

RESEARCH

Open Access



In silico studies of novel pyrazole-furan and pyrazole-pyrrole carboxamide as fungicides against *Sclerotinia sclerotiorum*

Yusuf Isyaku^{*} , Adamu Uzairu and Sani Uba

Abstract

Background: Pyrazole-furan and pyrazole-pyrrole moiety are among the molecular structures that were found to have an extensive range of applications in the field of medicine and agrochemical due to their wide spectrum of biological activities. These include antimicrobial activity, anti-glaucoma activity, ocular hypertension activity, and antifungal activity.

Results: An in silico study was carried out on 37 compounds of pyrazole-furan and pyrazole-pyrrole carboxamide derivatives against *Sclerotinia sclerotiorum*. Using Spartan 14 software, optimization of the compounds was performed at the DFT/B3LYP/6-31G* quantum mechanical method. PaDEL descriptor software was used to calculate the molecular descriptors, and a Generic Function Approximation (GFA) was employed to generate the model. Out of four models generated, model 1 was found to be the optimal and has the following statistical parameters; $R^2 = 0.83485$, $R^2_{adj} = 0.793563$, cross-validated $R^2 = 0.74037$, and external $R^2 = 0.58479$. Molecular docking study was carried out between the antifungal compounds, and the binding site of *S. sclerotiorum* (PDB CODE 2X2S) in which compound 7 was identified to have the highest binding energy of -7.5 kcal/mol. This compound "7" has a strong affinity with the macromolecular target point of the *S. sclerotiorum* (2x2s), producing H-bond and as well as the hydrophobic interaction at target point of the amino acid residue. Considering compound 7 as our scaffold, four (4) more potent compounds (7a, 7b, 7c, and 7d) were designed using optimization method of structure-based designed which have the following docking score, -7.7 , -7.8 , -7.7 , and -7.7 kcal/mol.

Conclusion: Statistical analyses including variance inflation factor (VIF), mean effect (ME), and applicability domain were conducted on the model. Considering an interpretation of the descriptors given in the discussion, the QSAR model provided an idea of ligand-based design while the molecular docking gave an insight on structure-based design of the new compounds with better activity against *S. sclerotiorum* in which four (4) compounds 7a, 7b, 7c, and 7d were designed and discovered to be of high quality and have greater binding affinity compared to the one obtained from the literature (compound 7).

1 Background

Sclerotinia sclerotiorum otherwise called cottony rot, blossom blight, stem rot, crown rot, or watery soft rot, is a fungal pathogen that results in a plant disease known as "white mold" under favorable conditions. This pathogen produces black resting structures (called sclerotia) on the affected plant. It can be found in different parts of the world with an extensive range of hosts [4]. *S. sclerotiorum* causes great losses when onset on a favorable environmental and extensive care or control measures

should drastically be taken [23]. Herbaceous, succulent plants (particularly flowers) and vegetables are the common hosts.

Pyrazole-furan and pyrazole-pyrrole moiety are among the molecular structures found to have an extensive wide range of applications in the field of medicine and agrochemical due to their wide range of biological activity such as antimicrobial activity [8], anti-glaucoma and ocular hypertension activity [20], antifungal activity [18], and as D-amino acid oxidase (DAAO) inhibitors [12] among others.

* Correspondence: isyakuyusuf01@gmail.com

Department of Chemistry, Ahamadu Bello University, Zaria, Nigeria

The quantitative analysis of the structure-activity relationship (QSAR) is among the major efficient methods to optimize the main compounds and design novel compounds. QSAR is used to predict bioactivity of the compounds such as toxicity, carcinogenicity, or mutagenicity, depending on the structural characteristics of the molecules and the actual mathematical models. Nowadays, one can easily and accurately calculate quantum chemical parameters of the compounds due to fast development in computer technology as well as theoretical quantum chemical study which helped in predicting the new compounds with better activity than the existing ones. This quantum chemical calculation is extensively applied while forming the QSAR models [13]. Molecular docking helped to investigate the capacity of the prepared compounds toward interaction with the protein residue of the target organism and to also predict the preferred orientation of the molecules.

The objective of the current work is to build a new model that can predict the activity of chemical products with much better activity against *S. sclerotiorum* using Genetic Function Approximation (GFA) and molecular docking techniques.

2 Methods

2.1 Dataset

In this work, a series of 37 compounds were used to generate the relationship between chemical traces of the compound and its antifungal activity. These 37 compounds of novel pyrazole-furan and pyrazole-pyrrole carboxamide were obtained from the previous work [24]. The logarithm of the measured EC_{50} (μM) against antifungal activity given by pEC_{50} ($pEC_{50} = -\log 1/EC_{50}$) was taken as a dependent variable; therefore, the data was linearly correlated with the independent variables (descriptors) [10]. The dataset was represented in Table 1 and Fig. 1.

2.2 Optimization

The dataset was optimized at a level of density function theory (DFT) by applying Becke's three-parameter read-Yang-Parr hybrid (B3LYP) function together with a "6-31G*" basis set of Spartan14 software [6]. Graphical-user-interface of Spartan14 software was used to draw the 2D molecular structures of the dataset which were later exported in the form of 3D. The optimized structures were then taken to PaDEL descriptor software to generate the quantum molecular descriptors [25].

2.3 Molecular descriptors calculations

Molecular descriptors are the properties of the molecule in numerical/mathematical values. PaDEL descriptor software was used to further calculate additional energy

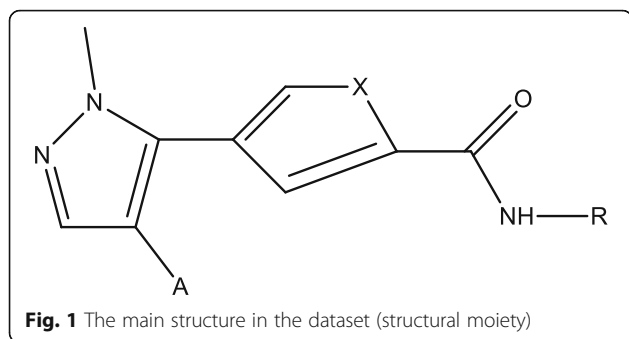
Table 1 Dataset

S/no	X	A	R	pEC_{50}
1	O	Cl	Ph	0.748188
2	O	Cl	2-Cl-Ph	0.70757
3	O	Cl	3-Cl-Ph	1.238046
4	O	Cl	4-Cl-Ph	0.919078
5	O	Cl	2-MeO-Ph	1.161368
6	O	Cl	3,4-(MeO) ₂ -Ph	2.180986
7	O	Cl	2-CF ₃ O-Ph	0.342423
8	O	Cl	4-CF ₃ O-Ph	1.139879
9	O	Cl	2-CF ₃ -Ph	1.235528
10	O	Cl	4-CF ₃ -Ph	1.227887
11	O	Cl	4-Me-Ph	1.209515
12	O	Cl	2,5-Me ₂ -Ph	1.143015
13	O	Cl	3-NO ₂ -Ph	0.94939
14	O	Cl	4-Cyano-Ph	1.11059
15	O	Cl	4-Cyanomethyl-Ph	1.113943
16	O	H	Ph	2.474508
17	O	H	3-Cl-Ph	2.150756
18	O	H	4-Cl-Ph	1.557507
19	O	H	2-MeO-Ph	2.232488
20	NMe	Cl	Ph	1.033424
21	NMe	Cl	2-Cl-Ph	0.755875
22	NMe	Cl	3-Cl-Ph	1.164353
23	NMe	Cl	4-Cl-Ph	1.532754
24	NMe	Cl	2-MeO-Ph	0.544068
25	NMe	Cl	3-MeO-Ph	1.912753
26	NMe	Cl	3,4-MeO ₂ -Ph	0.944483
27	NMe	Cl	2CF ₃ -Ph	1.834421
28	NMe	Cl	4-CF ₃ -Ph	2.115943
29	NMe	Cl	2-(4-Cl-Phenyl)-Ph	2.439806
30	NMe	Cl	4-t-Bu-PhCH ₂	0.973128
31	NMe	Cl	6-Cl-Pyridin-3-yl-CH ₂	0.875061
32	NMe	Cl	4-Cl-PhCH ₂ CH ₂	0.462398
33	NMe	Cl	1-naphthalene	1.884795
34	NMe	H	Ph	2.635383
35	NMe	H	2-MeO-Ph	1.751279
36	NMe	H	PhCH ₂	2.074085
37	NMe	H	4-t-Bu-PhCH ₂	2.121888

of those low-energy conformers, where a total of 1875 descriptors were calculated [1].

2.4 Data division

To get a validated model, the dataset was divided into training and test sets (3:1). Following the Kennard-Stone algorithm method, the division was performed in such a



way that the compounds forming the training set (70% of the data) and the test set (30% of the data) were shared within an entire descriptive space filled by the complete dataset [7].

2.5 Model building and validation

The generated molecular descriptors were taken for regression analysis, with experimental activities as dependent variables and the molecular descriptors served as independent variables. Using the Genetic Function Approximation method (GFA) incorporated in the Material Studio 2017 software [9], the training set compounds were utilized to develop the QSAR model. Four QSAR models were built, and the best model was chosen according to the one with the lowest score of lack of fit (LOF) given as follows:

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

where SSE represents the sum of squares of errors, d is a smoothing parameter defined by the user, c is the number of terms a model possessed in addition to the constant term, M gives the number of samples present in the training set, and p is the overall number of descriptors present in all terms of the model excluding the constant term [11].

2.5.1 Internal validation

The generated model was validated internally by the following parameters:

- The correlation coefficient (R^2): explain the division of overall variation ascribed to the built model. The accepted value of R^2 ranges from 0.5 to < 1 and the more the value of R^2 approaches 1.0 the better the model. Though there are other analyses that the model must pass before we can consider it a good model, being the most common internal validation pointer, R^2 is expressed as follows:

$$R^2 = 1 - \frac{\sum (Y_{\text{expt}} - Y_{\text{predt}})^2}{\sum (Y_{\text{expt}} - \bar{Y}_{\text{train}})^2} \quad (\text{ii})$$

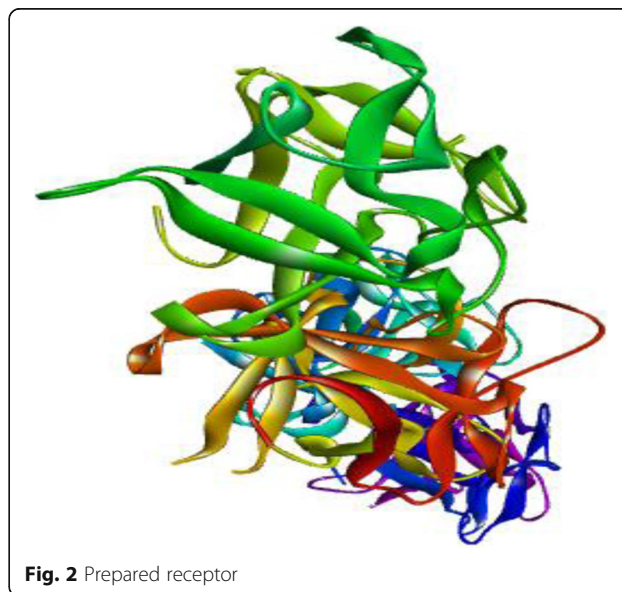
where Y_{expt} , Y_{predt} , and \bar{Y}_{train} represent the experimental, predictive, and average activities of the training set [3].

- Adjusted R^2 : The value of R^2 is inconsistent to evaluate the power of the built model; thus, R^2 is adjusted to restore and stabilize the model. This adjusted R^2 is defined in Eq. iii as:

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

where p presents the number of descriptors that constituted the model, while n gives the number of training set compounds [14].

- Cross-validated R^2 : The validity of the models was identified by a cross-validation test measured by predictive Q^2_{cv} . For a leave one out (LOO) cross-validation, a data point is eliminated (left-out) in the set and the model is readjusted and then compared the predicted value of the eliminated data point to its real value. This is repeated until each data removed. We can then calculate the value of Q^2_{cv} using the sum of the squares of these elimination residues as in the below equation:



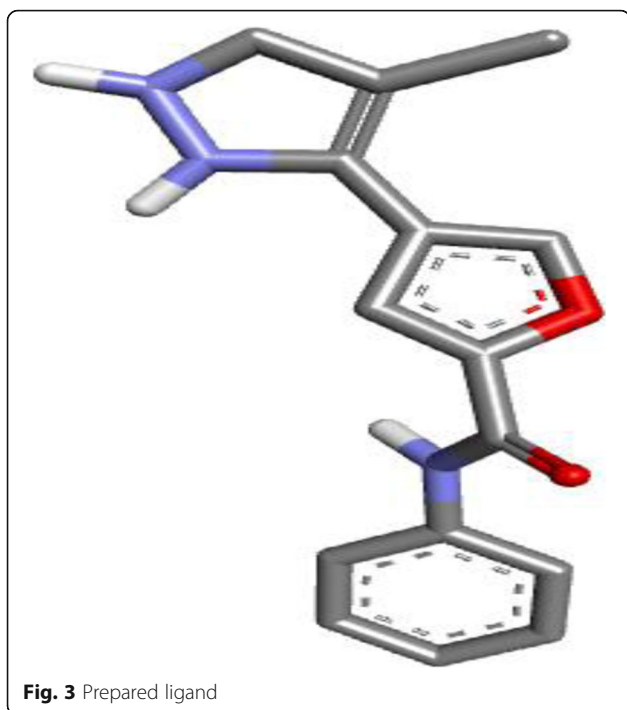


Fig. 3 Prepared ligand

$$Q_{cv}^2 = 1 - \frac{\sum (Y_{predt} - Y_{expt})^2}{\sum (Y_{expt} - \bar{Y}_{train})^2} \quad (\text{iv})$$

where Y_{expt} , Y_{predt} , and \bar{Y}_{train} represent the experimental, predictive, and average activities of the training set [2].

2.5.2 External validation

The prediction capacity of the model was examined by an external validation through the ability of the model to predict the activity values of the test set compounds as well as its application in the calculating the predicted value of R^2_{pred} according to the equation below:

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

where Y_{predt} and Y_{expt} are the test set's experimental and predicted activities while Y_{train} gives the average activity of the training set [5].

2.6 Statistical analysis of the descriptors

2.6.1 Variance inflation factor (VIF)

It was 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.

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

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 inter-correlation 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, it indicates the instability of the model and needs to be reexamined [16, 19].

2.6.2 Mean effect (ME)

The average effect (mean effect) correlates the effect or influence of given molecular descriptors to the activity of the compounds that made up the model. The sign of descriptors shows the direction of their deviation toward the activity of 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:

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

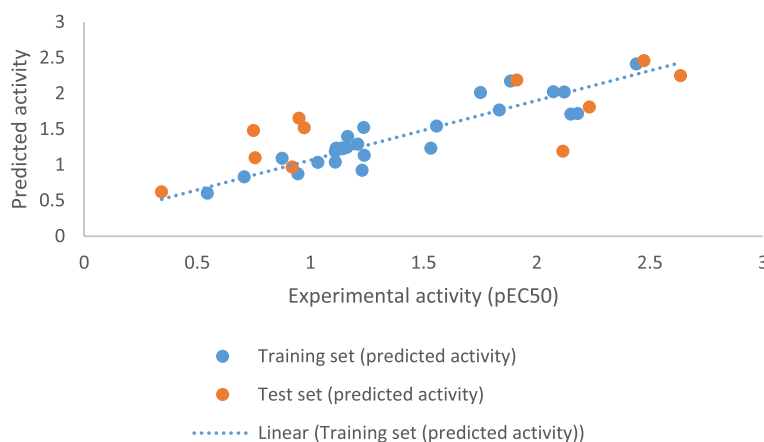


Fig. 4 A plot of predicted activity versus experimental activity

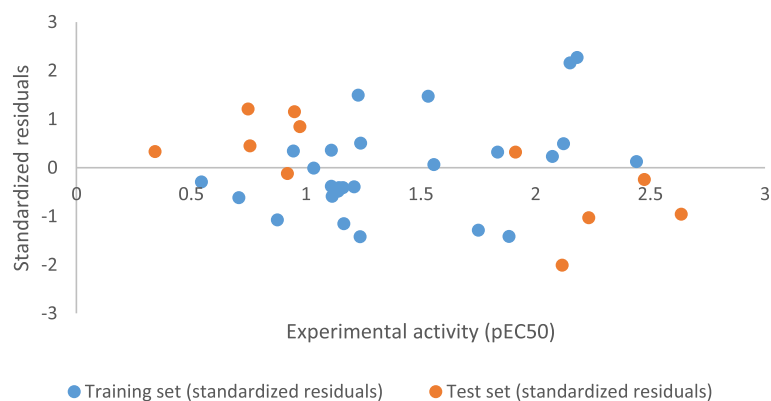


Fig. 5 A plot of standardized residual versus experimental activity

where B_j and D_j are the j -descriptor's coefficient in a model and the values of each descriptor in training set, while m and n stand for the number of molecular descriptors and the number of compounds in the training set. To evaluate the significance of the model, the ME of all the descriptors was calculated [11].

2.6.3 Applicability domain

To confirm the reliability of the model and to examine the outliers as well as the influential compounds, it was very important to evaluate its domain of applicability. It aimed to predict 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 sets of the compounds. This could be achieved by employing William's plot which was plotted using standardized residuals versus the leverages. The leverages for a particular chemical compound was given as:

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

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

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

where n = number of compounds in the training set and p is the number of descriptors present in the model [15].

2.7 Ligand and receptor preparation

From the RCSBPDB (www.rcsb.org), the PDB format of the receptor was successfully downloaded. This was then taken to the discovery studio for an appropriate

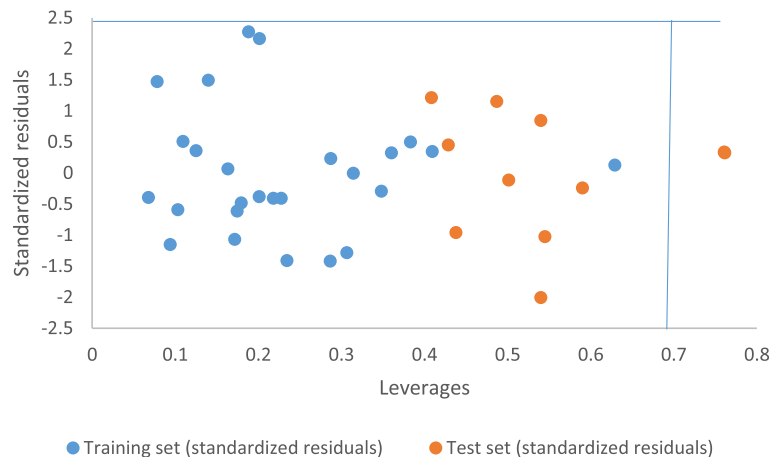


Fig. 6 William's plot

preparation whereby all the residues associated with the downloaded receptor (such as a ligand, water molecules, and other traces) were removed. The ligands (the optimized compounds) which were in the SDF file were transformed into a PDB file format. The prepared structure of *S. sclerotiorum* and prepared compounds were docked using Autodock Vina 4.2 [22]. Discovery Studio Visualizer was also used to visualize the docking results. Figures 2 and 3 showed the prepared receptor and ligand [15].

2.8 Optimization method of structure-based design

This refers to the optimization of known molecules through evaluating its proposed analogs within the binding cavity [17]. Discovery studio was used to visualize the receptor-ligand interactions in which different interactions such as H-bond and hydrophobic interaction formed between compound 7 and the receptor (PDB ID: 2x2s) were studied. Based on the knowledge of this interaction, the designed compounds were proposed in which they were drawn, optimized, converted to PDB, and later docked with the receptors to record their potency.

3 Results

3.1 QSAR model (Figs. 4 and 5)

The equation of the best and selected model (model 1):

$$\begin{aligned} \text{pEC}_{50} = & 3.742328478 * \text{ATSc5} - 0.055068016 * \text{ATSm1} \\ & + 0.308420856 * \text{nBondsM} \\ & + 0.128047514 * \text{MDEC-13} \\ & + 0.199822703 * \text{ndsN} - 0.389418304. \end{aligned}$$

(x)

3.2 Applicability domain (Fig. 6)

3.3 Docking studies (Figs. 7 and 8)

4 Discussion

4.1 QSAR model

QSAR examination was carried out to relate the structure-activity relationship of the novel pyrazole-furan and pyrazole-pyrrole carboxamide derivatives as potent inhibitors of *S. sclerotiorum*. Five descriptors were utilized in the construction of the QSAR model to predict the activities of the inhibitory compounds based on the Genetic Function Approximation (GFA). The first model (vi) was chosen as the optimal model due to its statistical significance.

All the validation parameters that signify the stability, robustness, and the prediction capability of the model were presented in Table 2. The names, symbols, and classes of the five (5) selected descriptors that made up the model are presented in Table 3.

From the Table 2 of validation parameters, the highly calculated R^2 value (0.835) for the predicted activities indicated that the model is internally good. Tables 4 and 5 represent the external validation and the calculation of predicted R^2 of the best-chosen model.

4.2 Interpretation of descriptors

ATSc5 and ATSm1 are 2D auto-correlation descriptors built by Todeschini and Consonni and defined by Moreau-Broto autocorrelation as ATS autocorrelation descriptor, weighted by charges (ATSc5), and weighted

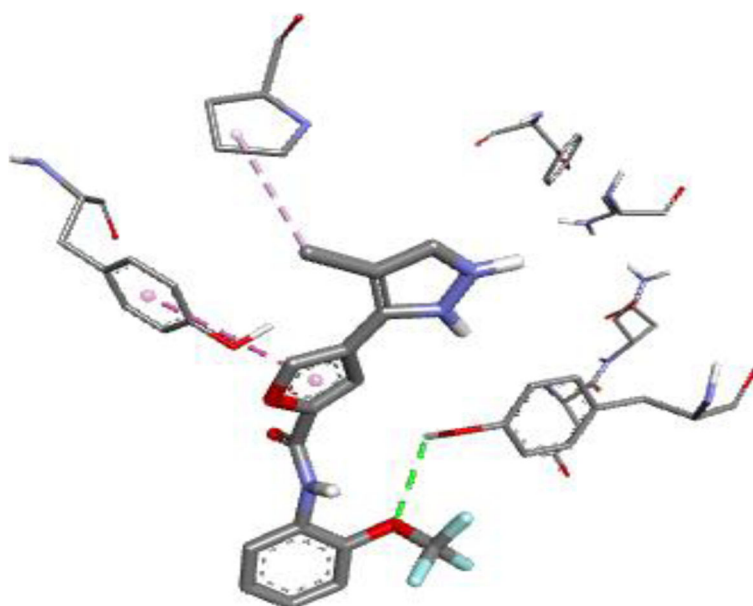


Fig. 7 The interaction between the compound with the highest docking score and receptor

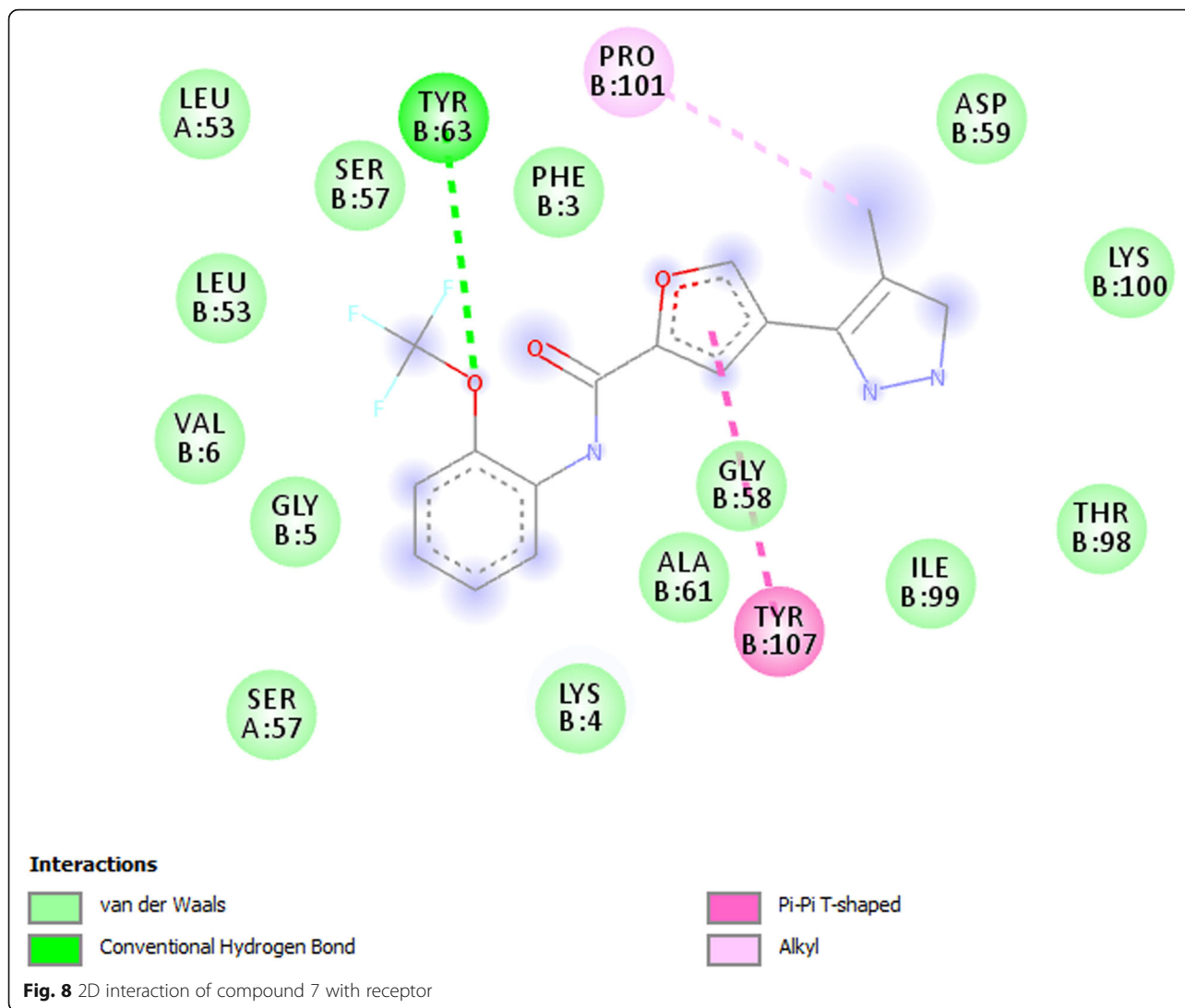


Table 2 Validation parameter of the model

Friedman LOF	0.236909
R-squared	0.83485
Adjusted R-squared	0.793563
Cross-validated R-squared	0.74037
Significant regression	Yes
Significance-of-regression F value	20.220461
Critical SOR F value (95%)	2.732939
Replicate points	0
Computed experimental error	0.0000000
Lack-of-fit points	20
Min expt. error for non-significant LOF (95%)	0.18129900

Table 3 List of descriptors with their descriptions and class

S/no	Descriptor	Description	Class
1	ATSc5	ATS autocorrelation descriptor, weighted by charges	2D
2	ATSm1	ATS autocorrelation descriptor, weighted by scaled atomic mass	2D
3	nBondsM	The total number of bonds that have a bond order greater than one (aromatic bonds have bond order 1.5).	2D
4	MDEC-13	Molecular distance edge between all primary and tertiary carbons	2D
5	ndsN	Count of atom-type E-State: = N-	2D

Table 4 External validation

S/no	Activity	ATSc5	ATSm1	nBondsM	MDEC-13	ndsN	Y_{pred}	$(Y_{\text{pred}} - Y_{\text{obs}})$
1	0.748188	-0.09095	30.34183	10	0	4	1.482858	0.73467
4	0.919078	-0.10029	39.05478	10	0	4	0.96808	0.049002
7	0.342423	-0.27593	40.62243	10	0	6	0.624095	0.281672
13	0.94939	-0.08001	35.25075	10	0	6	1.653117	0.703727
16	2.474508	-0.06526	21.62888	10	0	6	2.458433	-0.01607
19	2.232488	-0.18624	24.40335	10	1.23799	5	1.811596	-0.42089
21	0.755875	-0.12052	39.64029	10	1.857354	4	1.097961	0.342086
25	1.912753	-0.00494	33.70181	10	2.876191	5	2.187817	0.275064
28	2.115943	-0.0674	39.43347	9	1.782569	5	1.189993	-0.92595
30	0.973128	-0.08618	31.92734	9	1.668818	5	1.518469	0.545341
34	2.635383	-0.05827	22.21439	10	1.54067	4	2.250008	-0.38538

by scaled atomic mass (ATSm1) [21]. It is given by the equation below:

$$ATS_d = \sum_{i=1}^A \sum_{j=1}^A \delta_{ij} \cdot (w_i \cdot w_j)_d = w^T \cdot {}^m B \cdot w$$

where w = any atomic property, A = number of atoms, d = topological distance, δ_{ij} = Kronecker delta, ${}^m B$ = m th order binary sparse matrix, and w = A -dimensional vector of atomic properties. ATSc5 and ATSm1 consist of negative mean effects which imply that they harm the predictive activity of the model, which means that for any increase in these descriptors there will be a decrease in the activity of the compounds. The 2D descriptor "nBondsM" defined as the dimensionless total number of bonds that have bond order greater than one and is encoded to an unsaturation. It referred to as the number of non-kekulized structures. It has a positive mean effect which signified an increase in it, will enhance the activity of the chemical compound. A

simple cluster descriptor MDEC-13 (defined as molecular distance edge between all primary and tertiary carbons) is a 2D descriptor with a positive mean effect which signifies an increase of a simple cluster to a compound increases the activity of that compound. An E-State descriptor "ndsN" defined as Count of atom-type E-State: =N- is a 2D descriptor with positive mean effects which indicated that an increase in these descriptors enhances the activity of the compounds.

The experimental, predictive, and residual activities for both training and test sets are shown in Table 6. The residual value is the difference between the predicted and actual activities.

4.3 Pearson's correlation

To evaluate the relationships between each of the descriptors used in the model, Pearson's correlation was carried out on the descriptors of the built model and the results were presented in Table 7. The results show that the descriptors are significantly inter-correlated because none of their correlation coefficients are up to 0.5, and this indicates the robustness as well as the stability of the model. The Variance Inflation Factor (VIF) values for each of the five descriptors were not up to 2, which indicated that the descriptors and the model are stable and accepted.

Figure 5 presents a graph of observed activity versus standardized residual and shows a random dispersion at the baseline where the standardized residual is zero. This shows no systematic error occurred in the built model.

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 impact 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,

Table 5 Continuation of external validation

$(Y_{\text{pred}} - Y_{\text{obs}})^2$	$Y_{\text{meanTrain}}$	$Y_{\text{obs}} - Y_{\text{meanTr}}$	$(Y_{\text{pred}} - Y_{\text{MeanTrain}})^2$
0.53974	1.403369	-0.65518	0.429262
0.002401	1.403369	-0.48429	0.234538
0.079339	1.403369	-1.06095	1.125607
0.495232	1.403369	-0.45398	0.206097
0.000258	1.403369	1.071139	1.147338
0.17715	1.403369	0.829119	0.687438
0.117023	1.403369	-0.64749	0.419249
0.07566	1.403369	0.509384	0.259472
0.857383	1.403369	0.712574	0.507762
0.297397	1.403369	-0.43024	0.185108
0.148514	1.403369	1.232014	1.517859
$\Sigma(Y_{\text{pred}} - Y_{\text{obs}})^2 = 2.790097$			$\Sigma(Y_{\text{pred}} - Y_{\text{MeanTrain}})^2 = 6.719729$
$R^2 = (1 - 2.790097/6.719729) = 0.58479$			

Table 6 The experimental, predictive, and the residual activity for both training set and test sets

S/n	pEC50	Predicted pEC50	Residuals
1	0.748188	1.482858	0.73467
2	0.70757	0.832323	-0.12475
3	1.238046	1.134367	0.103679
4	0.919078	0.96808	0.049002
5	1.161368	1.243926	-0.08256
6	2.180986	1.718259	0.462726
7	0.342423	0.624095	0.281672
8	1.139879	1.237404	-0.09753
9	1.235528	1.523985	-0.28846
10	1.227887	0.923353	0.304534
11	1.209515	1.289041	-0.07953
12	1.143015	1.225679	-0.08267
13	0.94939	1.653117	0.703727
14	1.11059	1.036822	0.073768
15	1.113943	1.233488	-0.11955
16	2.474508	2.458433	-0.01607
17	2.150756	1.710297	0.44046
18	1.557507	1.54401	0.013497
19	2.232488	1.811596	-0.42089
20	1.033424	1.033728	-0.0003
21	0.755875	1.097961	0.342086
22	1.164353	1.398481	-0.23413
23	1.532754	1.232789	0.299966
24	0.544068	0.603402	-0.05933
25	1.912753	2.187817	0.275064
26	0.944483	0.873949	0.070533
27	1.834421	1.768216	0.066205
28	2.115943	1.189993	-0.92595
29	2.439806	2.413539	0.026267
30	0.973128	1.518469	0.545341
31	0.875061	1.092392	-0.21733
32	0.462398	1.057623	-0.59523
33	1.884795	2.171681	-0.28689
34	2.635383	2.250008	-0.38538
35	1.751279	2.012284	-0.26101
36	2.074085	2.026278	0.047807
37	2.121888	2.01996	0.101928

signs, and their mean effects as well. The p values of the five descriptors (at 95% c.l.) that made up the model are all <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.

Table 7 Pearson's correlation

	ATSc5	ATSm1	nBondsM	MDEC-13	ndsN	VIF
ATSc5	1					1.196977
ATSm1	0.24955	1				1.350698
nBondsM	-0.12262	0.377168	1			1.296405
MDEC-13	-0.28008	-0.20619	0.007194	1		1.179707
ndsN	0.18239	0.114442	-0.17875	-0.28875	1	1.156338

4.4 Applicability domain

A graph of leverages for each compound of dataset versus their standardized residuals (in terms of William's plot) was plotted to discover the outliers and the chemical influential values of the model. The domain of applicability was established within a box at ± 2.5 limits for the residuals and leverage threshold h^* (where h^* calculated to be 0.67). The result shows that except one (1) compound from the test set (with EC_{50} of 1.91275), all the molecules in the dataset are within the box of the applicability domain of the model. This may be characterized by differences in chemical structures considering the rest of the compounds highlighted in the dataset. Figure 6 shows William's plot.

4.5 Docking studies

A molecular docking study was performed between the ligands (compounds) and the target site of *S. Sclerotiorum* to investigate the binding affinity of the molecules with the macromolecular active site of the fungus. All the ligands show an interaction with the active site of the *S. sclerotiorum*, which is to say they inhibit the activity of the fungus. Some ligands show high binding energy that varies from -6.9 to -7.5 kcal/mol as presented in Table 8. However, compound 7 shows the highest docking score of -7.5kcal/mol. Compound 7 with the lowest binding free energy, possessed an interaction mode with H-bond of TYR63 and 2.86621 bond length and hydrophobic interaction of TYR107 and PRO101. The interaction between the compound with the highest docking score and the binding pocket of the receptor is shown in Fig. 7 while Fig. 8 is the 2D hydrogen bond interaction of compound 7 with the receptor. Table 9 shows the binding affinity,

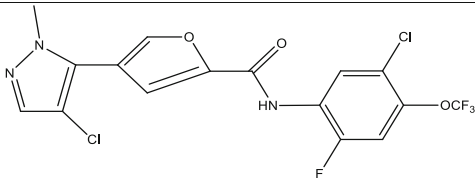
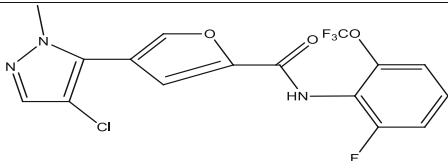
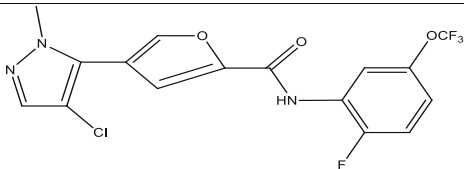
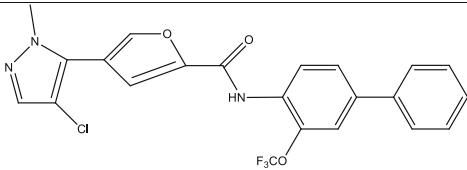
Table 8 The standard regression coefficients " b_j ", the values of mean effect (MF) and confidence interval (p values)

Descriptors	Standard regression coefficient (b_j)	Mean effect (MF)	P value (confidence interval)
ATSc5	3.742328	-0.22563	8.14E-05
ATSm1	-0.05507	-1.37276	1.96E-05
nBondsM	0.308421	2.130095	4.68E-07
MDEC-13	0.128048	0.135861	0.001867
ndsN	0.199823	0.609916	2.38E-06

Table 9 Ligands, binding affinity, H-bond, and hydrophobic interaction between high binding score compounds and receptor

Ligands	Binding affinity (kcal/mol)	Hydrogen bond		Hydrophobic interaction
		Amino acid	Bond distance (Å)	
1	-7.0	VAL6	3.00958	TYR107,PRO101, VAL6
2	-7.0	TRP24,ASN22, ASN20	3.01414, 2.40177, 2.19411	TRP24,TRP24, TRP24,TYR37
5	-7.1	ASN113,GLN121	2.58292,2.0082	THR111,LYS147, ALA140
7	-7.5	TYR63	2.86621	TYR107,PRO101
13	-7.0	ASN113,GLN121,SER117	2.53428,1.91709 3.41568	THR111,LYS147, ALA140
24	-7.4	GLY58,VAL6	2.96167,3.07231	GLY5,GLY58,ASP59,PRO101,VAL6
30	-6.9	ASN22,ASN20	2.32547,2.16498	TRP24,TRP24, TYR37
31	-7.1	ASN22,ASN20	2.13643,2.13056	TRP24,TRP24, TRP24

Table 10 Designed compounds

Serial No.	Compounds	Binding affinity (kcal/mol)
7a	 <p>4-(4-chloro-1-methyl-1H-pyrazol-5-yl)-N-(5-chloro-2-fluoro-4-(trifluoromethoxy)phenyl)furan-2-carboxamide</p>	-7.7
7b	 <p>4-(4-chloro-1-methyl-1H-pyrazol-5-yl)-N-(2-fluoro-6-(trifluoromethoxy)phenyl)furan-2-carboxamide</p>	-7.8
7c	 <p>4-(4-chloro-1-methyl-1H-pyrazol-5-yl)-N-(2-fluoro-5-(trifluoromethoxy)phenyl)furan-2-carboxamide</p>	-7.7
7d	 <p>4-(4-chloro-1-methyl-1H-pyrazol-5-yl)-N-(3-(trifluoromethoxy)-[1,1'-biphenyl]-4-yl)furan-2-carboxamide</p>	-7.7

hydrogen bond, and hydrophobic interaction of the high docking score compounds.

4.6 Design

In this study, an optimization method of structure-based drug design approach was employed to design new anti-fungal compounds with a better activity using compound 7 as a scaffold (which has binding free energy/docking score of -7.5 kJ/mol). On the bases of the interaction between the compound and the receptor, four (4) derivatives of the compound were designed through structural modification of the compound as understood from its docking analysis. All the proposed compounds were docked using autodock vina and their binding free energies were recorded. All the designed compounds were found to be more potent than the scaffold (compound 7) with docking scores of -7.7 , -7.8 , -7.7 , and also -7.7 kJ/mol as represented in Table 10. The name, structure, and binding affinity of the designed compounds are represented in Table 10 above.

5 Conclusion

This research involves a QSAR and molecular docking studies on 37 compounds of pyrazole-furan and pyrazole-pyrrole carboxamide derivatives against *S. sclerotiorum*. Density function theory (DFT) was used for molecules optimization, while Generic Function Approximation (GFA) was employed in generating the built model. Out of four models built, the first model was identified to be the optimal with the following statistical parameters; $R^2 = 0.83485$, $R^2_{adj} = 0.793563$, cross-validated $R^2 = 0.74037$, and external $R^2 = 0.58479$. A decrease in negative descriptors (like ATSc5 and ATSm1) and an increase in positive descriptors (like nBondsM, MDEC-13, and ndsN) will improve the activity of the compounds against *S. sclerotiorum*. According to the docking scores, most of the ligands (compounds) show good inhibitory activity against *S. sclerotiorum*. However, ligand 6 showed a higher binding affinity of -7.5 kcal/mol. This compound has a strong affinity with the macromolecular target point of the *S. sclerotiorum* (2x2s), producing hydrogen bond as well as the hydrophobic interaction at the target point of amino acid residue. QSAR model gave an idea of ligand-based design while the molecular docking gave an insight on structure-based design of the more potent compounds against *S. sclerotiorum* in which four (4) compounds 7a, 7b, 7c, and 7d were designed and discovered to be of high quality and have greater binding affinity compared to the one obtained from the literature (compound 7).

Abbreviations

QSAR: Quantitative structure-activity relationship; GFA: Genetic Function Approximation; DFT: Density Function Theory; B3LYP: Becke's three-parameter read-Yang-Parr hybrid; PDB: Protein data bank

Acknowledgments

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

Authors' contributions

YI contributed throughout the research work. AU gives directives and technical advices. SU had partaken in the technical activities. All authors have read and approved the final manuscript.

Funding

The authors declare no funding has been received.

Availability of data and materials

Not applicable

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interest.

Received: 10 October 2019 Accepted: 30 January 2020

Published online: 22 April 2020

References

- Abdulfatai U, Uzairu A, Uba S (2016) Quantitative structure-activity relationship study of anticonvulsant activity of α -substituted acetamido-N-benzylacetamide derivatives. Cogent Chemistry 2(1):1166538
- Adedirin O, Uzairu A, Shallangwa GA, Abechi SE (2018) QSAR and molecular docking based design of some n-benzylacetamide as Γ -aminobutyrate-aminotransferase inhibitors. J Eng Exact Sci 4(1):0065–0084
- Adeniji SE, Uba S, Uzairu A (2018) QSAR modeling and molecular docking analysis of some active compounds against mycobacterium tuberculosis receptor (Mtb CYP121). J Pathogens 2018(1018694):24
- Amselem J, Cuomo CA, Van Kan JA, Viaud M, Benito EP, Couloux A et al (2011) Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. PLoS Genet 7(8):e1002230
- Arthur DE (2017) Toxicity modelling of some active compounds against k562 cancer cell line using genetic algorithm-multiple linear regressions. J Turkish Chem Soc Sect A Chem 4(1):355–374
- Arthur DE, Uzairu A, Mamza P, Abechi S (2016a) Quantitative structure-activity relationship study on potent anticancer compounds against MOLT-4 and P388 leukemia cell lines. J Adv Res 7(5):823–837
- Arthur DE, Uzairu A, Mamza P, Abechi S, Shallangwa G (2016b) QSAR modelling of some anticancer PGI50 activity on HL-60 cell lines. Albanian J Pharmac Sci 3(1):4–9
- Bondock S, Khalifa W, Fadda AA (2011) Synthesis and antimicrobial activity of some new 4-hetarylpyrazole and furo [2, 3-c] pyrazole derivatives. European journal of medicinal chemistry 46(6):2555–2561
- BIOVIA DS (2017). BIOVIA Workbook, Release 2017; BIOVIA Pipeline Pilot, Release 2017. Dassault Systèmes: San Diego, CA, USA. Available online: <http://www.3dsbiovia.com/> (accessed on 4 December 2018)
- Edache EI, Arthur DE, Abdulfatai U (2017) Tyrosine activity of some tetraketone and benzyl-benzoate derivatives based on genetic algorithm-multiple linear regression. 6(1):2–12
- Edache EI, Uzairu A, Abeche SE (2015) Investigation of 5, 6-dihydro-2-pyrones derivatives as potent anti-HIV agent inhibitors. J Comput Methods Mol Des 5(3):135–149
- Fang QK, Hopkins S, Heffernan M, Chytil M (2009) U.S. Patent No. 7,488,747. U.S. Patent and Trademark Office, Washington, DC
- Gagic Z, Nikolic K, Ivkovic B, Filipic S, Agbaba D (2016) QSAR studies and design of new analogs of vitamin E with enhanced antiproliferative activity on MCF-7 breast cancer cells. J Taiwan Inst Chem Eng 59:33–44
- Ibrahim MT, Uzairu A, Shallangwa GA, Ibrahim A (2018a) In-silico studies of some oxadiazoles derivatives as anti-diabetic compounds. J King Saud Univ Sci:1018–3647

15. Ibrahim MT, Uzairu A, Shallangwa GA, Ibrahim A (2018b) Computational studies of some biscoumarin and biscoumarin thiourea derivatives Asa-glucosidase inhibitors. *J Eng Exact Sci* 4(2):0276–0285
16. 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
17. Klebe G (2000) Recent developments in structure-based drug design. *J Mol Med* 78(5):269–281
18. Liu XH, Qiao L, Zhai ZW, Cai PP, Cantrell CL, Tan CX et al (2019) Novel 4-pyrazole carboxamide derivatives containing flexible chain motif: design, synthesis and antifungal activity. *Pest Manag Sci* 2892-2900:75–11
19. 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
20. Smith RD (1993) U.S. Patent No. 5,189,051. U.S. Patent and Trademark Office, Washington, DC
21. Todeschini R, Consonni V, Maiocchi A (1999) The K Correlation Index: theory development and its applications in chemometrics. *Chemom Intell Lab Syst* 46(1):13–29
22. 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
23. Wikipedia Contributors (2019). Retrieved on 30th July, 2019 from https://en.wikipedia.org/wiki/Sclerotinia_sclerotiorum.
24. Yao TT, Xiao DX, Li ZS, Cheng JL, Fang SW, Du YJ et al (2017) Design, synthesis, and fungicidal evaluation of novel pyrazole-furan and pyrazole-pyrrole carboxamide as succinate dehydrogenase inhibitors. *J Agric Food Chem* 65(26):5397–5403
25. Yap CW (2011) PaDEL-descriptor: an open-source software to calculate molecular descriptors and fingerprints. *J Comput Chem* 32(7):1466–1474

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)
