QSAR STUDY OF 2-SUBSTITUTED PHENYL-2-OXO-, 2-HYDROXYL-AND 2-ACYLLOXYETHYLSULFONAMIDES AS FUNGICIDES

Y. ISYAKU¹,*, A. UZAIRU² and S. UBA³
¹ Ahmadu Bello University, Department of Chemistry, Zaria, Kaduna State, Nigeria

**A R T I C L E  I N F O**

Article history:
Received 2019-04-28
Accepted 2019-06-28
Available online 2019-06-30

**k e y w o r d s**
QSAR
2-substituted
Sulfonamides
Fungicides

**A B S T R A C T**

An insilico study was carried out on a series of thirty-five (35) sulfonyl-containing compounds for their antifungal activities against Botrytis Cinerea fungi using QSAR techniques. Using Spartan 14 molecular modelling software to draw the molecular structure of the compounds, the DFT/B3LYP/6-31G* quantum method of the software was used in optimizing the drawn compounds. The optimized compounds of the dataset were then underbring into PaDEL-Descriptor software for their molecular descriptors calculation. The calculated PaDel-descriptors were then subjected to data-Pretreatment and later splitted into 70% training set and 30% test set. The model was generated using the training set and the test set for the validation of the model built. Using Genetic Function Algorithm (GFA) the model was developed. Four models were developed in which model 1 was chosen as the optimum model with good statistical parameters; $R^2 = 0.954$, $R^2_{adj} = 0.941$, cross validation $R^2$, $Q^2_{cv} = 0.888$ and $R^2_{pred} = 0.839$. The model proposed was found to be stable, robust and showed a good internal and external validation. Other statistical analysis such as mean effect, variance inflation factor (VIF), Williams plot among others were also carried out for the applicability domain of the model.
1. INTRODUCTION

Sulfonyl-containing compounds are one of the major active compounds categorized to have a vast range of biological activity. They are extensively used in drugs and agrochemicals. Botrytis Cinerea (scientifically known as Gray mold, Flower capsule blight, Botrytis brown stain, Scape blight or Bunch not, is a plant pathogen that infects over 200 plant species, causing grey mould which lead to serious economic losses of $10billion to $100billion annually (Boddy, Lynne, 2016). The plant species found to be affected by B. Cinerea includes tomatoes, Lettuce, Grapes, Strawberries e.t.c causing a grey powdery mould on the affected plants. It has the ability to counteract a large range of plant defence chemicals. It is one of the most extensively studied necrotrrophic plant pathogens. B. Cinerea has some other relatives like B. byssoides, B. allii and B. squamosal which infect onions, B. Tulipae which infect Saffron and tulips, B. Fabae which affect beans and B. gladioli which infect gladioli and lilies. Due to development of several strains on many commercial antifugal compounds, the need of developing new antifungal with novel mode of action arised in order to specifically fight the resistance of these organism instead of the general fungicide. And also, to ensure the activity of the fungicides donot affect the beneficial organisms in the environment.

Sulfonyl-containing compounds excises a vital role in the field of agrochemicals as well as medicine. The first drugs that found to have selectivity on bacterial activity that could systematically be used to inhibit a specific bacterial infection was sulfonamides. Due to this great success, a considerable greater attention has been paid to develop more sulfonyl-containing compounds as agrochemicals and drugs. Some sulfonamides fungicides such as cyazofamid, tolnifamide and amisulbrom are commercially used. Sulfonyl-containing compounds such as sulfonamides drugs are used as anti-tumour (Huang et al., 2001). Also, sulfonamides were found to have anti-plasmodial activity (Fisher et al., 2017). Sulfonyl-containing compounds such as sulfonamides are extensively used in pharmaceutical industries as antiancer, anti-inflammatory and antiviral agents. There are over 30 drugs containing this functionality of sulfonamides that are clinically used. This includes, antibacterial, anticonvulsant, diuretics, hypoglycemic and HIV protease inhibitors.

Due to pathogenic activity of Botrytis Cinerea organisms on both plants and animals and an extensively wide range of antibacterial activity by sulfonyl-containing compounds, many researches are carried out to fight the existence and pathogenic activity of the fungi. Some of these researches are computational studies such as QSAR study. QSAR is a mathematical model which link the structure-derived characters of given compounds to their inhibitory activities. The studies of QSAR are intended at formulating a model (correlation models) using the activity and other informations from the chemical in the data in a statistical approch (Roy et al., 2015). QSAR studies were carried out to predict more active compounds that will inhibit the activity of fungal diseases (Saiz-Urra et al., 2009); (Singla et al., 2009).

The aims of the paper are to develop a model (QSAR) which can predict a better activities of sulfonyl-containing compounds against Botrytis Cinerea fungi.

2. MATERIALS AND METHOD

2.1 Data

Thirty-five (35) derivatives of 2-substituted phenyl-2-oxo-, 2-Hydroxy- and 2-Acloyoxethylsulfonamides used in the research are found in literature (Wang et al., 2017). The activities of these compounds were reported in EC$_{50}$ (mg/L), which were converted to pEC$_{50}$ (pEC$_{50}$ = -log1/ EC$_{50}$). The activity values and their corresponding molecular structure found in the date set are presented in the table 1 below.

3. TABLES 1- Compounds and pEC50 values

![Figure 2- N-(2-trifluoromethyl-4-chlorophenyl)-2-substituted-2-hydroxy-sulfonamides](image)

| Serial No. | Compound Structure | pEC50 |
|-----------|-------------------|-------|
| 1         | ![Structure 1](image) | -0.888741 |
| 2         | ![Structure 2](image) | -0.930949 |
| 3         | ![Structure 3](image) | -0.541579 |
| 4         | ![Structure 4](image) | -0.838219 |
| 5         | ![Structure 5](image) | -0.834421 |
| 6         | ![Structure 6](image) | -0.732394 |
| 7         | ![Structure 7](image) | -0.858537 |
| 8         | ![Structure 8](image) | -0.619093 |
| 9         | ![Structure 9](image) | -0.903090 |
| 10        | ![Structure 10](image) | -1.247973 |
| 11        | ![Structure 11](image) | -1.210853 |
| 12        | ![Structure 12](image) |       |
| 13        | ![Structure 13](image) |       |
| 14        | ![Structure 14](image) |       |
| 15        | ![Structure 15](image) |       |
| 16        | ![Structure 16](image) |       |
| 17        | ![Structure 17](image) |       |
| 18        | ![Structure 18](image) |       |
| 19        | ![Structure 19](image) |       |
| 20        | ![Structure 20](image) |       |
| 21        | ![Structure 21](image) |       |
| 22        | ![Structure 22](image) |       |
| 23        | ![Structure 23](image) |       |
| 24        | ![Structure 24](image) |       |
| 25        | ![Structure 25](image) |       |
| 26        | ![Structure 26](image) |       |
| 27        | ![Structure 27](image) |       |
| 28        | ![Structure 28](image) |       |
| 29        | ![Structure 29](image) |       |
| 30        | ![Structure 30](image) |       |
| 31        | ![Structure 31](image) |       |
| 32        | ![Structure 32](image) |       |
| 33        | ![Structure 33](image) |       |
| 34        | ![Structure 34](image) |       |
| 35        | ![Structure 35](image) |       |
The molecular structures were drawn in the graphical user interface of the Spartan 14 software (wavefunction, Inc. 2013) was employed to calculate 1D, 2D and 3D descriptors. These are the properties of the molecule in the training set that were calculated.

| Serial No. | R     | pEC50  |
|------------|-------|--------|
| 17         | H     | 1.520615 |
| 18         | 3-F   | 0.855519 |
| 19         | 4-NO₂ | 0.609594 |
| 20         | 3,4-F₂ | 0.396199 |
| 21         | 3,5-F₂ | 0.21467  |

For model to be externally validated, we calculate the cross-validation coefficient (Q^cv) where Y represents the predicted activity, D is the number of training set molecules, \( \bar{Y} \) is the average observed activity for the training set, \( \bar{Y}_{\text{pred}} \) is the average observed activity for the test set, \( M \) is the number of descriptors, \( \beta \) is the regression coefficient for the constant term, \( \gamma \) is the regression coefficient for the variable term, and \( c \) is the regression constant (Ibrahim et al., 2016). The regression model is in the form:

\[
Y = \beta_0 + \beta_1 D + \beta_2 Y + \cdots + \beta_M X_M + c
\]

The cross-validation coefficient (Q^cv) is calculated as:

\[
Q^cv = 1 - \frac{\sum (Y - \bar{Y}_{\text{pred}})^2}{\sum (Y - \bar{Y})^2}
\]

For a suitable and trustworthy model, \( Q^cv \) should have an external validation coefficient (Q^test) which describe the fitness of the model. The division is at 70% training set and 30% test set. The entire dataset was splitted into training set and test set 3:1 by Kennard Stone Algorithm. The training set are molecules that partakes in model building whereas the test set molecules are similar to that of training set structures. The unused data set and they are used to externally validate the models through was carried out by the employment of Genetic Function Approximation (GFA) procedure. The algorithms used in the training set for model generation of the built models through was carried out by the employment of Genetic Function Approximation (GFA) procedure. The internal and external validation of the model was carried out by calculating the cross-validation coefficient (Q^cv)

\[
Q^cv = 1 - \frac{\sum (Y_{\text{exp}} - Y_{\text{pred}})^2}{\sum (Y_{\text{exp}} - \bar{Y})^2}
\]

The models generated were internally validated while external validation was done through the test set. The nature of the chemicals used in the training set for model generation was influenced the predictive capacity of the built models. OECD principle 2007 using the test set. The nature of the chemicals used in the training set for model generation was validated the predictive capacity of the built models.
Mean effect is defined by the following:
\[ \text{Mean effect} = \frac{B_j}{\sum B_j} D_j \]  
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 \( n \) stands for the number of molecular descriptors as well as number of molecules in a training set. To evaluate the significance of the model, the mean effect of each descriptor was calculated (Edache et al., 2015).

Applicability Domain
Willian’s plot was employed to examine the outliers and of course the swayful (influential) compounds and also to assert positively the robustness and confidence of the generated model. The Willian’s plot was plotted using standardized residuals against the Leverage. In order to evaluate the model’s applicability domain, the approach of leverage was employed. For a given chemical compounds, leverage is given by the following equation;
\[ h_i = X_i (X^T X)^{-1} X_i^T \]  
where \( h_i \) is the leverage of each compound, \( X_i \) is the training set compounds of the matrix i. \( X \) is the matrix of nxn descriptor in the training set molecules. \( X^T \) is the transpose of \( X \)-matrix.

The warning leverage (\( h^* \)) defined as a boundary of normal values of an outlier X and is given by;
\[ h^* = 3 \left( \frac{d+3}{m} \right) \]  
The variable \( m \) stands for number of molecules in the training set and \( d \) is the descriptors describing the model.

2. RESULTS AND DISCUSSION

Descriptors calculation
QSAR was carried out to formulate a model which relate the structure of thirty-five (35) of sulfonyl-containing compounds (2-substituted phenyl-2-oxo-, 2-hyderoxy- and 2-acyloxyethylsulfonamides) with the respective activities to inhibit B. Cinerea fungi.

After optimizing the compounds in the dataset using Spartan 14 software, 32 quantum chemical descriptors were generated. These 32 descriptors were then combined with 1875 other descriptors obtained from PaDEL descriptor software giving a sum of 1907 descriptors.

Data Division
By employing Kennard-Stone method, the data was divided into training set 70% and test set 30% using the software “Data Division GUI 1.2”.

Model and its Validation
Five descriptors were used in generating the model through the employment of Genetic Function Approximation (GFA) available in Material Studio Software. The equation \( pEC_{50} \) below represent the best model with its statistically validation parameters.
\[ pEC_{50} = Y = 13.308368320 * FMF - 0.338596475 * RNCS - 0.012982836 * TPSA - 2.054638915 * WD.unity + 0.112869857 * Wgamma2.volume - 0.440310924 \]  
The validation parameters shown in the table (2) below i.e. the highly calculated \( R^2 \) values (0.954), \( R^2_{adj} \) value (0.941) and \( R^2_{cv} \) value (0.888) of the selected model indicates that the model possess the acceptability criteria.

Table 2 Validation parameter

| Validation parameter | Model | QSAR Standard |
|----------------------|-------|---------------|
| Friedman LOF         | 0.20012400 | -             |
| R-squared            | 0.95394800 | \( \geq 6 \) |
| Adjusted R-squared   | 0.94115600 | -             |
| Cross validated R-squared | 0.88762300 | \( \geq 5 \) |
| Significant Regression | Yes  | -             |
| Significant-of-regression F-value | 74.57284200 | -             |
| Critical SOR F-value (95%) | 2.79410900 | -             |
| Replicate points     | 0     | -             |
| Computed experimental error | 0.00000000 | -             |
| Lack-of-fit points   | 18    | -             |
| Min. exp. error for non-significant LOF (95%) | 0.16374300 | -             |

From the results of internal validation and that of the external validation [where the R-squares are 0.954 (internal) and 0.839 (external)] indicates a strong relationship between the observed and predicted activities. Additionally, the descriptors possesses of positive coefficient in the best chosen model ‘1’ such as FMF (Complexity of a molecule) and Wgamma2.volume (Directional WHIM, weighted by Van der Waal’s volumes) are to increase the inhibition activities of these compounds against B. Cinerea fungi while the negative once that is RNCS (Relative Negative Charge Surface area), TPSA (Sum of solvent accessible surface areas of atoms with absolute value of partial charges greater than or equal to 0.2) and WD.unity (Non-directional WHIM, weighted by unit weights) indicates that the inhibition activities of these compounds against B. Cinerea will be more when such descriptors reduces. Table (3) below is table of descriptions as well as the classes of descriptors that made up the built model.

Table (4) present the external validation while table 5 is for the calculation of predicted \( R^2 \) of the model1.
The observed and predicted activity of B. Cinerea inhibitors as a potential antifungal and their actual residual values are given in the table 6. This residual value is the different between the observed and predicted activities. The lower the residual values between the experimental and predicted activities signifies the higher prediction ability of the model.
Descriptors Correlation Matrix

The descriptors of the chosen model (model1) were selected and performed a correlation matrix on them as shown in table 7. The values indicated that some descriptors are inter-correlated while some are not for their correlation coefficients are greater than 0.5. The variance inflation factor (VIF) values are within the range of 1 to 5 which indicated that the descriptors and model are suitable and acceptable.

| Descriptors     | FMF   | RNCS  | TPSA  | WD.unity | Wgamma2.volume | VIF   |
|-----------------|-------|-------|-------|----------|----------------|-------|
| FMF             | 1.000 |       |       |          |                |       |
| RNCS            | -0.1103 | 1.000 |       |          |                | 4.1877|
| TPSA            | 0.3348 | -0.8345 | 1.000 |          |                | 4.1522|
| WD.unity        | -0.1538 | 0.5017 | -0.4809 | 1.000 |                | 1.6420|
| Wgamma2.volume  | -0.0575 | -0.5312 | 0.4036 | -0.5480 | 1.000 | 1.7059|

Some statistical parameters of the descriptors appeared in the built model are presented in the table 8 shown below. The magnitude of t-stat values for all descriptors are higher than 2 which signifies that the chosen descriptors are good (Adeniji et al., 2018). Also all the descriptors has p-values of less than 0.05 which signifies good relation between the descriptors and the inhibition concentration of the compounds.

| Descriptors     | Coefficients | Standard error | t-stat | p-value | Mean effect |
|-----------------|--------------|----------------|--------|---------|-------------|
| FMF             | 13.30837     | 1.409311       | 9.443172 | 2.14E-08 | 3.7879      |
| RNCS            | -0.3386      | 0.042274       | -8.0095 | 2.41E-07 | -0.5524     |
| TPSA            | -0.01298     | 0.002031       | -6.39088 | 5.11E-06 | -1.2873     |
| WD.unity        | -2.05464     | 0.289777       | -7.09043 | 1.31E-06 | -0.8852     |
| Wgamma2.volume  | 0.11287      | 0.035541       | 3.175746 | 0.005234 | 0.2179      |
Figure 3 – Plot of standardized residual versus experimental activity (pEC₅₀)

From figure 3, we witnessed a random disperse at a point where the standardized residual is zero which indicates the absence of systematic error while developing the model (Shola et al., 2018).

William’s Plot of Model 1

A Williams plot as shown in figure 4, is a plot of standardized residual versus leverages (for both the training set and test set compounds) of built model. The essence of this plot is to examine the presence of an outliers together with other influencing molecules present in the model. The result revealed that two (2) compounds from the test set were outside applicability domain of the compounds which indicated that the two compounds may have different structure from other in the dataset. Hence the compounds are beyond the threshold value or warning leverages h* which was calculated to be 0.75.

Figure 4 – Williams plot: a plot of standardized residual versus leverages (for both the training set and test set).

Conclusion

The QSAR model for 2-substituted phenyl-2-oxo-, 2-Hydroxy- and 2-Acloyxethylsulphonamides was successfully developed which predicted the toxic activity of the compounds against B. Cinerea by employing Genetic Function Approximation method. With model 1 being the best model, the R², R²_adj and Q²_cv are 0.954, 0.941 and 0.888 respectively, and the external validation R²_pred = 0.839. The research found that the toxicity of compounds was as a result of the molecular descriptors FMF, RNCS, TPSA, unity and Wgamma2 volume with their mean effect values of 3.787873, 0.55237, -1.28729, -0.88524 and 0.217949 respectively. This finding provides a guideline for development of new/novel sulfonyl compounds with excellent toxicity against B. Cinera fungi. Some of these compounds may include compounds 4, 6 and 18 (with pEC₅₀ of 0.71654, 0.77639 and 0.70044).

References

Abdulfatari, Usman, Adamu Uzairu, and Sani Uba. "Quantitative structure activity relationship study of anticonvulsant activity of α substituted acetamido-N-benzylacetamide derivatives." Cogent Chemistry 2, 2016.

Abdulfatari, Usman, Adamu Uzairu, and Sani Uba. "Investigation of some acetamido-N-benzylacetamide derivatives as potent anti-convulsant inhibitors." Journal of Computational Methods in Molecular Design 5, 2015.

Mustapha, Abdullahi, Gideon Shallangwa, Muhammad Tukur Ibrahim, Abdullahi Umar Bello, David Arthur Ebuka, Adamu Uzairu, and Paul Mamza. "QSAR studies on some C14-urea tetradrine compounds as potent anti-cancer against Leukemia cell line (K562)." Journal of the Turkish Chemical Society, Section A: Chemistry 5, 2018.

Adeniji, Shola Elijah, David Ebuka Arthur, and Adediran Oluwaseye. "Computational modeling of 4-
Phenoxynicotinamide and 4-Phenoxypyrimidine-5-carboxamide derivatives as potent anti-diabetic agent against TGR5 receptor. "Journal of King Saud University-Science, 2018.

Adeniji, Shola Elijah, Sani Uba, and Adamu Uzairu. "QSAR Modeling and Molecular Docking Analysis of Some Active Compounds against Mycobacterium tuberculosis Receptor (MtbcYP121)." "Journal of Pathogens (2018), 2018.

Adeniji, Shola Elijah, Sani Uba, and Adamu Uzairu. "In-silico study for investigating and predicting the activities of 1, 2, 4-triazole derivatives as potent anti-tubercular agents." "The Journal of Engineering and Exact Sciences 4, 2018.

Antifungal Activity Relationships of Clovane Derivatives against Botrytis cinerea." "Journal of Agricultural and Food Chemistry 57, 2009.

Arthur, David Ebuka. "Toxicity modelling of some active compounds against k562 cancer cell line using genetic algorithm-multiple linear regressions." "Journal of the Turkish Chemical Society, Section A: Chemistry 4, 2016.

Boddy, Lynne. "Pathogens of Autotrophs." In "The Fungi," pp. 245-292. Academic Press, 2016.

Dehmer, Matthias, Kurt Varmuza, Danail Bonchev, and Frank Emmert-Streib, eds. "Statistical modelling of molecular descriptors in QSAR/QSPR." Weinheim, Germany: Wiley-VCH, 2012.

Edache, Emmanuel Israel, David Ebuka Arthur, and Usman Abdulfatai. "tyrosine Activity of Some Tetraketone and Benzyl-benzoate Derivatives Based on Genetic Algorithm-Multiple Linear Regression." 2017.

Fisher, Robert S., J. Helen Cross, Carol D'souza, Jacqueline A. French, Sheryl R. Haut, Norimichi Higurashi, Edouard Hirsch et al. "Instruction manual for the ILAE 2017 operational classification of seizure types." "Epilepsia 58, 2017.

Gramatica, Paola, Stefano Cassani, Partha Pratim Roy, Simona kovarich, Chun Wie Yap, Ester Papa. "QSAR modeling is not “push button and find a correlation”: a case study of toxicity of (benzo-)triazoles on algae." "Molecular Informatics 31, 2012.

Huang, Zhaohua, Zhaoliang Lin, and Junlian Huang. "A novel kind of antitumour drugs using sulfonamide as parent compound." European journal of medicinal chemistry 36, 2001.

Ibrahim, Muhammad Tukur, Adamu Uzairu, Gideon Adamu Shallangwa, and Abdulkadir Ibrahim. "Computational studies of some bisoumarin and bisoumarin thiourea derivatives as α-glucosidase inhibitors." "The Journal of Engineering and Exact Sciences 4, 2018.

Ibrahim, Muhammad Tukur, Adamu Uzairu, Gideon Adamu Shallangwa, and Abdulqadir Ibrahim. "In-silico studies of some oxadiazoles derivatives as anti-diabetic compounds." "Journal of King Saud University-Science, 2018.

Pourbasheer, Eslam, Reza Aalizadeh, Mohammad Reza Ganjali, and Parviz Norouzi. "QSAR study of IKKB inhibitors by the genetic algorithm: multiple linear regressions." "Medicinal Chemistry Research 23, 2014.

Roy, Kunal, and Anis Sattwa Mandal. "Development of linear and nonlinear predictive QSAR models and their external validation using molecular similarity principle for anti-HIV indolyl aryl sulfones." "Journal of enzyme inhibition and medicinal chemistry 23, 2008.

Roy, Kunal, Supratik Kar, and Rudra Narayan Das. A primer on QSAR/QSPR modeling: Fundamental concepts. Springer, 2015.

Singla, Rajeev K., and Varadaraj Bhat G. "QSAR model for predicting the fungicidal action of 1,2,4-triazole derivatives against candida albicans." Journal of enzyme inhibition and medicinal chemistry 25, 2010.

Singla, Rajeev K., and Varadaraj Bhat G. "QSAR model for predicting the fungicidal action of 1, 2, 4-triazole derivatives against Candida albicans." "Journal of enzyme inhibition and medicinal chemistry 25, 2010.

Saiz-Urra, Liane, Juan C. Racero, Antonio J. Macias-Sanchez, Rosario Hernandez-Galan, James R. Hanson, Maykel Perez-Gonzalez, and Isidro G. Collado. "Synthesis and Quantitative Structure—Antifungal Activity Relationships of Clovane Derivatives against Botrytis cinerea." "Journal of agricultural and food chemistry 57, 2009.

Wang, Minlong, Peng Rui, Caixiu Liu, Ying Du, Peiwen Qin, Zhiqiu Qi, Mingshan Ji, Xinghai Li, and Zining Cui. "Design, synthesis and fungicidal activity of 2-substituted phenyl-2-oxo-, 2-hydroxy-and 2-acetoxyethylsulfonamides." "Molecules 22, 2017.

Wavefunction, Inc. "Spartan’14, version 1.1. 2." 2013.

Yap, Chun Wei. "PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints." "Journal of computational chemistry 32, 2011."