Journal of Pharmaceutical and Medicinal Research 2(1) (2016) 43–47

ISSN: 2455-0280

Contents List available at JACS Directory

Journal of Pharmaceutical and Medicinal Research

journal homepage: www.jacsdirectory.com/jpmr

QSAR Studies of Neolignans Derivatives using Physicochemical Descriptors: MLR and GFA-Modeling

A. Ajala*, A. Uzairu, I.O. Suleiman

Department of Chemistry, Ahmadu Bello University Zaria, Kaduna, Nigeria.

ARTICLE DETAILS

Article history:
Received 11 March 2016
Accepted 25 March 2016
Available online 05 April 2016

Keywords:
Density Functional Theory (DFT)
Genetic Function Approximation (GFA)
Neolignans
Epidermophyton floccosum
Antifungi

ABSTRACT

The Density Functional Theory (DFT) method (B3LP/6-31G) was employed to calculate a set of molecular properties (variables or descriptors) of 18 neolignans compounds with activities against *Epidermophyton floccosum* a most susceptible species of dermophytes. The correlation between biological activity and structural properties was obtained by using the multiple linear regression method. The model obtained showed not only statistical significance but also predictive ability. The aim of this study was to correlate the chemical structure of compounds with experimental data from biological activity anti-*Epidermophyton floccosum*. Eighteen (18) descriptors were calculated and multiparameter model was obtained through Genetic Function Approximation method. The results showed that thermodynamic, dimensional and steric parameters are important in elucidating of action mechanism compounds. Four descriptors (Solvation energy, Maximum Elpotential, Standard Enthalpy and Standard Gibbs energy) were selected and good model (n = 18; R^2 = 0.9215; R^2_{cv} = 0.7779; RMSE = 0.6612, PRESS = 0.5465; F = 20.5407; LOF = 0.016 and R^2_{adj} = 0.8954; R^2_{pred} = 0.7207; k = 0.9897; k' = 1.00; $|R^2_0 - R'^2_0|$ = 0.003 was built with Four variables describing the original information. Internal and External validation analysis were performed in order to confirm the robustness of the model. The proposed model may provide a better understanding of the anti-*Epidermophyton floccosum* activity of neolignans and can be used as guidance for proposition of new chemopreventive agents.

1. Introduction

Fungal infections of the skin and nails are a common global problem. The high prevalence of superficial mycotic infections shows that 20-25% of the world's population has skin mycoses making these one of the most frequent forms of infection. Pathogens responsible for skin mycoses are primarily anthropophilic and zoophilic dermatophytes from the genera *Trichophyton (T.)*, *Microsporum (M.)* and *Epidermophyton (E.)*. There appears to be considerable inter-continental and intra-continental variability in the global incidence of these fungal infections. *Trichophyton rubrum, T. interdigitale* (mentagrophytes var. interdigitale), *M. canis, M. audouinii, T. tonsurans and T. verrucosum* are the most common [1] but the attack rates and incidence of specific mycoses can vary widely. Local socio-economic conditions and cultural practices can also influence the prevalence of a particular infection in a given area. For example, tinea pedis (athletes foot) is more prevalent in developed countries than in emerging economies and is likely to be caused by the anthropophilic germ *T. rubrum*. In poorer countries, scalp infections (tinea capitis) caused by *T. soudanense* or *M. audouinii* are more prevalent. This review summarises current epidemiological trends for fungal infections and focuses on dermatomycosis of glabrous skin on different continents [2]. Chemical investigations of species of *Virola* and related genera from the Amazonian region lead to the hypothesis that the alleged usefulness of plasters made from their leaves or bark resin in the treatment of skin fungal infections, may be due to the fungistatic or fungitoxic activity of neolignans [3]. Amongst the wide variety of known neolignans, the 8.O.4′-type represents a small group whose members were isolated exclusively from plants of the Myristicaceae. Different ketones and alcohols of *threo* and *erythro* relative configuration have been isolated from *Virola surinamensis* (Roland) Warb [4]. In preliminary studies made with agar-dilution assays, we carried out an evaluation of the antifungal activities of 8.O.4′-neolignans and reported that alcohols, 1-12, but not the ketones 13-18 possess significant antifungal activity (MICs 5 - 250 μg/mL), against dermatophytes a group of fungi which characteristically infect the keratinized areas of human skin. This activity was dependent upon relative stereochemistry (*erythro* up to three times more active than *threo* alcohols) and upon substitution patterns at rings A and B. In addition, *Epidermophyton floccosum* was the most susceptible species [5].

QSAR methodologies have the potential of decreasing substantially the time and effort required for the discovery of the new medicines [6]. A major step in constructing the QSAR models is to find a set of molecular descriptors that represents variation of the structural properties of the molecules [7]. The QSAR analysis employs statistical methods to drive quantitative mathematical relationships between chemical structure and biological activity [8]. Thus, the use of the QSAR in the development of a theoretical model to predict the biological activity of a set of compounds is very important. The strategy used in the QSAR methodology includes the following steps: (i) selection of a data set; (ii) generation of the molecular structures; (iii) optimization of the geometry of the molecular structures by appropriate method; (iv) generation of several structural descriptors; (v) application of variable selection or/and methods data reduction of the calculated descriptors; (vi) regression analysis and finally (vii) evaluation of the validity and predictability of the developed QSAR models [8].

2. Experimental Methods

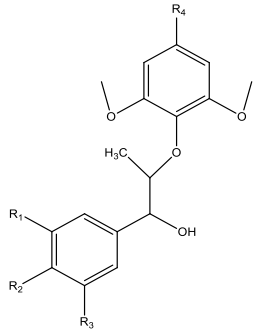
The purpose of the present work is to perform a quantum chemical QSAR study of the neolignans derivatives [Table 1a, b and c] to investigate the binding mode of these compounds and properties that are relevant for their activity. We use the approach of Hansch [9] in classical QSAR analysis for obtain linear model by the Multilinear regression Genetic function (MLR-GF) method to predict the experimental activity.

2.1 Chemical Data

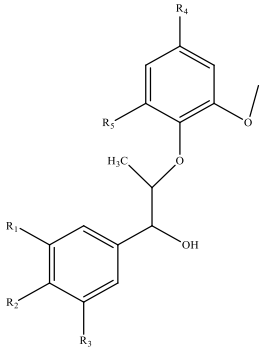
Biological data on the activity of ketone derivatives has been obtained from the literature [1-3] (Table 1a, b and c). The activity data refers pMIC, which indicates the biological activity of compounds experimentally determined necessary for the inhibition of *Epidermophyton floccosum* resistant. The -log MIC (molar) scale refers pMIC.

*Corresponding Author
Email Address: abdulajala39@gmail.com (Abduljelil Ajala)

(a)



(b)



(c)

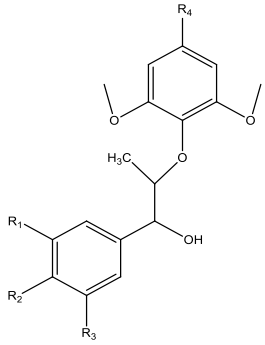


Table 1 Class a

Compound	R_1	R_2	R_3	R_4	MIC	pMIC
1	OCH_3	OCH_3	H	Allyl	0.103	0.99
2	OCH_3	OCH_3	H	Allyl	0.154	0.81
3	OCH_3	OCH_3	OCH_3	Allyl	1.120	0.92
4	OCH_3	OCH_3	OCH_3	Allyl	0.143	0.84
5	OCH_2O	-	H	Allyl	0.013	1.89
6	OCH_2O	-	H	Allyl	0.040	1.40

Class b

Compound	R_1	R_2	R_3	R_4	MIC	pMIC
7	OCH_3	OCH_3	H	Allyl	0.042	1.38
8	OCH_3	OCH_3	H	Allyl	0.670	0.17
9	OCH_3	OCH_3	H	t-propenyl	0.195	0.71
10	OCH_3	OCH_3	H	t-propenyl	0.279	0.55
11	OCH_3	OCH_3	OCH_3	t-propenyl	0.064	1.19
12	OCH_3	OCH_3	OCH_3	t-propenyl	0.129	0.89

Class c

Compound	R_1	R_2	R_3	R_4	R_5	MIC	pMIC
13	OCH_3	OCH_3	H	Allyl	OCH_3	0.647	0.65
14	OCH_3	OCH_3	OCH_3	Allyl	OCH_3	0.600	0.60
15	OCH_2O	-	H	Allyl	OCH_3	0.675	0.68
16	OCH_3	OCH_3	H	Allyl	H	0.701	0.15
17	OCH_3	OCH_3	H	t-propenyl	H	0.751	0.70
18	OCH_3	OCH_3	OCH_3	t-propenyl	H	0.647	0.65

2.2 Geometry Optimization

The Core-Seven personal computer equipped with the operating system Windows® Seven was used for making calculations of this work. The molecular structures of the dataset were sketched using Spartan '14 version 1.1.4 developed by Wave Function. The first step consisted in obtaining the molecular geometry of all the derivatives from the dataset (Table 1a, b and c) was energy minimization [10] and geometry optimization using Merck Molecular Force Field (MMFF) in semi-empirical PM3 Method [11]. We initially performed geometry optimization, which was done using semi-empirical PM3 (Parametric Method) Hamiltonian method in the Spartan '14 program. After this first procedure, the stability of the molecular geometry was obtained by Density Functional Theory method (B3LP/6-31G), using Spartan '14 software v1.1.4.

2.3 Structural Descriptors

In the quantum chemical analysis we calculated 18 properties of all compounds. The calculated physical-chemical parameters types are: hydrophobic, electronic, steric, thermodynamic, dimensional, topological and geometric. All the molecular properties were calculated by Spartan '14 program and MMPP computational package.

2.4 Computational and Statistical Details

The genetic function approximation (GFA) algorithm offers a new approach to the problem of building quantitative structure-activity relationship (QSAR) and quantitative structure-property relationship (QSPR) models. Replacing regression analysis with the GFA algorithm enables the construction of models competitive with or superior to those produced by standard techniques and makes available additional information not provided by other techniques. Unlike most other analysis algorithms, GFA gives multiple models, where the populations of the models are created by evolving random initial models using a genetic algorithm [12]. GFA can build models using not only linear polynomials but also higher-order polynomials, splines and other nonlinear functions. The genetic algorithms are search algorithms that take inspiration from natural genetics and evolution. In this section, the ideas underlying genetic algorithms are briefly described, emphasizing the aspects relevant to the genetic function approximation (GFA) approach to model building. The GFA algorithm itself applies these ideas to the problem of function approximation [13] given a large number of potential factors influencing a response including several powers and other functions of the raw inputs, to find the subset of terms that correlates best with the response. The central ideas of genetic algorithms are simple. The region to be searched is coded into one or multiple strings. In the GFA, these strings are sets of terms-powers and splines of the raw inputs. Each string represents a location in the search space. The algorithm works with a set of these strings called a population. This population is evolved in a manner that leads it toward the objective of the search. This requires that a measure of the fitness of each string, corresponding to a model in the GFA, be available following this, three operations are performed iteratively in succession: selection, crossover and mutation. Newly added members are scored according to a fitness criterion. In the GFA, the scoring criteria for models are all related to the quality of the regression fit to the data [14]. The selection probabilities must be re-evaluated each time a new member is added to the population. Stability and convergence In common with other iterative minimization algorithms, there are issues with the stability and convergence of the GFA algorithm. An indication of the stability of the GFA algorithm can be obtained by generating a plot showing the evolution of variable usage with time. Such a plot shows the number of occurrences of each variable in the population for each generation of the evolution. For practical reasons to reduce the amount of data that would be collected, such a plot is generated only for those variables that occur most commonly in the final population and the data are not normally collected for every generation. The GFA algorithm is assumed to have converged when no improvement is seen in the score of the population over a significant length of time, either that of the best model in each population or the average of all the models in each population. When this criterion has been satisfied, no further generations are calculated. Advantages of GFA the GFA algorithm approach has several important advantages over other techniques [15]:

1. It builds multiple models rather than a single model.
2. It automatically selects which features are to be used in the models.
3. It is better at discovering combinations of features that take advantage of correlations between multiple features.
4. It incorporates Friedman's lack-of-fit (LOF) [16] error measure, which estimates the most appropriate number of features, resists over fitting and allows control over the smoothness of fit.
5. It can use a large variety of equation term types in construction of its models e.g., splines, step functions, high order polynomials.
6. It provides, through study of the evolving models, additional information not available from standard regression analysis, such as the preferred model length and useful partitions of the data set. The procedure continues for a user-specified number of generations, unless convergence occurs in the interim. Convergence is triggered by lack of progress in the highest and average scores of the population.

2.5 Building a Structure-Activity Relationship or Model

The genetic function approximation (GFA) algorithm offers a new approach to the problem of building quantitative structure-activity relationship (QSAR) and quantitative structure-property relationship (QSPR) models. Replacing regression analysis with the GFA algorithm enables the construction of models competitive with or superior to those produced by standard techniques and makes available additional information not provided by other techniques. Unlike most other analysis algorithms, GFA gives multiple models, where the populations of the models are created by evolving random initial models using a genetic algorithm. GFA can build models using not only linear polynomials but also

higher-order polynomials, splines and other nonlinear functions [16]. Friedman's MARS algorithm is a statistical technique for modeling data. It provides an error measure called the lack-of-fit (LOF) score, Eq. (1) that automatically penalizes models with too many features. It also inspired the use of splines as a powerful tool for nonlinear modeling [16]. The equation below is for LOF:

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

Where SSE is the sum of squares of errors, c is the number of terms in the model, other than the constant term, d is a user-defined smoothing parameter, p is the total number of descriptors contained in all model terms (again ignoring the constant term) and M is the number of samples in the training set. Unlike the commonly used least squares measure the LOF measure cannot always be reduced by adding more terms to the regression model. While the new term may reduce the SSE, it also increases the values of c and p, which tends to increase the LOF score. Thus, adding a new term may reduce the SSE, but actually increases the LOF score. By limiting the tendency to simply add more terms, the LOF measure resists over fitting better than the SSE measure [17].

3. Results and Discussion

The aim of the current study is mainly to study the different physicochemical parameters linear models and their regression analysis by using indicator variable and different parameters.

A QSAR analysis was performed to explore the structure activity relationship of different 18 neolignans derivatives acting as antifungi. In a QSAR study, generally, the quality of a model is expressed by its fitting and prediction ability. In order to build and test model, a data set of 18 compounds was separated into a training set of 13 compounds, which was used to build model and a test set of 5 compounds, which was applied to evaluate the built model. The GA-MLR analysis led to the derivation of one model, with four descriptors. With the selected descriptors, we have built the linear model using the training set data and obtained the best model given by Eq. (2) below:

$$pMIC = -0.05424a - 0.003097b + 23.45799c - 23.4555d + 1.139116 \quad (2)$$

where, a = solvation energy, b = maximum Electropotential, c = enthalpy and d = Gibbs energy. N = 13; Friedman LOF = 0.0160; R^2 = 0.9215; R^2_{adj} = 0.8954; R^2_{cv} = 0.7779; F = 20.5407 and R^2_{pred} = 0.7202

Table 2 List of descriptors used in this study

Descriptors	Type	Significance
Solvation energy	Quantum Electronics	This is the minimum energy of the molecular conformation
Maximum Elpotential	Dimensional	It describe the van der waals surface area of the molecule
Standard Enthalpy (H°)	Thermodynamic	In theses descriptors, the molecule atoms represent a set of diverse points in space and the atomic property and function are evaluated at those points
Standard Gibbs Energy (G°)	Thermodynamic	It depends on the spatial array of the aromatic ring in the synthesized compounds also necessary to study the interaction of the ligand with the receptor

Table 3 Table for test set

Compound	Observed Activity	Predicted Activity	Residual
2	0.81	1.1269	0.3169
4	0.84	0.884675	0.044675
6	1.4	1.01247	-0.38753
8	0.17	1.380989	1.210989
10	0.55	0.780332	0.230332

To investigate the observed data, the distribution of the data must be first investigated. Most regression algorithm relies on the data that is being normally investigated in case the data are not normally distributed we should consider applying a numerical transformation to achieve a normal distribution. Observed data in Table 1(a, b and c) show acceptable normal distribution, so no need to perform a numerical transformation. Table 4 shows a univariate analysis for the inhibition data. Table 4 contains several statistical measures that describe the observed activity data. The most important parameters in Table 4 are the skewness and kurtosis. Skewness is the third moment of the distribution, which indicates the symmetry of the distribution.

Table 4 Univariate analysis of the observed data

S.No.	Statistical parameters	pMIC
1	Number of sample points	13
2	Range	1.23000000
3	Maximum	1.38000000
4	Minimum	0.15000000
5	Mean	0.79333333
6	Median	0.70500000
7	Variance	0.09030560
8	Standard deviation	0.31387100
9	Mean absolute deviation	0.23388900
10	Skewness	-0.01826450
11	Kurtosis	-0.26285600

Table 5 Observed and predicted activity for training set

Observed Activity	Predicted Activity	Residual
0.99000000	1.00890000	-0.01890000
0.92000000	1.07022400	-0.15022400
1.89000000	1.78829400	0.10170600
1.38000000	1.38493100	-0.00493100
0.71000000	0.72571600	-0.01571600
1.19000000	0.96620300	0.22379700
0.89000000	0.96620300	-0.07620300
0.65000000	0.75221400	-0.10221400
0.60000000	0.51273700	0.08726300
0.68000000	0.68004800	-4.800000e-5
0.15000000	0.06590000	0.08410000
0.70000000	0.80034900	-0.10034900
0.66000000	0.68828000	-0.02828000

Table 5 shows the experimental pMIC and the predicted pMIC using the GFA approach of the training set. This shows how the GFA method predicted the pMIC.

The genetic function approximation analysis gives a summary of the input parameters used for the calculation that is the square of the correlation coefficient (R^2), the cross validated correlation coefficient (R^2_{cv}), the Fisher criterion-(F) and the R^2 predicted were used as criteria for the stability and the robustness of the models. Also it reports whether the GFA algorithm converged in the specified number of generations. Convergence is achieved when there has been no improvement in the scoring function for a number of generations. It can be seen from Table 6 that the accuracy of the model, indicated by the R^2 value is reasonably high therefore the predictive power of the model as indicated by the adjusted R^2 and cross validated R^2 values, is also high even though the regression is significant according to F-test. The Friedman's lack-of-fit (LOF) score [17], which evaluates the QSAR model by considering the number of descriptors as well as the quality of fitness is chosen the lower the LOF, the less likely it is that GFA model will fit the data. The significant regression is given by F-test and the higher the value the better the model.

Table 6 Validation table of the genetic function approximation

S.No.	Equation Parameter	Values
1	Friedman LOF	0.016
2	R^2	0.9215
3	R^2_{adj}	0.8954
4	R^2_{cv}	0.7779
5	F-value	20.5407

Fig. 1 shows a relation between the predicted values using the Eq. (1) above and the experimental data in Table 1. Also Table 2 shows the distribution of the residual values against the observed activity values. A residual can be defined as the difference between the predicted value in the generated model and the measured value for observed activity. To test the constructed QSAR model, potential outliers have been identified in Figs. 3a and b. An outlier can be defined as a data point whose residual value is not within two standard deviations of the mean of the residual values. Although the number of outliers can vary depending on the quality of the dataset (e.g., incorrect measurements of physical properties or errors in molecular structures will reduce the data set quality), it still a

good test of QSAR model is to identify potential outliers. Fig. 3a and b contains two charts. One contains the residual values plotted against the observed pMIC and the other displays the residual values plotted against Table 1 raw number. Each chart contains a dotted line that indicates the critical threshold of two standard deviations beyond which a value may be considered to an outlier. Inspection of Figs 3a and b shows that there is no points appeared outside the dotted lines which make the QSAR model acceptable.

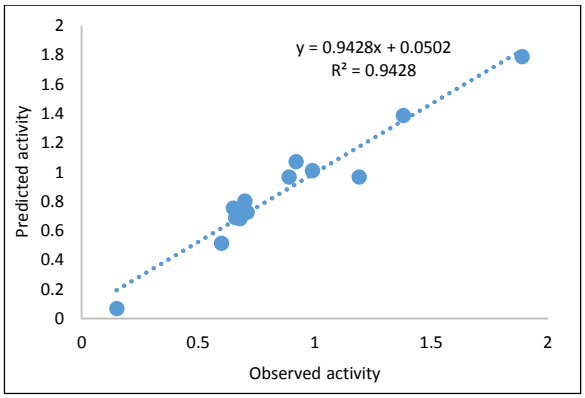


Fig. 1 Plot of predicted activity against observed activity for training set

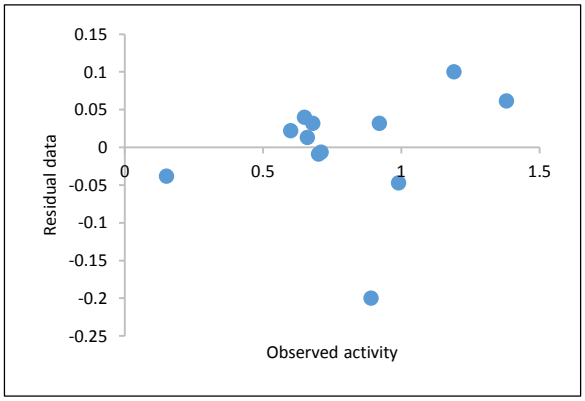
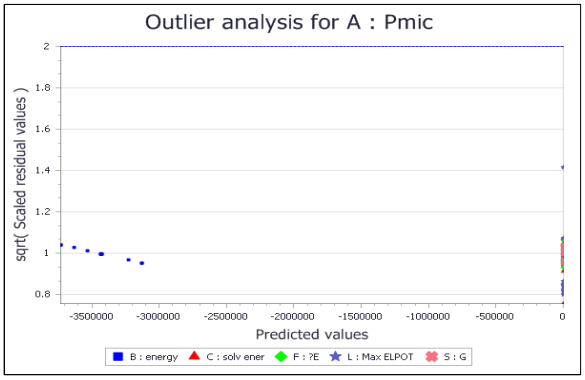
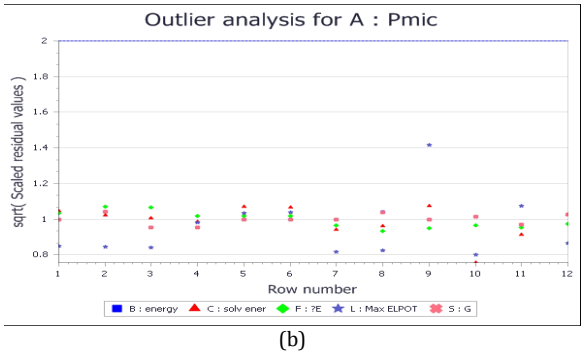


Fig. 2 plot of Residual data against observed activity for training set



(a)



(b)

Fig. 3 Outlier analysis a) It contains the residual values plotted against the Observed pMIC and b) displays the residual values plotted against Table 1 raw number

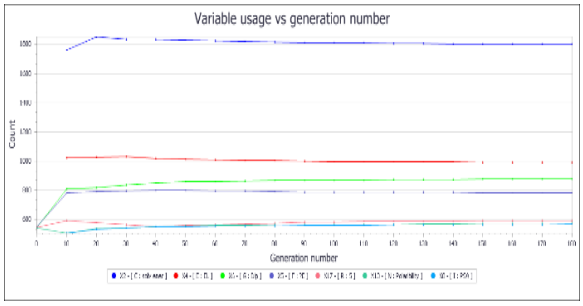


Fig. 4 The graph of the variable usage against generation number

In Fig. 4 the Y-axis represents the different molecular descriptors used in this study as shown on the left side of the graph. On the other hand the X-axis represents the number of the generations we could generate for each of these molecular descriptors. According to Fig. 4, at each step the GFA uses the current population to create the children that makes up the next generation. The algorithm selects a group of individuals in the current population called parents, who contribute their genes the entries of their vectors to their children. The algorithm usually selects individuals that have better fitness values as parents. User can specify the function that the algorithm uses to select the parents. The GFA creates three types of children for the next generation: Elite children, crossover children and mutation children. In our QSAR study, the algorithm stops when the number of generations reaches the value of 180 generations.

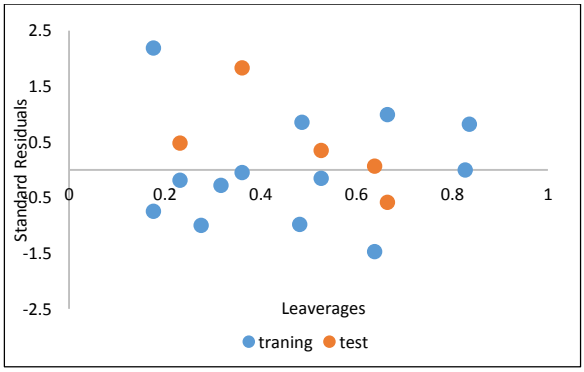


Fig. 5 The Williams plot, the plot of the standardized residuals versus the leverage value for training and test set

The leverage values can be calculated for every compound and plotted vs. standardized residuals and it allows a graphical detection of both the outliers and the influential chemicals in a model. Fig. 5, shows the Williams plot the applicability domain is established inside a squared area within ± 2.5 bound for residuals and a leverage threshold h*, [18].

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

Where h^* the warning leverage, k is the number of descriptor in the model and n is the number of observation that make up the training set. It demonstrates that all the compounds of the training set and test set are inside of this square area. From Fig. 5, it is obvious that there are no outlier compounds with standard residuals > 3d for both the training and test sets. Furthermore, all the chemicals have a leverage lower than the warning h* value of 1.154. All results confirm that the build model is a valid model and can be utilized to predict the activity of neolignans.

4. Conclusion

A QSAR model was derived using GFA for a series of neolignans antifungals. The best model generated correlates with the antifungal activity. The model has moderate internal and external predictivity. Significant regression equations were obtained by MLR method for 13 neolignan compounds according to their antifungal activity. The best regression equation obtained was based on the following descriptors: solvation energy, maximum electropotential, standard enthalpy and standard Gibbs energy. The model obtained showed not only statistical significance but also predictive ability and robust. These variables allowed a physical explanation of electronic molecular properties contributing to antifungal inhibitory potency as the electronic character relates directly to the electron distribution of interacting molecules at the active site. On the basis of the developed QSAR models, novel molecules can be designed as potential antifungi.

References

[1] D.J. Lodge, D.L. Hawksworth, B.J. Ritchie, Microbial diversity and tropical forest functioning, In Biodiversity and ecosystem processes in tropical forests, Springer, Berlin Heidelberg, 1996, pp.69-100.

[2] S. Zacchino, G. Rodríguez, G. Pezzenati, G. Orellana, R. Enriz, S.M. Gonzalez, In vitro evaluation of antifungal properties of 8. O. 4'-neolignans, J. Nat. Prod. 60(7) (1997) 659-662.

[3] J.S. Tkacz, Glucan biosynthesis in fungi and its inhibition, In Emerging targets in antibacterial and antifungal chemotherapy, Springer, US, 1992, pp.495-523.

[4] S. Zacchino, G. Rodrıguez, C. Santecchia, G. Pezzenati, F. Giannini, R. Enriz, In vitro studies on mode of action of antifungal 8. O. 4′-neolignans occurring in certain species of Virola and related genera of Myristicaceae, J. Ethnopharmacol. 62(1) (1998) 35-41.

[5] A.B. Macura, Dermatophyte infections, Int. Jour. Dermatology 32(5) (1993) 313-323.

[6] W. Tong, H. Hong, Q. Xie, L. Shi, H. Fang, R. Perkins, Assessing QSAR limitations-A regulatory perspective, Curr. Comput.- Aided Drug Des. 1(2) (2005) 195-205.

[7] L. He, P.C. Jurs, Assessing the reliability of a QSAR model's predictions, J. Mol. Graphics Modell. 23(6) (2005) 503-523.

[8] T. Ghafourian, M.T. Cronin, The impact of variable selection on the modelling of oestrogenicity, SAR and QSAR in Env. Res. 16(1-2) (2005) 171-190.

[9] S.H. Unger, C. Hansch, Model building in structure-activity relations. Reexamination of adrenergic blocking activity of. beta.-halo-. beta-arylalkylamines, J. Med. Chem. 16(7) (1973) 745-749.

[10] J.C. Byrd, R.R. Furman, S.E. Coutre, I.W. Flinn, J.A. Burger, K.A. Blum, et al, Targeting BTK with ibrutinib in relapsed chronic lymphocytic leukemia, New England J. Med. 369(1) (2013) 32-42.

[11] T.A. Halgren, Merck molecular force field. I. Basis, form scope parameterization and performance of MMFF94, J. Comput. Chem. 17(5-6) (1996) 490-519.

[12] J. Devillers, Prediction of toxicity of organophosphorus insecticides against the midge, Chironomus riparius, via a QSAR neural network model integrating environmental variables, Toxicol. Methods 10(1) (2000) 69-79.

[13] T.J. Hou, J.M. Wang, N. Liao, X.J. Xu, Applications of genetic algorithms on the structure-activity relationship analysis of some cinnamamides, J. Chem. Inf. Comput. Sci. 39(5) (1999) 775-781.

[14] D. Niculescu, B. Nath, Ad hoc positioning system (APS) using AOA. In INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications, IEEE Societies 3 (2003)1734-1743.

[15] D. Rogers, A.J. Hopfinger, Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships, J. Chem. Inf. Comput. Sci. 34(4) (1994) 854-866.

[16] K.F. Khaled, N.S. Abdel-Shafi, Quantitative structure and activity relationship modeling study of corrosion inhibitors: genetic function approximation and molecular dynamics simulation methods, Int. J. Electrochem. Sci, 6(9) (2011) 4077-4094.

[17] J.H. Friedman, Multivariate adaptive regression splines, The annals of statistics 15(8) (1991)1-67.

[18] P. Gramatica, Principle of QSAR model validation: internal and external, QSAR Combinatorial Sci. 26(5) (2007) 694-701.

Cite this Article as: A. Ajala, A. Uzairu, I.O. Suleiman, QSAR studies of neolignans derivatives using physicochemical descriptors: MLR and GFA-modeling, J. Pharm. Med. Res. 2(1) (2016) 43–47.