

RESEARCH

Open Access



QSAR, molecular docking, and design of novel 4-(*N,N*-diarylmethyl amines) Furan-2(5*H*)-one derivatives as insecticides against *Aphis craccivora*

Yusuf Isyaku*, Adamu Uzairu, Sani Uba, Muhammad Tukur Ibrahim and Abdullahi Bello Umar

Abstract

Background: *Aphis craccivora* has many plant hosts, though it seemingly forechoice to groups of bean family. Other plants it hosts are families of Solanaceae, Rosaceae, Malvaceae, Chenopodiaceae, Caryophyllaceae, Ranunculaceae, Cucurbitaceae, Brassicaceae, and Asteraceae.

Result: A computational study was carried out on a series of twenty compounds of novel 4-(*N,N*-diarylmethylamines) furan-2(5*H*)-one derivatives against *Aphis craccivora* insect. Optimization of the compounds was performed with the aid of Spartan 14 software using DFT/B3LYP/6-31G** quantum mechanical method. Using PaDel descriptor software to calculate the descriptors, Generic Function Approximation (GFA) was employed to generate the model. Model 1 found to be the optimal out of four models generated which has the following statistical parameters; $R^2 = 0.871489$, $R^2_{adj} = 0.83644$, cross-validated $R^2 = 0.790821$, and external $R^2 = 0.550768$. Molecular docking study occurred between the compounds and the complex crystal structure of the acetylcholine (protein AChBP) (PDB CODE 2zju) in which compound 13 was identified to have the highest binding energy of $-8.4 \text{ kcalmol}^{-1}$. Statistical analyses, such as variance inflation factor, mean effect, and the applicability domain, were conducted on the model. This compound has a strong affinity with the macromolecular target point of the *A. craccivora* (2zju) producing H-bond and as well the hydrophobic interaction at the target point of amino acid residue. Molecular docking gave an insight into the structure-based design of the new compounds with better activity against *A. craccivora* in which three compounds A, B, and C were designed and discovered to be of high quality and have greater binding affinity compared to the one obtained from the literature.

Conclusion: The QSAR model was generated by the employment of Genetic Function Approximation (GFA). The model was found to be robust and possessed a good statistical parameter. Furthermore, a molecular docking study was performed to get an idea for structure-based design in which three (3) compounds A, B, and C were designed and were found to be more active than the template (compound 13, i.e., the one with highest docking score). QSAR model was developed to give an insight into the ligand/template-based design of computer-aided drug design.

Keywords: QSAR, Molecular docking, Design, Insecticides

* Correspondence: isyakuyusuf01@gmail.com

Department of Chemistry, Ahamadu Bello University, P.M.B, Zaria 1044, Nigeria

Background

Aphis craccivora, known as cowpea aphid, peanut (groundnut) aphid, or black legume aphid, is one of the most dangerous agricultural pests which directly causes harm to plants by delaying and deforming the growth of the plants through malnutrition. The molasses manufactured by the vector are placed on plants and stimulate the growth of molds with soot that limits photosynthesis. *Aphis craccivora* has many plant hosts, though it seemingly forechoice to groups of bean family. Other plants it hosts are families of Solanaceae, Rosaceae, Malvaceae, Chenopodiaceae, Caryophyllaceae, Ranunculaceae, Cucurbitaceae, Brassicaceae, and Asteraceae. Aphid is a vector of a series of plant viruses that include peanut rosette virus, groundnut mottle virus, mosaic virus common bean, alfalfa mosaic virus, and cucumber mosaic virus (Wikipedia, 2018).

Furan- and amide-containing compounds are among the molecular structures found to have an extensively wide range of applications in the field of medicine and agrochemical due to their extensive range of biological activity like antimicrobial/anti-inflammatory activity (Huczynski et al. 2012; Ravindra et al. 2006; Özden et al. 2005), as antibiotic activity (Arjona et al. 1999), as pesticides such antifungal (Yao et al. 2017), and insecticides (Teixeira et al. 2015; Wang et al. 2013) among others. The furan-2(5H)-one was examined to be a potential inhibitor of nicotinic acetylcholine receptor (Tian et al., 2019), and this was the reason for the docking study on the crystal structure of this protein.

The quantitative analysis of the structure-activity relationship (QSAR) is among the most efficient ways to optimize the main compounds and design new compounds. QSAR can be used to predict bioactivities, like toxicity, carcinogenicity, and 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 helps in predicting the new compounds with better activity than the existing ones. This quantum chemical calculation is extensively applied while forming the QSAR models (Gagic et al. 2016). Molecular docking helps to investigate the capacity of the prepared compounds toward the interaction with the protein residue of the target organism and to also predict the preferred orientation of the molecules.

The objective of the research is to discover a new model that predicts the activity of chemical products with better activity capable of destroying *Aphis craccivora* using Genetic Function Approximation (GFA) or molecular docking techniques.

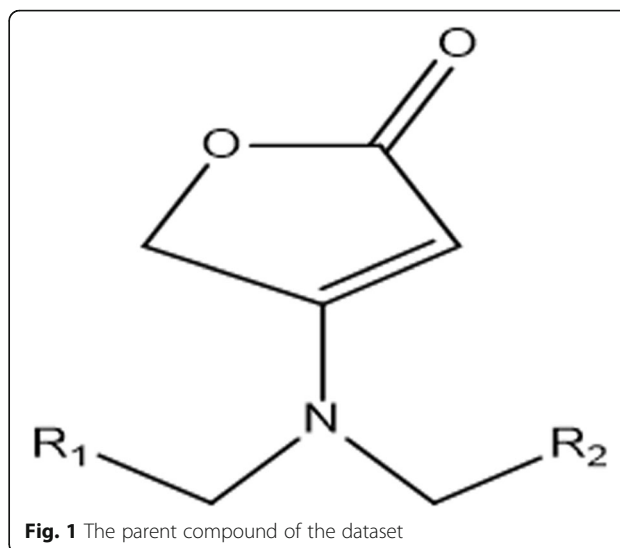


Fig. 1 The parent compound of the dataset

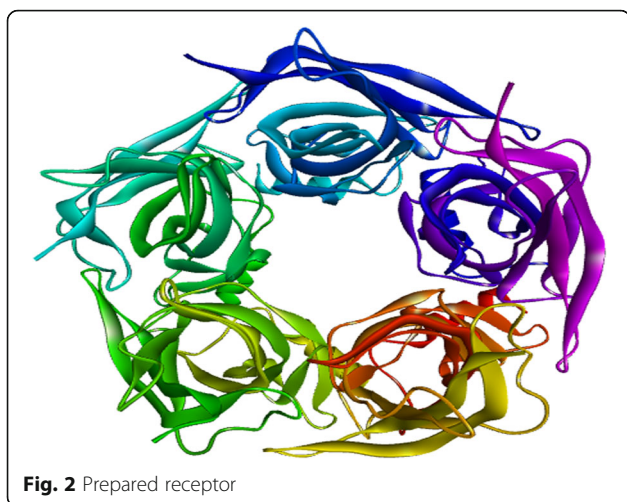
Material and methods

Dataset

In this work, we used a dataset of 20 compounds to design a relation between the chemical traces of compounds and their insecticidal activity. These 20 compounds of novel 4-(*N,N*-diarylmethylamines) furan-2(5H)-one derivatives were obtained from the literature (Tian et al., 2019). The logarithm of the measured LC₅₀ (μg mL⁻¹) against

Table 1 Dataset compounds

S/No	R1	R2	pLC50
1	6-Cl-Pyrid	2-FluoroPhenyl	0.695482
2	6-Cl-Pyrid	3,4-DiFluoroPhenyl	0.848189
3	3-Pyrid	4-FluoroPhenyl	1.558829
4	6-Cl-Pyrid	3-Chloro-4-FluoroPhenyl	0.637490
5	6-Cl-Pyrid	3-CN-4-FluoroPhenyl	0.872739
6	6-Cl-Pyrid	3,4-DiChloroPhenyl	1.146748
7	6-Cl-Pyrid	2-ChloroPhenyl	1.024896
8	6-Cl-Pyrid	2,4,5-TriFluoroPhenyl	1.559787
9	6-Cl-Pyrid	2-Thiophen	1.061075
10	6-Cl-Pyrid	4-FluoroPhenyl	0.725095
11	6-Cl-Pyrid	3-BromoPhenyl	0.735599
12	6-Cl-Pyrid	3-ChloroPhenyl	0.836324
13	6-Cl-Pyrid	4-TriFluoromethy-Phenyl	1.257198
14	6-Cl-Pyrid	4-F-3-NO ₂ -Phenyl	1.296665
15	6-Cl-Pyrid	2,3-DiFluoroPhenyl	1.128399
16	6-Cl-Pyrid	5-Cl-2-FluoroPhenyl	1.492900
17	6-Cl-Pyrid	4-CN-Phenyl	0.825426
18	6-Cl-Pyrid	3-NO ₂ -Phenyl	1.107549
19	6-Cl-Pyrid	4-((3,3-Dichloroallyl) oxy-Phenyl)	1.697317
20	6-Cl-Pyrid	3-Br-4-F-Phenyl	0.235528



insecticidal activity given by $p\text{ LC}_{50}$ ($p\text{ LC}_{50} = -\log 1/\text{LC}_{50}$) was taken as a dependent parameter; therefore, the data was linearly correlated with the independent parameter/descriptors (Edache et al. 2017).

Optimization/molecular descriptor calculation

The database (see Fig. 1 and Table 1) was optimized at a density function theory level using the “Becke’s three-parameter read-Yang-Parr hybrid” (B3LYP) function together with “6-31G**” basis set of Spartan14 software (Arthur et al. 2016a). Graphical-user-interface of Spartan14 was utilized in drawing 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 calculate the quantum molecular descriptors (Yap 2011).

Data division

To get a validated model, the dataset was split into (3:1) train test sets. Accordingly, the split was done in such a way that the compounds forming the train (70% of the data) and the test sets (30% of the data) are shared

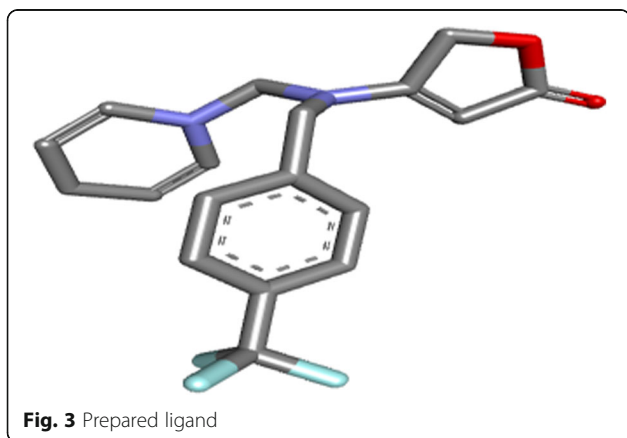


Table 2 Validation parameter of the model

Friedman LOF	0.064842
R-squared	0.871489
Adj R-squared	0.83644
Cv R-squared	0.790821
Significant regression	Yes
Significance of regression (<i>F</i> value)	24.86516
Critical SOR <i>F</i> value (95%)	3.748716
Replicate points	0
Computed experimental error	0
LOF points	11
Min expt. error for non-significant LOF (95%)	0.090103

within an entire descriptive space filled by the complete dataset as described by Kennard-Stone Algorithm method (Arthur et al. 2016b).

The generated molecular descriptors were taken for regression analysis, with experimental activities as dependent parameters where the molecular descriptors served as independent parameters. With the Genetic Function Approximation method (GFA) incorporated in “Material Studio 2017” software, the compounds of train sets were utilized to develop the QSAR model. Four QSAR models were built where 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 = number of terms a model possessed in addition to the constant term, M is equal to the number of samples present in the training set, and p = overall number of descriptors present in all terms of the model excluding the constant term (Edache et al. 2015).

Internal validation

The generated model was validated internally by the following parameters:

(a) The correlation coefficient (R^2): explain the division of overall variation ascribed to the built model. The

Table 3 Descriptors and their description

S/no	Descriptor	Description	Class
1	nCl	Number of chlorine atoms	2D
2	ATSc4	ATS autocorrelation descriptor, weighted by charges	2D
3	Weta3,polar	Directional WHIM, weighted by atomic polarizabilities	3D

Table 4 External validation

Name	pLC50	nCl	ATSc4	Weta3.polar	Ypred	(Y _{pred} - Y _{obs})
3	1.558828525	1	- 0.1014418	0.293087860	0.952953889	- 0.60587464
8	1.559786968	1	- 0.1615480	0.274664020	1.382496168	- 0.1772908
12	1.29666519	1	- 0.1538037	0.385254528	0.71496466	- 0.58170053
15	1.128399269	1	- 0.1126655	0.253145528	1.239656943	0.111257674
19	1.697316542	3	- 0.2024771	0.325435894	1.840893036	0.143576494

accepted value of R^2 ranges from 0.5 to <1 and more the value of R^2 and the model considered to be a better model as R^2 approaches 1.0, though there are other analyses that the model passed to be a better one. 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 (Adeniji et al. 2018).

(b) 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 equation 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 constituted the model,” while n = number of training set compounds (Ibrahim et al. 2018a).

(c) 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; the predicted value of the eliminated data point is compared to its real value. This is repeated until each data removed. We can then calculate the value of Q^2 using the sum of the squares of these elimination residues as in the equation below:

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

where Y_{expt} , Y_{predt} , and \bar{Y}_{train} represent the experimental, predictive, and average activities of the training set (Adedirin et al. 2018).

External validation

The prediction ability 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:

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

where Y_{predt} and Y_{expt} are the test set's experimental and predicted activities while \bar{Y}_{train} indicates the average activities of the training set (Edache et al., 2017).

Statistical analysis of the descriptors

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.

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

where R^2 gives multiple correlation coefficient between the variables within the model. If the VIF is equal to 1, it

Table 5 Continuation of external validation

(Y _{pred} - Y _{obs}) ²	Y _{meanTrain}	Y _{obs} - Y _{meanTr}	(Y _{pred} - Y _{MeanTrain}) ²
0.367084074	0.900149221	0.65867930	0.433858425
0.031432028	0.900149221	0.65963775	0.435121957
0.338375507	0.900149221	0.39651597	0.157224914
0.01237827	0.900149221	0.22825005	0.052098084
0.02061421	0.900149221	0.797167321	0.635475737
Σ(Y _{pred} - Y _{obs}) ² = 0.769884088			Σ(Y _{pred} - Y _{MeanTrain}) ² = 1.713779118
$R^2 = (1 - 0.7698841/1.713779) = 0.550768$			

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 (Pourbasheer et al. 2015; Karthikeyan et al. 2009).

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 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:

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

where B_j and D_j are the j -descriptor coefficient in a 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 compounds in the training set. To evaluate the significance of the model, the ME of all the descriptors was calculated (Edache et al. 2015).

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 its domain of applicability. It aimed 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 train 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 are given as follows:

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

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

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

where n = number of compounds in the training set whereas p gives the number of descriptors present in the model (Ibrahim et al. 2018a).

Table 6 The experimental, predictive, and residual activity

Name	pLC50	predicted pLC50	Residual
1	0.695482	0.733838	− 0.038360
2	0.848189	0.751294	0.096895
3	1.558828	0.952954	− 0.605870
4	0.637490	0.651660	− 0.014170
5	0.872739	0.815291	0.057448
6	1.146748	1.123891	0.022857
7	1.024896	1.094485	− 0.069590
8	1.559787	1.382496	− 0.177290
9	1.061075	1.032018	0.029057
10	0.725095	0.573817	0.151278
11	0.735599	0.622145	0.113453
12	0.836324	0.797338	0.038986
13	1.257198	1.254363	0.002835
14	1.296665	0.714965	− 0.581700
15	1.128399	1.239657	0.111258
16	1.492900	1.493841	− 9.41E−04
17	0.825426	0.992587	− 0.16716
18	1.107549	1.061123	0.046426
19	1.6973165	1.8408930	0.143576
20	0.2355280	0.5045470	− 0.269020

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 preparation where all the residues associated with it such as a ligand, water molecules, and other traces associated with the receptor were removed. The ligands (the optimized compounds) which were in the SDF file were transformed into the PDB file format. Figures 2 and 3 showed the prepared receptor and ligand (Ibrahim et al. 2018b).

Results

QSAR model

Genetic function algorithm (GFA) was used to generate the three QSAR models which predicted the activity of the compounds. The first model was chosen as the optimal model due to its statistical significance was presented in equation (x) below:

Table 7 Pearson's correlation

	nCl	ATSc4	Weta3.polar	VIF
nCl	1			1.064331
ATSc4	0.12939	1		1.043468
Weta3.polar	0.19040	−0.1303	1	1.064594

Table 8 Standard regression coefficients “*b_j*”, the values of mean effect (ME) and confidence interval (*p* values)

Descriptors	Standard regression coefficient (<i>b_j</i>)	Mean effect (MF)	<i>P</i> value (confidence interval)
nCI	0.262046	0.40756	0.000416
ATSc4	− 5.41238	0.666756	0.000535
Weta3.polar	− 5.65705	− 2.07385	1.87E−05

$$Y = 0.262046205 \cdot \text{nCI} - 5.412377232 \cdot \text{ATSc4} - 5.657053582 \cdot \text{weta3.polar} + 1.799880140 \quad (x)$$

All the validation/statistical parameters that signified the stability, robustness, and the prediction capability of the model were presented in Table 2.

The name, symbol, and class of the three selected descriptors that made up the model are presented in Table 3.

Since the selected model is internally valid, then an external validation is the next step. Tables 4 and 5 represent the external validation as well as the calculation of predicted R^2 of the best-chosen model.

The experimental, predictive, and residual activity for both training sets and test sets are shown in Table 6. This table is represented to show the robustness of the model considering the lower residual values.

Statistical analyses

Statistical analyses on the model's descriptors are very necessary in order to know how related they are. For that, Pearson's correlation, variance inflation factor, mean effect (which contains regression analysis), and applicability domain were carried out.

Pearson's correlation (Tables 7 and 8)

Figures 4 and 5 are the plot of predicted activity against the experimental activity (pLC₅₀) and plot of standardized residual against the experimental activity (pLC₅₀).

Applicability domain

A graph of leverages of each compound of dataset versus their standardized residuals terms William's plot was presented in the Fig. 6 below.

Docking studies

Molecular docking analysis was carried out between the ligands (compounds) and the receptor to evaluate the binding affinity at the ligand-receptor interface.

Result of the design

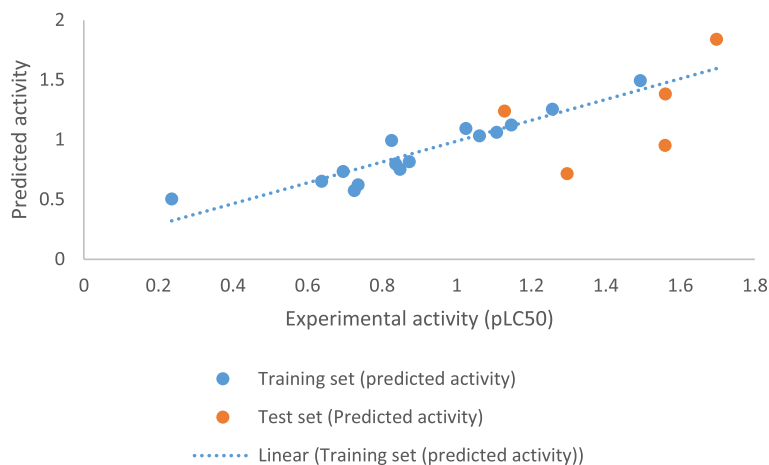
The structure of the three (3) compounds which were designed using an optimization method of structure-based design. The structure of the chosen scaffold (compound 13) was presented in Fig. 9 below.

Discussion

QSAR model

The QSAR examination was carried out to relate the structure-activity relationship of novel 4-(*N,N*-diaryl-methylamines) furan-2(5*H*)-one derivatives as a potent inhibitor of *Aphis craccivora*. Three descriptors were utilized in constructing the QSAR model which predicted the activity of the compounds based on the genetic function algorithm (GFA). The first model was chosen as the optimal model due to its statistical significance. The best-chosen model constructed was presented in equation.

All the validation/statistical parameters that signified the stability, robustness, and the prediction capability of

**Fig. 4** A plot of predicted activity against experimental activity (pLC₅₀)

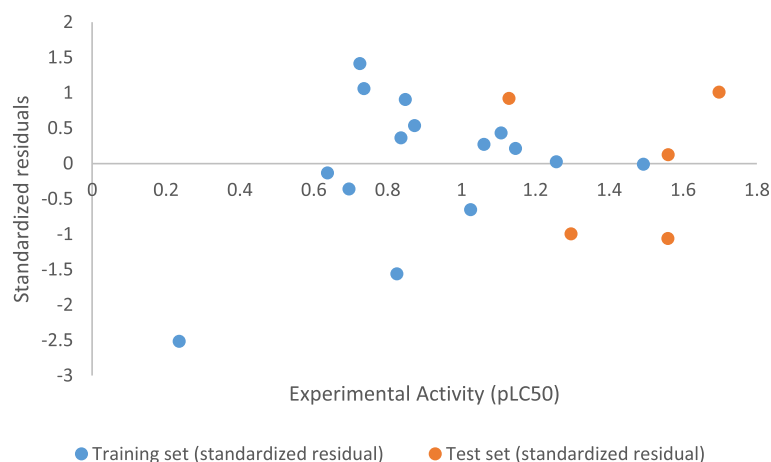


Fig. 5 A plot of standardized residual against experimental activity (pLC₅₀)

the model were presented in Table 2. From the table, the highly calculated R^2 value (0.8715) for the predicted activity indicated the robustness of the model. The descriptors, definitions, and their classes are represented in Table 3 below. The fact that 2D and 3D descriptors are present in the model implies that the descriptors used in the model can determine a better insecticidal activity of the compounds. The individual capability and inducing power of the selected descriptors toward the activity of the compounds depend on their values, signs, and as well their mean effects. Tables 4 and 5 represent the external validation as well as the calculation of predicted R^2 of the best-chosen model.

Statistical analysis

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

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

Table 8 showed the standard regression coefficients " b_j ," the values of mean effect (ME), 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, signs, and as well their mean effects. The p values of the three

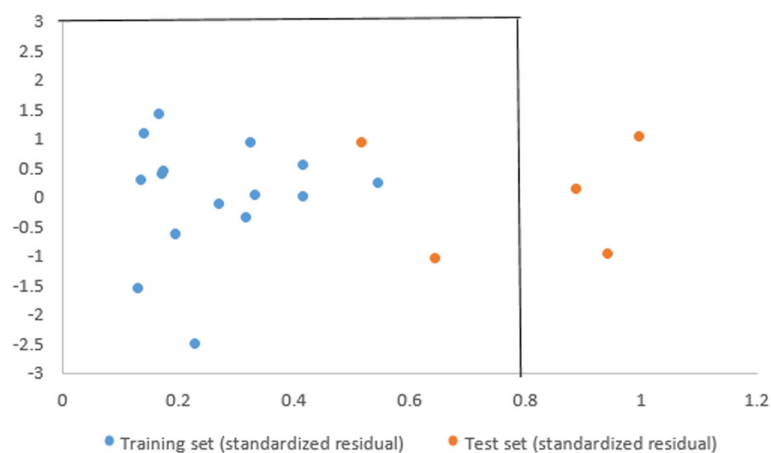


Fig. 6 William's plot

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

Ligands	Binding affinity	Hydrogen bond		Hydrophobic interaction
		Amino acid	Bond distance (Å)	
1	− 7.5	TYR164	3.40384	TYR89, LYS139, VAL183
3	− 7.5	ARG137, ARG137, LYS139	2.36455, 2.46429, 3.61889	TYR89
4	− 7.9	ARG137, ARG137, TYR164	2.77808, 2.62344, 3.57359	LYS139, VAL183
5	− 8.2	ALA88, GLN119	2.29854, 2.63975	VAL83, ALA87, PRO95, PRO100
6	− 7.9	ARG137, ARG137, SER122, TYR164	2.89161, 2.67782, 3.65143, 3.56844	VAL183, LYS139, VAL183, TRP53, TYR164, TYR89, TYR185
7	− 7.5	SER166	3.67404	LYS139, VAL183, TYR164, TYR89
13	− 8.4	ARG137	2.60716	TYR89, TYR89, ASN90, VAL183, TRP53, TYR89, TYR185

descriptors (at 95% c.l.) that made up the model are all < 0.05; this implies that there is a significant relationship among these descriptors (as contrary to the null hypothesis) and the inhibitory concentration of the compounds.

Figure 5 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.

A graph of leverages of each compound of dataset versus their standardized residual terms William's plot was plotted to discover the outliers as well as the chemical influential values of the model. The domain of applicability was established within a box at ± 3.0 limit for the residuals and leverage threshold h^* (where h^* calculated to be 0.8). The result indicates that except three compounds from the test set, all the molecules in the dataset are within the box of the applicability domain of the model. This may be characterized by their clear differences in chemical structures by considering the rest of the compounds highlighted in the dataset.

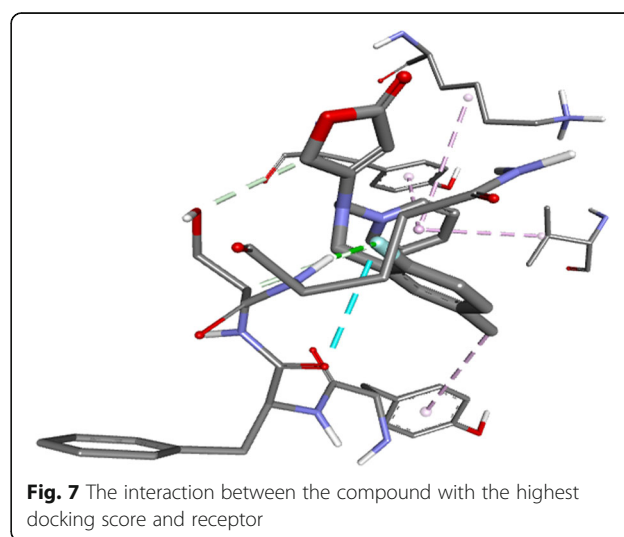
Molecular docking study

Due to unavailability of the crystal structure of *Aphis craccivora*, the complex crystal structure of the acetylcholine (protein AChBP) from *Lymnaea stagnalis* and imidacloprid with PDB code 2ZJU was utilized for the docking analysis since it possess "high homology extracellular domain" of the *A. craccivora* protein (nAChR) which has been used for many docking studies involving *A. craccivora*. The docking studies were performed between 2ZJU and the ligands (compounds) of novel 4-(*N*, *N*-diarylmethylamines) furan-2(5*H*)-one derivatives to investigate the binding energy of the compounds to the target site of the insect. The ligands show a good interaction with the active site of the *Aphis craccivora* that is to say they inhibit the activity of the insect. Some ligands show high binding energy that varies from − 7.9 to − 8.4 kcalmol^{−1} as presented in Table 9. However, compound 13 with the highest binding score (− 8.4 kcal/mol) possessed an interaction mode with H-bond of ARG137 and 2.60716 bond length and hydrophobic interaction of

TYR89, TYR89, ASN90, VAL183, TRP53, TYR89, and TYR185. The interaction between the compound with highest binding energy and the binding pocket of the receptor is shown in Fig. 7 while Figs. 8 and 9 is the 2D hydrogen bond interaction of compound 13 with the receptor.

Design

In our research, we utilized the method of structure-based design to design a new (novel) insecticidal compound with a better activity by taking the compound with the highest docking score which is compound 13 (with binding energy of − 8.4 kcal/mol) as our template compound and thus provides suitable insecticidal activity and appeared very inspiring as a noteful scaffold. Compound 13 was selected as a synthetically best structure in which some structural advancement was performed on it. The newly designed compounds (A) 4-(((2-chloro-4-(trichloromethyl)pyridine-1(2*H*)-yl)methyl)(2-chloro-4-(trifluoromethyl)benzyl)amino)furan-2(5*H*)-one, (B) 4-(((2-chloro-4-(trichloromethyl)pyridine-1(2*H*)-yl)methyl)(3-chloro-4-(trifluoromethyl)benzyl)amino)furan-2-(5*H*)-one, and (C) 4-(((2-chloro-4-(2,2,2-trichloroethyl)pyridin-1(2*H*)-yl)methyl)(4-(trifluoromethyl)benzyl)amino)furan-2(5*H*)-one



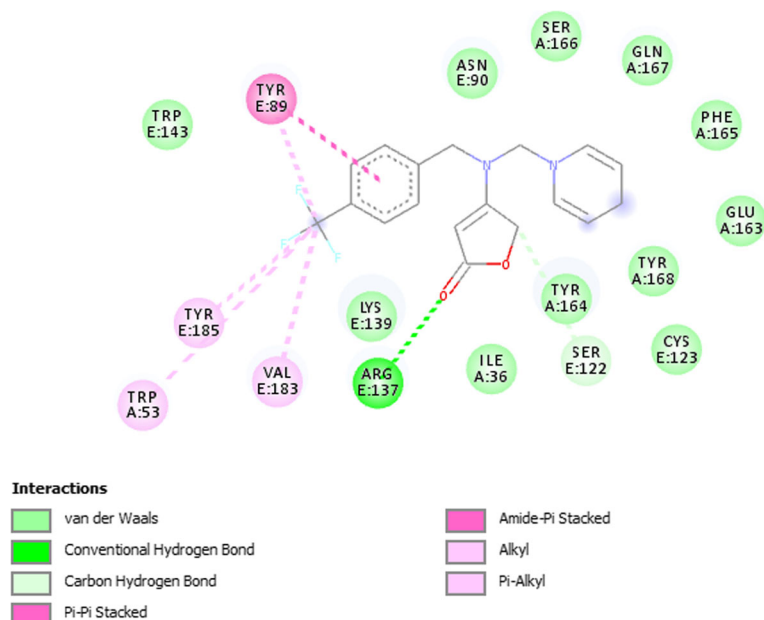


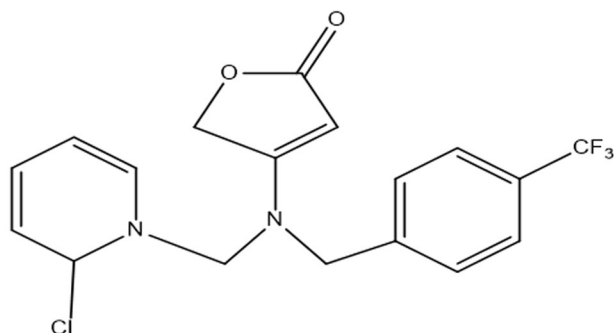
Fig. 8 2D interaction of compound 13

with their binding energy of -8.9 , -9.1 , and -9.0 kcal/mol (as shown in Figs. 10, 11, and 12) were discovered to be of high quality and have greater binding affinity compared to the one obtained from the literature.

Conclusion

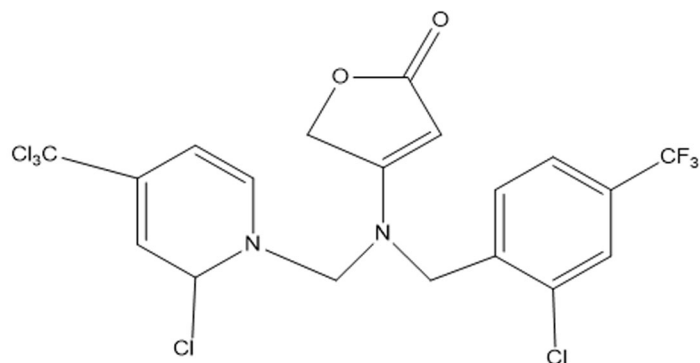
This research involves a QSAR and molecular docking studies on 20 compounds of novel 4-(*N,N*-diarylmethylamines) furan-2(5*H*)-one derivatives against *Aphis craccivora*. Using DFT for molecule optimization, Genetic Function Approximation (GFA) was employed in generating the built model. Out of three models built, the first model was identified to be the optimal constituted with good statistical parameters such as $R^2 = 0.871489$, $R^2_{adj} = 0.83644$, cross-validated $R^2 = 0.790821$, and external $R^2 = 0.550768$. A decrease in negative coefficient descriptors

(like **ATSc4** and **Weta3.polar**) and an increase in positive coefficients descriptors (like **nCl**) will improve the activities of the compounds against *A. craccivora*. According to the docking scores, most of the ligands (compounds) show good inhibitory activity against *A. craccivora* protein. However, ligands 13 showed a higher binding affinity of -8.4 kcal/mol. This compound has a strong affinity with the macromolecular target point of the *A. craccivora* (2zju) producing H-bond and as well the hydrophobic interaction at the target point of amino acid residue. Molecular docking gave an insight into the structure-based design of the new compounds with better activity against *A. craccivora* in which three compounds A, B, and C were designed and discovered to be of high quality and have greater binding affinity compared to the one obtained from the literature.



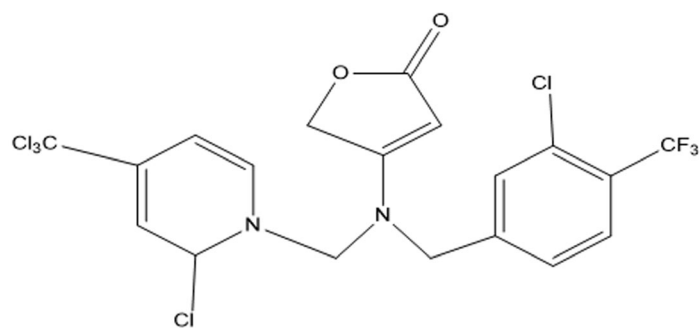
4-(((2-chloropyridin-1(2H)-yl)methyl)(4-(trifluoromethyl)benzyl)amino)furan-2(5H)-one

Fig. 9 2D structure of the template



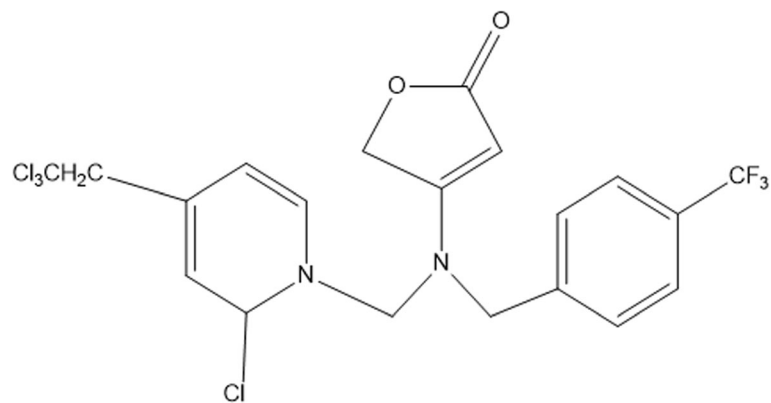
4-(((2-chloro-4-(trichloromethyl)pyridin-1(2H)-yl)methyl)(2-chloro-4-(trifluoromethyl)benzyl)amino)furan-2(5H)-one

Fig. 10 2D structure of the designed compound (A)



4-(((2-chloro-4-(trichloromethyl)pyridin-1(2H)-yl)methyl)(3-chloro-4-(trifluoromethyl)benzyl)amino)furan-2(5H)-one

Fig. 11 2D structure of the designed compound (B)



4-(((2-chloro-4-(2,2,2-trichloroethyl)pyridin-1(2H)-yl)methyl)(4-(trifluoromethyl)benzyl)amino)furan-2(5H)-one

Fig. 12 2D structure of the designed compound (C)

Abbreviations

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

Author's contributions

YI: contributed throughout the research work. AU: gives directives and technical advices. SU: partake in technical activities. MTI: partake in technical activities. ABU: partake in technical activities. All authors have read and approved the final manuscript.

Funding

No fund was collected on this research work.

Availability of data and materials

The section is not applicable to this research work

Ethics approval and consent to participate

The section is not applicable to this research work

Consent for publication

The section is not applicable to this research work

Competing interests

The authors declared no competing interest.

Received: 30 January 2020 Accepted: 6 March 2020

Published online: 26 March 2020

References

- 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*
- Arjona O, Iradier F, Medel R, Plumet J (1999) Enantioselective synthesis of antibiotic (+)-rancinamycin III derivative and two protected carbasugars of the α -D-talo-series from furan. *Tetrahedron Asymmetry* 10(17):3431–3442
- Arthur DE, Uzairu A, Mamza P, Abechi E, Shallangwa G (2016a) QSAR modelling of some anticancer PGI50 activity on HL-60 cell lines. *Albanian J Pharmaceut Sci* 3(1):4–9
- Arthur DE, Uzairu A, Mamza P, Abechi S (2016b) Quantitative structure–activity relationship study on potent anticancer compounds against MOLT-4 and P388 leukemia cell lines. *J Adv Res* 7(5):823–837
- Edache EI, Arthur DE, Abdulfatai U (2017). Tyrosine activity of some tetraketone and benzyl-benzoate derivatives based on genetic algorithm-multiple linear regression.
- Edache EI, Uzairu A, Abeche SE (2015) Investigation of 5, 6-dihydro-2-pyrones derivatives as potent anti-HIV agents inhibitors. *J Comput Methods Mol Des.* 5(3):135–149
- 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 Instit Chem Eng* 59:33–44
- Huczyski A, Janczak J, Stefańska J, Antoszczak M, Brzezinski B (2012) Synthesis and antimicrobial activity of amide derivatives of polyether antibiotic—salinomycin. *Bioorg Med Chem Lett* 22(14):4697–4702
- Ibrahim MT, Uzairu A, Shallangwa GA, Ibrahim A (2018a) Computational studies of some biscoumarin and biscoumarin thiourea derivatives ASa-glucosidase inhibitors. *J Eng Exact Sci* 4(2):0276–0285
- Ibrahim MT, Uzairu A, Shallangwa GA, Ibrahim A (2018b) In-silico studies of some oxadiazoles derivatives as anti-diabetic compounds. *Journal of King Saud University-Science*
- 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
- Özden S, Atabey D, Yıldız S, Göker H (2005) Synthesis and potent antimicrobial activity of some novel methyl or ethyl 1H-benzimidazole-5-carboxylates derivatives carrying amide or amidine groups. *Bioorg Med Chem* 13(5):1587–1597

- Pourbasheer E, Aalizadeh R, Ganjali MR, Norouzi P (2015) QSAR study of IKK β inhibitors by the genetic algorithm: multiple linear regressions. *Med Chem Res* 23(1):57–66
- Ravindra et al. (2006) Synthesized a series of 3-acetyl-5-naphthol[2,1-b]furan-2-yl-2-aryl-2,3-dihydro-1,3,4-oxadiazoles. The synthesized compounds showed very good anti-inflammatory activity [61].
- Teixeira MG, Alvarenga ES, Pimentel MF, Picanço MC (2015) Synthesis and insecticidal activity of lactones derived from furan-2 (5H)-one. *J Braz Chem Soc* 26(11):2279–2289
- Tian P, Liu D, Liu Z, Shi J, He W, Qi P, Song B (2019) Design, synthesis, and insecticidal activity evaluation of novel 4-(N,N-diarylmethylamines) furan-2 (5H)-one derivatives as potential acetylcholine receptor insecticides. *Pest Manag Sci* 75(2):427–437
- Wang BL, Zhu HW, Ma Y, Xiong LX, Li YQ, Zhao Y, Zhang JF, Chen YW, Zhou S, Li ZM (2013) Synthesis, insecticidal activities, and SAR studies of novel pyridylpyrazole acid derivatives based on amide bridge modification of anthranilic diamide insecticides. *J Agric Food Chem* 61(23):5483–5493
- Wikipedia contributors (2018). Retrieved August 6, 2019, from https://en.wikipedia.org/wiki/Aphis_craccivora#cite_note-ITIS-2
- Yao TT, Xiao DX, Li ZS, Cheng JL, Fang SW, Du YJ, Zhao JH, Dong XW, Zhu GN (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
- 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)