiadiazines and [1, 2, 4] triazolo [3, 4-b] [1, 3, 4] thiadiazin-6-ones: synthesis, characterization and antibacterial activity evaluation. *Eur J Med Chem* 45(7):3191–3195
- Arthur DE (2017) Toxicity modelling of some active compounds against k562 cancer cell line using genetic algorithm-multiple linear regressions. *J Turkish Chem Soc, Section A: Chem* 4(1):355–374
- Benarous N, Cherouana A, Aubert E, Durand P, Dahaoui S (2016) Synthesis, characterization, crystal structure and DFT study of two new polymorphs of a Schiff base (E)-2-(2, 6-dichlorobenzylidene) amino) benzonitrile. *J Mol Struct* 1105:186–193
- CABI International, (2018). *Phytophthora infestans* (Phytophthora blight). Retrieved from <https://www.cabi.org/isc/datasheet/40970>.
- Cui ZN, Shi YX, Zhang L, Ling Y, Li BJ, Nishida Y, Yang XL (2012) Synthesis and fungicidal activity of novel 2, 5-disubstituted-1, 3, 4-oxadiazole derivatives. *J Agric Food Chem* 60(47):11649–11656
- Cui Z, Li X, Tian F, Yan X (2014) Synthesis and bioactivity of 5-substituted-2-furoyl diacylhydrazide derivatives with aliphatic chain. *Int J Mol Sci* 15(5): 8941–8958
- Cui ZN, Li YS, Hu DK, Tian H, Jiang JZ, Wang Y, Yan XJ (2016) Synthesis and fungicidal activity of novel 2, 5-disubstituted-1, 3, 4-thiadiazole derivatives containing 5-phenyl-2-furan. *Sci Rep* 6:20204
- Edache EI, Hambali HU, Arthur DE, Oluwaseye A, Chinweuba OC (2016) In-silico discovery and simulated selection of multi-target Anti-HIV-1 inhibitors. *Int Res J Pure Appl Chem* 11(1):1–15
- Fry W (2008) *Phytophthora infestans*: the plant (and R gene) destroyer. *Mol Plant Pathol* 9(3):385–402
- Fry WE, Grünwald NJ (2010) Introduction to oomycetes. *The Plant Health Instructor*. <https://doi.org/10.1094/PHI-I-2010-1207-01>.
- Gramatica P, Cassani S, Roy PP, Kovarich S, Yap CW, Papa E (2012) QSAR modeling is not "push a button and find a correlation": a case study of toxicity of (benzo-) triazoles on algae. *Mol Inform* 31(11-12):817–835
- Griffiths RG, Dancer J, O'Neill E, Harwood JL (2003) A mandelamide pesticide alters lipid metabolism in *Phytophthora infestans*. *New Phytol* 158(2):345–353
- Haas BJ, Kamoun S, Zody MC, Jiang RH, Handsaker RE, Cano LM et al (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461(7262):393
- Henfling JW (1987) Late blight of potato: *Phytophthora infestans*, Technical Information Bulletin 4. International Potato Center, Lima
- Ibezim EC, Duchowicz PR, Ibezim NE, Mullen LMA, Onyishi IV, Brown SA, Castro EA (2009) Computer-aided linear modeling employing QSAR for drug discovery. *Sci Res Essays* 4(13):1559–1564
- Ibrahim MT, Uzairu A, Shallangwa GA, Ibrahim A (2018) Computational studies of some biscoumarin and biscoumarin thiourea derivatives as α -glucosidase inhibitors. *J Eng Exact Sci* 4(2):0276–0285

18. Jatav V, Mishra P, Kashaw S, Stables JP (2008) CNS depressant and anticonvulsant activities of some novel 3-[5-substituted 1,3,4-thiadiazole-2-yl]-2-styryl quinazoline-4 (3H)-ones. *Eur J Med Chem* 43(9):1945–1954
19. Kadi AA, Al-Abdullah ES, Shehata IA, Habib EE, Ibrahim TM, El-Emam AA (2010) Synthesis, antimicrobial and anti-inflammatory activities of novel 5-(1-adamantyl)-1,3,4-thiadiazole derivatives. *Eur J Med Chem* 45(11):5006–5011
20. Karthikeyan C, Moorthy NHN, Trivedi P (2009) QSAR study of substituted 2-pyridinyl guanidines as selective urokinase-type plasminogen activator (uPA) inhibitors. *J Enzyme Inhib Med Chem* 24(1):6–13
21. Mustapha A, Shallangwa G, Ibrahim MT, Bello AU, Ebuka DA, Uzairu A, Mamza P (2018) QSAR studies on some C14-urea tetrandrine compounds as potent anti-cancer against leukemia cell line (K562). *J Turkish Chem Soc, Section A: Chem* 5(3):1387–1398
22. Nowicki M, Lichocka M, Nowakowska M, Kłosińska U, Kozik EU (2012) A simple dual stain for detailed investigations of plant-fungal pathogen interactions. *Vegetable Crops Res Bull* 77:61–74
23. Oruç EE, Rollas S, Kandemirli F, Shvets N, Dimoglo AS (2004) 1,3,4-thiadiazole derivatives. Synthesis, structure elucidation, and structure–antituberculosis activity relationship investigation. *J Med Chem* 47(27):6760–6767
24. Parvatham K, Veerakumari L, Shoba G (2015) Molecular docking studies of acetate-succinate CoA-transferase of *Ascaris lumbricoides* with a few phytochemicals and anthelmintics. *J Comput Methods Mol Design* 5(4):1–10
25. Plech T, Wujec M, Kosikowska U, Malm A, Kaproń B (2012) Studies on the synthesis and antibacterial activity of 3,6-disubstituted 1,2,4-triazolo [3, 4-b]1,3,4-thiadiazoles. *Eur J Med Chem* 47:580–584
26. Pourbasheer E, Aalizadeh R, Ganjali MR, Norouzi P (2014) QSAR study of IKK β inhibitors by the genetic algorithm: multiple linear regressions. *Med Chem Res* 23(1):57–66
27. Roy K, Kar S, Das RN (2015) A primer on QSAR/QSPR modeling: fundamental concepts. Springer
28. Taher AT, Georgey HH, El-Subbagh HI (2012) Novel 1,3,4-heterodiazole analogues: synthesis and in-vitro antitumor activity. *Eur J Med Chem* 47:445–451
29. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31(2):455–461
30. Vio L, Mamolo MG, Laneve A (1989) Synthesis and antihypertensive activity of some 1,3,4-thiadiazole derivatives. *Farmaco (Societa chimica italiana: 1989)* 44(2):165–172

Publisher's Note

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

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
