Regular Article

PHYSICAL CHEMISTRY RESEARCH
Published by the
Iranian Chemical Society
www.physchemres.org
info@physchemres.org

***Phys. Chem. Res.**, Vol. 6, No. 1, 209-223, March 2018*
DOI: 10.22036/pcr.2017.95755.1409

Quantitative Structure Activity Relationship Analysis of Coumarins as Free Radical Scavengers by Genetic Function Algorithm

I.O. Alisi[a,*], A. Uzairu[b], S.E. Abechi[b] and S.O. Idris[b]

[a]*Department of Applied Chemistry, Federal University Dutsinma, Katsina State, Nigeria*
[b]*Department of Chemistry, Ahmadu Bello University Zaria, Kaduna State, Nigeria*

(Received 20 August 2017, Accepted 13 November 2017)

The antioxidant properties of coumarin derivatives using the 2,2'-diphenyl-1-picrylhydrazyl (DPPH) radical scavenging assay were investigated by the application of quantitative structure activity relationship (QSAR) studies. The molecular structures were optimized and submitted for the generation of quantum chemical and molecular descriptors. Genetic function algorithm (GFA) was employed in model development. Also, the applicability domain of the developed models were accessed by the leverage approach. The variation inflation factor (VIF), mean effect (MF) and degree of contribution (DC) of these descriptors were estimated. Five predictive QSAR models were developed and subjected to various validation tests. The developed models gave highly encouraging results upon validation (R 0.938, R^2 = 0.879, Q^2 = 0.845, $R^2_{pred} = 0.700$ and $R^2_p = 0.832$. This research indicates that the most crucial descriptors influencing the free radical scavenging activities and essential in the design of new set of the coumarin antioxidants are the HBDCount (hydrogen-bond donor count), AATS3e (average Broto-Moreau autocorrelation-lag3/weighted by Sanderson electronegativities), and the MW (molecular weight) descriptors.

Keywords: Antioxidants, Coumarins, Descriptors, Model validation, QSAR model

INTRODUCTION

Recently, the chemistry of coumarins (1,2-benzopyrones) and their derivatives have arisen considerable interest due to their wide range of biological and pharmacological activities. These include antioxidant, anticancer, enzyme inhibition, vasorelaxant, anti-inflammatory, anti-HIV, antifilarial, cytotoxic, lipid-lowering, anti-gastrointestinal, antibacterial, anticoagulant, and acetylcholinesterase inhibitory activities [1,2]. In foods and pastries, coumarins have also been used as aroma enhancer [3].

Chemical entities containing unpaired electrons in their atomic or molecular orbitals are called free radicals [4]. They are produced as a result of normal cellular activities in the body, or due to external agents such as exposure to X-rays, industrial chemicals, heavy metals and cigarette smoking [5,6]. Free radicals are highly unstable, and readily react with other substances in their environment in order to acquire electron to achieve stability. At moderate levels, free radicals are beneficial to living cells. However, at high levels they lead to oxidative stress which is harmful to cell structures. Thus, maintaining a balance in the formation and neutralization of free radicals in the human system is critical for the wellbeing of an individual. In order to maintain a balance in the level of free radicals, antioxidants are employed.

Antioxidants are natural or synthetic compounds that scavenge free radicals by stabilizing or deactivating them by one or more mechanisms [7]. Compounds possessing antioxidant activities are used as free radical scavengers. The free radical scavenging activities of various compounds have been explored and reported by the 2,2'-diphenyl-1-picrylhydrazyl (DPPH) radical, hydroxyl (OH) radical,

*Corresponding author. E-mail: ialisi@fudutsinma.edu.ng

nitric oxide (NO) radical, hydrogen peroxide (H_2O_2) radical, superoxide anion (O_2^-) radical, 2,2'-azino-bis(3-ethylbenzo-thiazoline-6-sulfonic acid) (ABTS) cation radical assays and ferric ion reducing antioxidant power (FRAP) test methods [8-12].

Quantitative structure activity relationship (QSAR) is a technique widely employed to correlate molecular structure and chemical activity [13-16]. The fundamental principle underlying the use of QSAR is that the biological activities of compounds is a function of their structural properties.

This research is born out of the need to develop an efficient predictive model which can be employed successfully for the design of new set of coumarin antioxidants. A data set of 62 coumarin antioxidants with their respective DPPH IC_{50} antioxidant activities were generated from literature and subjected to quantitative structure activity relationship studies. Data optimization was carried out at the density functional theory (DFT) level together with the Becke's three-parameter Lee-Yang-Parr (B3LYP) exchange-correlation hybrid functional in combination with the 6-311G* basis set. Quantum chemical and molecular descriptors were calculated for the optimized structures and subsequently subjected to model development by employing Genetic Function Algorithm. Internal and external validation techniques were employed to validate the developed models. The variation inflation factor (VIF), mean effect (MF) and degree of contribution (DC) of each descriptor in the developed model were calculated. Also, the applicability domain of the model was accessed by the leverage approach.

EXPERIMENTAL

Data Set Generation, Optimization, Descriptors Calculation and Normalization

The coumarins data set of 62 molecular structures with their corresponding DPPH IC_{50} (μg ml^{-1}) antioxidant activities were generated from literature [17-20]. Their IC_{50} values were subsequently converted to their corresponding pIC_{50} values (Eq. (1)). This is to ensure that a more uniformly generated data is obtained.

$$pIC_{50} = -\log(IC_{50} \times 10^{-6}) \quad (1)$$

The chemical structures of the compounds were drawn using the Chemdraw program [21], while the molecular geometries were first minimized and subsequently optimized using Spartan 14 V1.1.4 wave function program at the DFT level [22]. Quantum chemical descriptors were also generated during the DFT computation. Molecular descriptors were calculated from the low energy conformers using the "PaDel-Descriptor package V2.20" [23].

Various descriptors such as electronic, spatial, structural, thermodynamic and topological were calculated for each molecule. The entire data set was subjected to data pre-treatment in order to eliminate variable pairs whose correlation coefficient are greater than 0.9 [24,25].

The resulting data set were normalized by scaling between the interval N(0,1) to eliminate the possibility of a descriptor dominating the model as a result of lower or higher pre-scaled value compared to the other descriptors in the developed model [26,27].

Model Development

The tool "Dataset Division GUI 1.2" [28], was employed in splitting the data into training and test sets by Kennard Stone algorithm. The resulting training set was employed in model development by genetic function algorithm (GFA) using the material studio tool. In this method, the dependent variable (antioxidant activities) and the independent variables (quantum chemical and molecular descriptors) were subjected to multivariate analysis using 50,000 crossovers, a smoothness value of 0.5 and other default settings. The Friedman lack-of-fit (LOF) which measures the fitness of a model was calculated through (Eq. (2));

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

where p is the total number of descriptors in the model, M is the number of training set samples, d is the smoothing parameter, c is the number of terms in the model excluding the constant term while SSE is the sum of squares of errors [29].

Model Validation

Internal and external validation techniques were employed to validate the developed models. The leave- one-out (LOO) cross-validation technique was employed to internally validate the models [30]. Various internal validation parameters were calculated as presented below:

The correlation coefficient, R. This parameter measures the variation in the calculated data with respect to the observed data [30].

The cross-validated squared correlation coefficient, $R^2_{cv}(Q^2)$. For the calculation of this parameter, Eq. (3) was employed.

$$Q^2 = 1 - \frac{\sum(Y_{obs} - Y_{pred})^2}{\sum(Y_{obs} - \overline{Y})^2} \quad (3)$$

where Y_{obs} represents the observed activity of the training set compounds, Y_{pred} is the predicted activity of the training set and $\overline{Y}$ represents the mean observed activity of the training set.

The adjusted R^2 (R^2_a) was computed using Eq. (4). This parameter is a modification of R^2 in which an improvement in its value upon the addition of a new term only gives rise to a result greater than what is obtainable by chance [32];

$$R^2_a = \frac{(n-1)R^2 - p}{n - p - 1} \quad (4)$$

where the predictor variables employed in the model development is represented as p.

The variance ratio, F, which is employed to judge the overall significance of the regression coefficients was also calculated (Eq. (5)).

$$F = \frac{\frac{\sum(Y_{cal} - \overline{Y})^2}{p}}{\frac{\sum(Y_{obs} - Y_{cal})^2}{N - P - 1}} \quad (5)$$

The standard error of the estimate (s) was calculated using Eq. (6);

$$s = \sqrt{\frac{RSS}{n - p'}} \quad (6)$$

where $RSS = \Sigma (Y_{obs} - Y_{pred})^2$ is the sum of squares of the residuals for training set predictions, p' is the number of model variables plus one, and n is the number of objects used to calculate the model [33,34].

External validation gives the predictive capacity of the developed model as judged by its application for the prediction of test set activity values and calculation The ability of the developed model to predict the test set activities is reflected in the value of the predictive R^2 (R^2_{pred}) value (Eq. (7));

$$R^2_{pred} = 1 - \frac{\sum(Y_{pred(Test)} - Y_{(Test)})^2}{\sum(Y_{(Test)} - \overline{Y}_{(Training)})^2} \quad (7)$$

where $Y_{pred(Test)}$ and $Y_{(Test)}$ are the predicted and observed activity values, respectively, of the test set compounds. $\overline{Y}_{(Training)}$ is the training set mean activity. R^2_{pred} is the predicted squared correlation coefficient obtained from the predicted activity of the test set. From Eq. (7), we observe that the R^2_{pred} value is controlled by the expression $\sum(Y_{(Test)} - \overline{Y}_{(Training)})^2$. Thus, it depends on the training set mean, and for new data set, it may not be truly predictive and may result in significant numerical difference between the observed and predicted values [35]. For a better measurement of the external predictivity of the developed model, a modified R^2 called r^2_m which determines the correlation between the predicted activity and the corresponding observed activity range was introduced, [16,36] as defined in Eq. (8);

$$r^2_m = r^2\left(1 - \sqrt{r^2 - r^2_0}\right) \quad (8)$$

where r^2_0 and r^2 represent squared correlation coefficients of the linear correlations between the observed and predicted values of the compounds with intercept set to zero and intercept not set to zero, respectively. When the axes are interchanged, the parameter r'^2_m is obtained (Eq. (9));

$$r'^2_m = r^2 \times \left(1 - \sqrt{r^2 - r'^2_0}\right) \quad (9)$$

where r'^2_0 bears the same meaning as r^2_0 but in the reversed axes. A plot of the observed values of test set compounds against the predicted values with intercept set to zero has

slope equal to k. Interchange of the axes gives slope equal to k' [32].

Also calculated was the root mean square error in prediction (RMSEP) which indicates the error between the mean of the experimental values and the predicted values (Eq. (10)) [37];

$$RMSE = \sqrt{\frac{\sum (Y_{obs(test)} - Y_{pred(test)})^2}{n_{ext}}} \quad (10)$$

where n_{ext} = number of compounds in the test set. The program: "DTC-MLR Plus Validation GUI 1.2" [27,38-40] was employed for the computation of the external validation parameters.

Randomization Test

The robustness of the developed QSAR model was checked using the Y-randomization technique in which model randomization was employed. Y-randomization test which checks the robustness of the developed model was performed by permuting the activity values with respect to the descriptor matrix [36].

The R_p^2 parameter which gives the deviation in the values of R_r^2 from R^2 was calculated using Eq. (11);

$$R_p^2 = R \times \sqrt{R^2 - R_2^2} \quad (11)$$

where R_r^2 represent the squared mean correlation coefficient of the randomized model, and R^2 is the squared correlation coefficient of the non-random model [36,41,42].

Estimation of the Variation Inflation Factor, Mean Effect and Degree of Contribution of the Descriptors

The observed variation inflation factors (VIF) for a given model indicates the multi-collinearity, among the descriptors [43]. This factor was computed for each descriptor using Eq. (12);

$$VIF = \frac{1}{1 - r^2} \quad (12)$$

where r is the correlation coefficient of multiple regressions of one descriptor with the other descriptors in the QSAR model [44]. For the computation of the mean effect (MF) of each descriptor in the model, Eq. (13) was employed;

$$MF_j = \frac{\beta_j \sum_{i=1}^{i=n} d_{ij}}{\sum_j^m \beta_j \sum_1^n d_{ij}} \quad (13)$$

where MF_j represents the mean effect for the considered descriptor j, β_j is the coefficient of the descriptor j, d_{ij} stands for the value of the target descriptors for each molecule, while m is the number of descriptors in the model [45]. Also, calculated for each descriptor is the degree of contribution (DC) (standardized regression coefficient).

Applicability Domain Investigation

The applicability domain of a QSAR model is the vicinity of chemical structures where the predictions of the developed model can reliably be applied [46]. A developed model cannot make predictions for molecules that fall outside the applicability domain [47-48]. Thus, prediction results that are interpolations in the chemical space are acceptable. On the other hand, prediction results that are extrapolations in the chemical space are rejected.

The leverage approach was employed in accessing the applicability domain of the developed QSAR model [49]. The leverage (hat) matrix (H), as presented in Eq. (14), was used to calculate the leverage value of all compounds in the dataset (X);

$$H = X(X^T X)^{-1} X^T \quad (14)$$

where X represents the training set two-dimensional n × k descriptor matrix, *n* is the number of compounds, *k* is the number of descriptors while X^T is the transpose of X. The i th compound leverage value (h_i) is the ith diagonal element of H (Eq. (15)).

$$h_i = x_i (X^T X)^{-1} x_i^T \qquad (i = 1, \ldots, m) \quad (15)$$

The limit of normal values for X outliers is represented as the leverage threshold, warning leverage or cut-off leverage value, h*, (Eq. (16)) [50].

Table 1. Coumarin Data Set and their Corresponding IC_{50} and pIC_{50} Values

Comp. No.	Compounds	IC_{50} ($\mu g\ ml^{-1}$)	pIC_{50}		
			Observed	Predicted	Residual
M01[a]	4-Methyl-2H-chromen-2-one	15568.700	1.808	1.906	-0.098
M02	4,6-Dimethyl-2H-chromen-2-one	13465.600	1.871	1.721	0.150
M03	4,7-Dimethyl-2H-chromen-2-one	14876.600	1.827	1.595	0.233
M04	6-Methoxy-4-methyl-2H-chromen-2-one	2624.730	2.581	3.327	-0.746
M05	7-Methoxy-4-methyl-2H-chromen-2-one	2795.910	2.553	3.243	-0.689
M06	6-Hydroxy-4-methyl-2H-chromen-2-one	1032.360	2.986	3.832	-0.845
M07	7-Hydroxy-4-methyl-2H-chromen-2-one	1312.470	2.882	3.675	-0.793
M08	7-Hydroxy-4,8-dimethyl-2H-chromen-2-one	1329.480	2.876	3.483	-0.607
M09	7-Hydroxy-4,5-dimethyl-2H-chromen-2-one	1363.720	2.865	3.065	-0.199
M10	6-Hydroxy-4,7-dimethyl-2H-chromen-2-one	456.475	3.341	3.506	-0.166
M11	7,8-Dihydroxy-4-methyl-2H-chromen-2-one	1.922	5.716	5.414	0.303
M12	4-Methyl-2-oxo-2H-chromen-6-yl acetate	473.511	3.325	2.558	0.767
M13	4-Methyl-2-oxo-2H-chromen-7-yl acetate	822.644	3.085	2.465	0.619
M14	4,8-Dimethyl-2-oxo-2H-chromen-7-yl acetate	14746.900	1.831	2.361	-0.53
M15	4,5-Dimethyl-2-oxo-2H-chromen-7-yl acetate	17696.300	1.752	1.953	-0.201
M16	4,7-Dimethyl-2-oxo-2H-chromen-6-yl acetate	3553.200	2.449	2.311	0.138
M17	4-Methyl-2-oxo-2H-chromen-7,8-diyl diacetate	314.918	3.502	3.456	0.045
M18	4-Methyl-2-oxo-2H-chromen-6-yl benzoate	6978.950	2.156	2.121	0.035
M19	4-Methyl-2-oxo-2H-chromen-7-yl benzoate	9053.010	2.043	2.022	0.021
M20	4,8-Dimethyl-2-oxo-2H-chromen-7-yl benzoate	11065.900	1.956	2.034	-0.078
M21	4,5-Dimethyl-2-oxo-2H-chromen-7-yl benzoate	14273.800	1.845	1.653	0.192
M22	4,7-Dimethyl-2-oxo-2H-chromen-6-yl benzoate	7887.400	2.103	1.993	0.11
M23	4-Methyl-2-oxo-2H-chromen-7,8-diyl dibenzoate	8528.220	2.069	2.579	-0.509
M24	2-((2-Oxo-2H-chromen-7-yl)oxy)-N-(4-sulfamoylphenyl) acetamide	587.766	3.231	3.377	-0.146
M25	2-((4-Methyl-2-oxo-2H-chromen-7-yl)oxy)-N-(4-sulfamoyl phenyl) acetamide	256.344	3.591	2.913	0.678
M26	2-((2-Oxo-2H-chromen-4-yl)oxy)-N-(4-sulfamoylphenyl) acetamide	460.479	3.337	3.382	-0.045

Table 1. Continued

M27	N-(4-(N-(5-methylisoxazol-3-yl)sulfamoyl)phenyl)-2-((2-oxo-2H-chromen-7-yl)oxy)acetamide	432.675	3.364	3.302	0.062
M28	Ethyl 2-oxo-7-(2-oxo-2-((4-sulfamoylphenyl)amino) ethoxy)-2H-chromene-3-carboxylate	781.263	3.107	3.054	0.053
M29	Ethyl 2-oxo-6-(2-oxo-2-((4-sulfamoylphenyl)amino) ethoxy)-2H-chromene-3-carboxylate	513.401	3.290	3.089	0.201
M30	2,2'-((4-Methyl-2-oxo-2H-chromene-7,8-diyl)bis(oxy) bis (N -(4-sulfamoylphenyl)acetamide)	499.469	3.301	3.555	-0.253
M31	2-((2-Oxo-2H-chromen-7-yl)oxy)-N-phenylacetamide	676.223	3.17	3.32	-0.15
M32	N-(4-chlorophenyl)-2-((2-oxo-2H-chromen-7-yl)oxy) acetamide	375.903	3.425	3.609	-0.184
M33	N-(4-bromophenyl)-2-((2-oxo-2H-chromen-7-yl)oxy) acetamide	351.739	3.454	3.495	-0.041
M34	2-((2-Oxo-2H-chromen-4-yl)oxy)-N-phenylacetamide	327.776	3.484	3.379	0.106
M35[a]	N-(4-chlorophenyl)-2-((2-oxo-2H-chromen-4-yl)oxy) acetamide	405.579	3.392	3.663	-0.271
M36	2-((4-Methyl-2-oxo-2H-chromen-7-yl)oxy)-N-phenyl acetamide	640.295	3.194	2.895	0.299
M37	N-(4-chlorophenyl)-2-((4-methyl-2-oxo-2H-chromen-7-yl)oxy) acetamide	715.033	3.146	3.158	-0.013
M38[a]	N-(4-bromophenyl)-2-((4-methyl-2-oxo-2H-chromen-7-yl) oxy)acetamide	652.205	3.186	3.05	0.136
M39[a]	2-((2-Oxo-4-phenyl-2H-chromen-7-yl)oxy)-N-phenyl acetamide	690.789	3.161	3.219	-0.058
M40[a]	N-(4-chlorophenyl)-2-((2-oxo-4-phenyl-2H-chromen-7-yl) oxy) acetamide	397.720	3.400	3.431	-0.030
M41	3-Hydroxy-2H-chromen-2-one	3.270	5.485	4.78	0.706
M42	4-Hydroxy-2H-chromen-2-one	14.022	4.853	4.407	0.446
M43	6-Hydroxy-2H-chromen-2-one	127.598	3.894	4.378	-0.483
M44[a]	7-Hydroxy-2H-chromen-2-one	26.704	4.573	4.344	0.229
M45	7,8-Dihydroxy-2H-chromen-2-one	1.701	5.769	6.051	-0.281
M46[a]	7,8-Dihydroxy-4-methyl-2H-chromen-2-one	8.329	5.079	5.414	-0.334
M47[a]	7-Methoxy-2H-chromen-2-one	7.420	5.130	3.812	1.317
M48	4-Hydroxy-6-methyl-2H-chromen-2-one	9.353	5.029	4.188	0.841
M49[a]	6-Hydroxy-4-methyl-2H-chromen-2-one	93.762	4.028	3.832	0.196
M50[a]	7-Hydroxy-5-methyl-2H-chromen-2-one	78.039	4.108	3.686	0.422
M51	5,7-Dihydroxy-4-methyl-2H-chromen-2-one	38.601	4.413	4.565	-0.151
M52[a]	4-Hydroxy-6,7-dimethyl-2H-chromen-2-one	32.771	4.485	3.900	0.584
M53[a]	7-Hydroxy-3,4,8-trimethyl-2H-chromen-2-one	96.689	4.015	3.263	0.752
M54[a]	7-Hydroxy-4-methyl-3-phenyl-2H-chromen-2-one	43.186	4.365	3.586	0.779

Table 1. Continued

M55	4,5,7-Trihydroxy-3-phenyl-2H-chromen-2-one	23.333	4.632	4.935	-0.303
M56	7-(Benzyloxy)-4-((phenylamino)methyl-2H-chromen-2-one	20.629	4.686	3.943	0.743
M57	5,7-Bis(benzyloxy)-4-((phenylamino)methyl-2H-chromen-2-one	35.400	4.451	4.453	-0.002
M58	7-(Benzyloxy)-4-(((2,4-dinitrophenyl)amino)methyl)-2H-chromen-2-one	4.7020	5.328	4.973	0.355
M59	7-(Benzyloxy)-4-(((4-hydroxyphenyl)amino)methyl)-2H-chromen-2-one	13.678	4.864	4.975	-0.111
M60	5,7-Bis(benzyloxy)-4-(((4-hydroxyphenyl)amino)methyl)-2H-chromen-2-one	3.314	5.480	5.311	0.169
M61	7-(Benzyloxy)-4-(((2-hydroxyphenyl)amino)methyl)-2H-chromen-2-one	12.431	4.906	4.704	0.201
M62	5,7-Bis(benzyloxy)-4-(((2-hydroxyphenyl)amino)methyl)-2H-chromen-2-one	7.773	5.109	5.054	0.056

[a]Test set.

Table 2. Developed Models for Coumarin Antioxidants by Genetic Function Approximation

S/No	Equation
1	2.074 * HBDCount + 2.067 * AATS3e + 1.085 * ATSC5s - 2.370 * GATS1c + 1.269 * MW + 2.204
2	1.517 * HBDCount + 1.974 * AATS3e + 1.019 * MATS5s - 2.660 * GATS1c + 1.157 * MW + 2.638
3	2.380 * HBDCount + 2.210 * AATS3e + 1.1558 * ATSC5s - 2.148 * GATS1c - 0.560 * MDEO-22 + 1.616 * MW + 1.926
4	2.378 * HBDCount + 2.115 * AATS3e + 0.636 * ATSC5s - 2.072 * GATS1c - 1.212 * GGI3 + 1.646 * MW + 2.777
5	2.155 * HBDCount + 2.060 * AATS3e + 0.606 * MATS5s - 2.141 * GATS1c - 1.463* GGI3 + 1.656 * MW + 3.061

$$h^* = \frac{3(k+1)}{n} \quad (16)$$

The standardized residuals were calculated by using Eq. (17);

$$Standardized\ Residual = \frac{Residual}{RMSE} \quad (17)$$

where RMSE is the root mean square error.

For the detection of the response outliers (Y outliers) and structurally influential chemicals (X outliers) in the developed model the Williams plot was generated [51].

RESULTS AND DISCUSSION

Model Development and Validation

The chemical name of the entire data set together with their IC_{50} and pIC_{50} values are presented in Table 1. An alternative table that gives the chemical structure of these compounds is presented in Table S1 of the supplementary data. The execution of optimization of the molecular

structures and computation of molecular descriptors resulted in 32 quantum chemical descriptors and 1875 molecular descriptors. This gives a total of 1907 descriptors. Upon feature selection, the entire descriptors were reduced to 946. The resulting data set were then normalized. Also, after data division, 49 training set compounds and 13 test set compounds were generated.

Five QSAR models were developed by GFA (Table 2). The molecular descriptors in these models fall under autocorrelation, constitution, topological and molecular distance edge descriptors. A list of all the descriptors and their meanings is presented in Table S2 of the supplementary data. The generated models were employed in the prediction of the training set and test set activities and their corresponding residuals as presented respectively in Tables S3 and S4 of the supplementary data.

The results of internal validation of the developed models are summarized in Table 3. From this table, we observe that all the five models met the requirements for acceptability based on the internal validation. For instance, their R^2 values are by far greater than 0.6. Also the difference between R^2 and Q^2 for model 1, 2, 3, 4 and 5 is 0.036, 0.044, 0.035, 0.039 and 0.040, respectively. This indicates that the number of descriptors in the developed models are acceptable and do not suffer from overfitting. Recall that for model acceptability, the difference between R^2 and R_a^2 should be less than 0.3 for the number of descriptors in the developed model to be acceptable [52,53]. Model 3 gave the highest values for R^2 (0.879), R_a^2 (0.862) and $R_{cv}^2(Q^2)$ (0.845). It also gave the lowest standard error value of 0.435.

The Y-randomization test results are presented in Table 4. A model whose Y-randomization test results satisfies the conditions: $R \geq 0.8$, $R^2 \geq 0.6$, $Q^2 > 0.5$, $cR_p^2 \geq 0.5$ [16], is said to be robust and has good predictive power [30]. The five developed models satisfied these conditions. From Table 4, we observe that model 3 has the highest cR_p^2 value of 0.832, while model 5 has the lowest value of 0.809. This result indicates that model 3 is the most robust among all the five models. Thus, in terms of internal validation results, as presented in Tables 3 and 4, model 3 is the best of the five models.

The external validation results for the developed models are given in Table 5. These models passed all the Golbraikh and Tropsha criteria for the model acceptability: $R_{pred}^2 > 0.5$, $r^2 > 0.6$, $r_m^2 \geq 0.5$, Delta $r_m^2 < 0.2$, $\left|r_0^2 - r_0'^2\right| < 0.3$, $(r^2 - r_0^2)/r^2 < 0.1$ and $0.85 \leq k \leq 1.15$, or $(r^2 - r_0^2)/r^2 < 0.1$ and $0.85 \leq k' \leq 1.15$ [27,39].

From Table 5, model 3 has the highest value for R_{pred}^2 (0.700) and the lowest value for rmsep (0.536). Since very high values of R_{pred}^2 and very low values of rmsep are the main parameters employed in external validation to judge the predictive power of the model, model 3 is thus recognised as the most predictive of all the models. Judging from these excellent internal and external validation results for model 3, we hereby recognise this model as the best of the five models. Subsequently, the results of this model were those represented as the predicted and residual values in Table 1. They were also employed in generating the plot in Fig. (1). This model is presented below:

pIC_{50} = 2.380 * HBDCount + 2.210 * AATS3e + 1.156 * ATSC5s - 2.148 * GATS1c - 0.560 * MDEO-22 + 1.616 * MW + 1.926

$R = 0.938$, $R^2 = 0.879$, $Q^2(R_{cv}^2) = 0.845$, $R_{pred}^2 = 0.700$, $cR_p^2 = 0.832$, $s = 0.435$ and rmsep = 0.536

A plot of the predicted activities against the experimental activities for the Coumarin antioxidants training and test sets are presented in Fig. (1). The observed R^2 values of 0.879 and 0.726 for the training and test sets are indicative of good correlation between the experimental and predicted activities. This is an indication of the good fitting power of our model.

The standard residuals and corresponding leverage values for the training and test set compounds are respectively presented in Tables S5 and S6 of the supplementary data. Also, the plot of standard residuals against leverage values (William's plot) for the coumarin antioxidants is presented in Fig. (2). The computed warning leverage (h*) for the coumarin antioxidant model is 0.429. From Fig. (2), we observe that no response outliers were identified for both training and test set compounds. Thus, all the compounds had cross-validated standard residuals that

Table 3. Summary of Internal Validation Results for Coumarin Antioxidants

Validation Parameters	Model 1	Model 2	Model 3	Model 4	Model 5
Friedman LOF	0.362	0.3904	0.393	0.394	0.396
R-squared	0.868	0.857	0.879	0.879	0.879
Adjusted R-squared	0.852	0.840	0.862	0.862	0.861
Cross validated R-squared	0.831	0.8135	0.845	0.840	0.839
Significant Regression	Yes	Yes	Yes	Yes	Yes
Significance-of-regression F-value	56.309	51.543	50.996	50.844	50.679
Critical SOR F-value (95%)	2.459	2.4591	2.339	2.339	2.339
Replicate points	0.000	0.000	0.000	0.000	0.000
Computed experimental error	0.000	0.000	0.000	0.000	0.000
Lack-of-fit points	43.000	43.000	42.000	42.000	42.000
Min expt. error for non-significant LOF (95%)	0.383	0.398	0.370	0.370	0.371
Standard Error of Estimate	0.451	0.468	0.435	0.436	0.436

[a]The criteria for model acceptability is: $R^2 \geq 0.6$.

Table 4. Results of Y-randomization for Coumarin Antioxidants

Parameters	Model 1	Model 2	Model 3	Model 4	Model 5
R	0.931	0.926	0.938	0.938	0.937
R^2	0.868	0.857	0.879	0.879	0.879
Q^2	0.831	0.813	0.845	0.840	0.839
Random model parameters					
Average r	0.303	0.297	0.304	0.341	0.365
Average r^2	0.101	0.102	0.100	0.120	0.140
Average Q^2	-0.177	-0.193	-0.240	-0.204	-0.197
cR_p^2	0.820	0.812	0.832	0.819	0.809

[a]Model acceptability criteria: $R \geq 0.8$, $R^2 \geq 0.6$, $Q^2 \geq 0.5$, $cR_p^2 \geq 0.5$ [16].

Table 5. External Validation Results for Coumarin Antioxidants

Validation Parameters	Model 1	Model 2	Model 3	Model 4	Model 5
r^2	0.708	0.744	0.726	0.733	0.762
r_0^2	0.706	0.741	0.719	0.724	0.751
Reverse r_0^2	0.633	0.696	0.681	0.697	0.740
r_m^2	0.674	0.703	0.666	0.664	0.684
Reverse r_m^2	0.514	0.580	0.572	0.594	0.65
Average r_m^2	0.594	0.642	0.619	0.629	0.667
Delta r_m^2	0.160	0.123	0.094	0.070	0.033
r^2 - r_0^2/ r^2	0.003	0.004	0.009	0.012	0.014
r^2 - $r_0'^2$ / r^2	0.106	0.065	0.062	0.049	0.028
k	1.08	1.084	1.073	1.104	1.109
k'	0.913	0.911	0.92	0.894	0.891
$\lvert r_0^2$ - $r_0'^2 \rvert$	0.073	0.046	0.038	0.027	0.011
rmsep	0.558	0.541	0.536	0.592	0.585
R_{pred}^2	0.675	0.694	0.700	0.634	0.643

The acceptable threshold values are as follows: $R_{pred}^2 > 0.5$, $r^2 > 0.6$, $r_m^2 \geq 0.5$, Delta $r_m^2 < 0.2$, $\left|r_0^2 - r_0'^2\right| <$ 0.3, and $0.85 \leq k \leq 1.15$, or $(r^2 - r_0^2)/ r^2 < 0.1$ and $0.85 \leq k' \leq 1.15$ [27].

Table 6. Specifications of Coefficient, Standard Error, Mean Effect, Variation Inflation Factor and Degree of Contribution of the Descriptors

S/N0	Descriptor	Coefficient	Standard Error	P-Value	DC	MF	VIF
1	HBDCount	2.380	0.377	1.44E-07	6.306	0.512	1.926
2	AATS3e	2.210	0.373	5.21E-07	5.918	0.475	1.696
3	ATSC5s	1.156	0.302	0.000427	3.825	0.248	1.342
4	GATS1c	-2.148	0.295	5.96E-09	-7.271	-0.462	2.022
5	MDEO-22	-0.560	0.277	0.04915	-2.026	-0.12	1.581
6	MW	1.616	0.319	8.62E-06	5.065	0.347	2.086

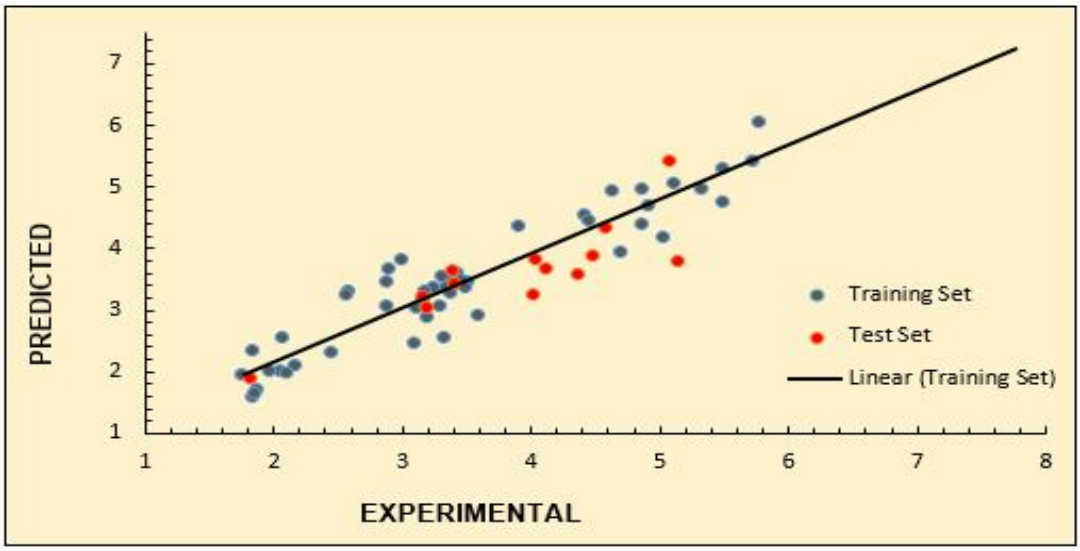

Fig. 1. Plot of the predicted pIC_{50} against experimental values for coumarin antioxidants training and test sets.

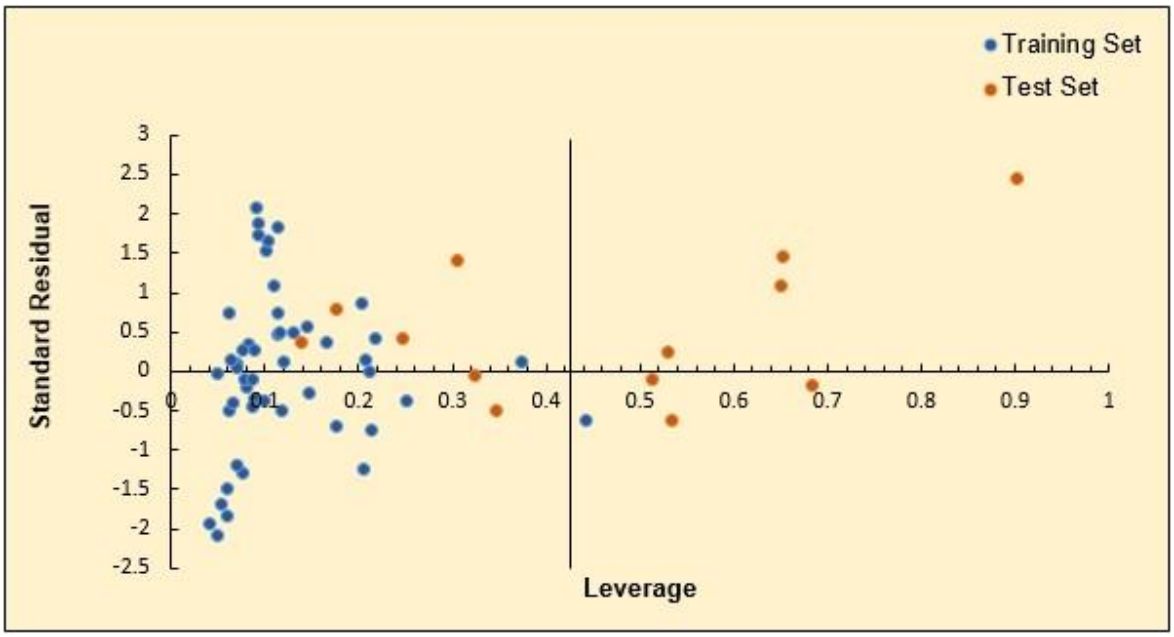

Fig. 2. William's plot for Coumarin antioxidants data set.

are within ±2.5 standard deviation units. We also observe that for the training set, only one structural outlier corresponding to compound No. 30 is identified. For the test, seven structural outliers corresponding to compounds No. 1, 38, 39, 46, 47, 52 and 54 were identified. A close observation indicates that compound No. 30 has a peculiar molecular structure with two similar pendant groups attached to the parent coumarin moiety.

Interpretation and Significance of the Descriptors in the Developed QSAR Model

The results of the computed coefficient, standard error, mean effect, variation inflation factor and degree of contribution of each descriptor in the developed model are presented in Table 6. The lowest value for VIF is 1.342 corresponding to the descriptor ATSC5s, while the highest value is 2.085629042 corresponding to the descriptor MW. These results are within the acceptable range for VIF values. Recall that a VIF value in the range of 1-5 implies that the generated model is acceptable, while a VIF value that is larger than 10 implies that the developed model is unstable and a recheck is necessary [43].

AATS3e (average Broto-Moreau autocorrelation-lag3/ weighted by Sanderson electronegativities), ATSC5s (centered Broto-Moreau autocorrelation-lag5/weighted by I-state) and GATS1c (geary autocorrelation-lag1/weighted by charges). These are 2D autocorrelation descriptors that explain how a considered property such as Sanderson electronegativity, charges, *etc.* is distributed along the topological structure and calculated at a given spatial lag usually ranging from 1 up to 8.

The GATS1c descriptor accounts for the difference on the atomic charge at the topological distance of 1. In our developed model, the DC, MF and coefficient for this descriptor are -7.271, -0.462 and -2.148, respectively. These results indicate that this descriptor has the weakest contribution and lowest significance in the coumarin developed model in addition to being negatively correlated with the antioxidant activity.

HBDCount (hydrogen-bond donor count). This descriptor signifies the number of hydrogen-bond donor sites in the molecule. The developed model indicates that the HBDCount descriptor is positively correlated with the free radical scavenging activity of the Coumarin antioxidants. Thus, increasing the number of hydrogen donor sites in the Coumarin antioxidants moiety increases their antioxidant activities. Also, this descriptor has the highest values for coefficient (2.380) and DC (6.306). These results are in very good agreement with the positive correlation of this descriptor. They indicate the strength of this descriptor towards influencing the antioxidant potency of the coumarins.

MDEO-22 (molecular distance edge between all secondary oxygens). From Table 6, we observe that the MF, DC and coefficient results for the descriptor MDEO-22 are -0.120, -2.0260 and -0.560, respectively. These results suggest that this descriptor has very weak contribution, low influence and negative correlation on the antioxidant activities of the coumarins. Thus, increasing the distance between secondary oxygens among the coumarin antioxidant molecules negatively influence their antioxidant activities.

MW (molecular weight). This descriptor indicates the influence of the molecular weight of each Coumarin molecule on the antioxidant activity. Since MW is positively correlated with the antioxidant activity, increasing the molecular weight of the coumarins increases the antioxidant activity. Also, we observe that this descriptor has a DC value of 5.065, a MF value of 0.347 and a coefficient value of 1.616. With these results, it has a moderate contribution and influence toward the antioxidant activities of the coumarins.

CONCLUSIONS

In this research, we have successfully deigned a QSAR model for the investigation of the free radical scavenging potentials of the coumarin antioxidants by genetic function algorithm. The developed models met all the necessary requirements for model acceptability with highly impressive results (R = 0.938, R^2 = 0.879, $Q^2(R^2_{cv})$ = 0.845, R^2_{pred} = 0.700, cR^2_p = 0.832, s = 0.435 and rmsep = 0.536). The results of the developed models highlight the important descriptors that influence the antioxidant activities of the coumarins. These descriptors in conjunction with the developed model are excellent basis in the design of novel coumarin antioxidants with improved free radical scavenge. The HBDCount descriptor was observed to be the most significant descriptor that influence the free radical scavenging activities of the coumarins.

ACKNOWLEDGMENTS

The authors are grateful to the members of the physical and Theoretical Chemistry unit of the department of Chemistry Ahmadu Bello University Zaria for their cooperation.

REFERENCES

[1] Yang, J.; Liu, G. Y.; Dai, F.; Cao, X. Y.; Kang, Y.; Hu, L. M.; Tang, J. J.; Li, X. Z.; Li, Y.; Jin, X. L.; Zhou, B., Synthesis and biological evaluation of hydroxylated 3-phenylcoumarins as antioxidants and antiproliferative agents. *Bioorg. Med. Chem. Lett.,* **2011**, *21*, 6420-6425, DOI: 10.1016/ j.bmcl.2011.08.090.

[2] Zhang, Y.; Zou, B.; Chen, Z.; Pan, Y.; Wang, H.; Liang, H.; Yi, X., Synthesis and antioxidant activities of novel 4-Schiff base-7-benzyloxy-coumarin derivatives. *Bioorg. Med. Chem. Lett.,* **2011**, *21*, 6811-6815, DOI: 10.1016/j.bmcl.2011.09.029.

[3] Sproll, C.; Ruge, W.; Andlauer, C.; Godelmann, R.; Lachenmeier, D. W., HPLC analysis and safetyassessment of coumarin in foods, *Food Chem.,* **2008**, *109*, 462-469, Doi: 10.1016/ j.foodchem.2007.12.068.

[4] Gutteridge, J. M. C.; Halliwell, B., Antioxidants: Molecules, medicines, and myths. *Biochem. Biophys. Res. Comm.* **2010**, *393*, 561-564, DOI: 10.1016/ j.bbrc.2010.02.071.

[5] Lone, A. A.; Ganai, S. A.; Ahanger, R. A.; Bhat, H. A.; Bhat, T. A.; Wani, I. A., Free radicals and antioxidants: Myths, facts and Mysteries. *Afr. J. Pure Appl. Chem.,* **2013**, *7*, 91-113, DOI: 10.5897/ AJPAC12.074.

[6] Kabel, A. M., Free radicals and antioxidants: Role of enzymes and nutrition. *World J. Nutrition and Health,* **2014**, *2*, 35-38. DOI: 10.12691/jnh-2-3-2.

[7] Brewer, M. S., Natural antioxidants: Sources, compounds, mechanisms of action, and potential applications. *Compr. Rev. Food Sci. Food Saf.,* **2011**, *10*, 221-247, DOI: 10.1111/j.1541-4337.2011.00156.x.

[8] Prakash, G.; Manikandan, R.; Viswanathamurthi, P.; Velmurugan, K.; Nandhakumar, R., Ruthenium(III) S-methylisothiosemicarbazone Schiff base complexes bearing PPh3/AsPh3 coligand: Synthesis, structure and biological investigations, including antioxidant, DNA and protein interaction, and in vitro anticancer activities. *J. Photochem. Photobiol., B,* **2014**, *138*, 63-74, DOI: 10.1016/j.jphotobiol.2014.04.019.

[9] Ilies, D. C.; Shova, S.; Radulescu, V.; Pahontu, E.; Rosu, T., Synthesis, characterization, crystal structure and antioxidant activity of Ni(II) and Cu(II) complexes with 2-formilpyridine N(4)-phenylthio-semicarbazone. *Polyhedron,* **2015**, *97*, 157-166. DOI: 10.1016/j.poly.2015.05.009.

[10] Aziz, A. N.; Taha, M.; Ismail, N. H.; Anouar, E. H.; Yousuf, S.; Jamil, W.; Awang, K.; Ahmat, N.; Khan, K. M.; Kashif, S. M., Synthesis, crystal structure, DFT studies and evaluation of the antioxidant Activity of 3,4-dimethoxybenzenamine Schiff bases. *Molecules* **2014**, *19*, 8414-8433, DOI: 10.3390/ molecules19068414.

[11] Zhang, Y.; Fang, Y.; Liang, H.; Wang, H.; Huc, K.; Liu, X.; Yi, X.; Peng, Y., Synthesis and antioxidant activities of 2-oxo-quinoline-3-carbaldehyde Schiff-base derivatives. Bioorg. Med. Chem. Lett., 2013, 23 107-111, DOI: 10.1016/j.bmcl.2012.11.006.

[12] [12] Parameswaran, K.; Sivaguru, P.; Lalitha, A., Synthesis of novel bis(pyrimido[5,4-c]quinoline-2,4(1H,3H)-dione) and its derivatives: Evaluation of their antioxidant properties. *Bioorg. Med. Chem. Lett.,* **2013**, *23*, 3873–3878, DOI: 10.1016/ j.bmcl.2013.04.068.

[13] [13] Arthur, D. E.; Uzairu, A.; Mamza, P.; Abechi, S., Quantitative structure–activity relationship study on potent anticancer compounds against MOLT-4 and P388 leukemia cell lines. *J. Adv. Res.,* **2016**, *7*, 823-837, DOI: 10.1016/j.jare.2016.03.010.

[14] Ferna´ndez, M.; Caballeroa, J.; Tundidor-Camba, A., Linear and nonlinear QSAR study of N-hydroxy-2-[(phenylsulfonyl)amino]acetamide derivatives as matrix metalloproteinase inhibitors. *Bioorg. Med. Chem.,* **2006**, *14*, 4137-4150, DOI: 10.1016/ j.bmc.2006.01.072.

[15] Solomon, K. A.; Sundararajan, S.; Abirami, V., QSAR studies on N-aryl derivative activity towards alzheimer's disease. *Molecules* **2009**, *14*, 1448-1455, DOI: 10.3390/molecules14041448.

[16] Mitra, I.; Saha, A.; Roy, K., Chemometric modeling of free radical scavenging activity of flavone derivatives. *Eur. J. Med. Chem.,* **2010**, *45,* 5071-5079. DOI: 10.1016/j.ejmech.2010.08.016.

[17] Sanja, C.; Franci, K.; Milka, M., Evaluation of the

antioxidant activity of a series of 4-methylcoumarins using different testing methods. *Food Chem.,* **2012**, *133*, 930-937. DOI: 10.1016/j.foodchem.2012.02.006.

[18] Saeedi, M; Goli, F; Mahdavi, M; Dehghan, G; Faramarzi, M. A; Foroumadi, A.; Shafiee, A., Synthesis and biological investigation of some novel sulfonamide and amide derivatives containing coumarin moieties. *Iran. J. Pharm. Res.,* **2014**, *13*: 881-892.

[19] Gacche, R. N.; Jadhav, S. G., Antioxidant activities and cytotoxicity of selected coumarin derivatives: preliminary results of a structure activity relationship study using computational tools. *J. Exp. Clin. Med.,* **2012**, *4*, 165-169, DOI: 10.1016/j.jecm.2012.04.007.

[20] Zhang, Y.; Zou, B.; Chen, Z.; Pan, Y.; Wang, H.; Liang, H.; Yi, X., Synthesis and antioxidant activities of novel 4-Schiff base-7-benzyloxy-coumarin derivatives. *Bioorg. Med. Chem. Lett.,* **2011**, *21*, 6811-6815, DOI: 10.1016/j.bmcl.2011.09.029.

[21] ChemDraw Ultra 6.0 and Chem3D Ultra version 12.0, Cambridge Soft Corporation, Cambridge, USA, 2010.

[22] Spartan 14v1.1.2. Wavefunction, Inc. USA, 2014.

[23] Yap, C. W., PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.,* **2011**, *32*, 1466-1474, DOI: 10.1002/jcc.21707.

[24] Ballabio, D.; Consonni, V.; Mauri, A.; Claeys-Bruno, M.; Sergent, M.; Todeschini, R., A novel variable reduction method adapted from space-filling designs. *Chemom. Intell. Lab. Syst.,* **2014**, *136*, 147-154. DOI.org/10.1016/j.chemolab.2014.05.010.

[25] Ambure, P.; Aher, R. B.; Gajewicz, A.; Puzyn, T., "NanoBRIDGES" software: Open access tools to perform QSAR and nano-QSAR modeling. *Chemom. Intell. Lab. Syst.,* **2015**, *147*: 1-13, DOI: 10.1016/ j.chemolab.2015.07.007.

[26] Brignole, M.; Auricchio, A.; Baron-Esquivias, G.; Bordachar, P.; Boriani, G.; Breithardt, O., *et al*., ESC Guidelines on cardiac pacing and cardiac resynchronization therapy. *Eur Heart J.,* **2013**, *34*, 2281-329, DOI: 10.1093/eurheartj/eht150.

[27] Golbraikh, A.; Tropsha, A., Beware of q2! *J. Mol. Graph. Model.,* **2002**, *20*, 269-276, DOI.org/ 10.1016/ S1093-3263(01)00123-1.

[28] Todd, M. M; Harten, P.; Douglas, M. Y.; Muratov, E. N.; Golbraikh, A.; Zhu, H.; Tropsha, A., Does rational slection of training and test sets improve the outcome of QSAR modeling? *J. Chem. Inf. Model.*, **2012**, *52*, 2570-2578, DOI.org/10.1021/ci300338w.

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

[30] Ravichandran, V.; Harish, R.; Abhishek, J.; Shalini, S.; Christapher, P. V.; Ram, K. A., Validation of QSAR models-strategies and importance. *Int. J. Drug Des. Discovery,* **2011**, *2*, 511-519.

[31] Mitra, I.; Saha, A.; Roy, K., Chemometric QSAR modeling and in silico design of antioxidant NO donor phenols. *Sci Pharm.,* **2011**, *79*, 31-57, DOI: 10.3797/scipharm.1011-02.

[32] Rudra, N. D.; Kunal, R., Development of classification and regression models for vibrio fischeri toxicity of ionic liquids: Green solvents for the future. *Toxicol. Res.,* **2012**, *1*, 186-195, DOI.org/ 10.1039/c2tx20020a.

[33] Eriksson, L.; Jaworska, J.; Worth, A.; Cronin, M.; McDowell, R. M.; Gramatica, P., Methods for reliability, uncertainty assessment, and applicability evaluations of regression based and classification QSARs. *Environ. Health Perspect.,* **2003**, *111*, 1361-1375,

[34] Tropsha, A.; Gramatica, P.; Gombar, V. K., The importance of being Earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.,* **2003**, *22*, 69-76.

[35] Supratik, K.; Kunal, R., Development and validation of a robust QSAR model for prediction of carcinogenicity of drugs. *Indian J. Biochem. Biophys.,* **2011**, *48*, 111-122.

[36] Roy, P. P.; Roy, K., On some aspects of variable selection for partial least squares regression models. *QSAR Comb. Sci.,* **2008**, *27*, 302-313.

[37] Roy, K.; Kar, S.; Das, R. N., A primer on QSAR/QSPR modeling, springer briefs in molecular science: *Statistical Methods in QSAR/QSPR.* **2015**, 37-59. DOI.org/10.1007/978-3-319-17281-1_2.

[38] Roy, K.; Mitra, I., On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. *Comb. Chem. High Throughput Screening,* **2011**, *14*, 450-474, DOI.org/10.2174/138620711795767893.

[39] Roy, K.; Chakraborty, P.; Mitra, I.; Ojha, P. K.; Kar, S.; Das, R. N., Some case studies on application of "rm2" metrics for judging quality of quantitative structure-activity relationship predictions: Emphasis on scaling of response data. *J. Compu. Chem.,* **2013**, *34*, 1071-1082, DOI.org/10.1002/jcc.23231.

[40] Tropsha, A., Best practices for QSAR model development, validation, and exploitation. Mol. Inf., 2010, 29, 476-488. DOI.org/10.1002/minf.201000061.

[41] Todeschini R Milano Chemometrics. Italy (personal communication), 2010.

[42] Pravin, A., Drug Theoretics & Cheminformatics (DTC) Laboratory, Jadavpur University, 2013.

[43] Agrawal, V.K.; Khadikar, P.V., QSAR Prediction of Toxicity of Nitrobenzenes. *Bioorg. Med. Chem.,* **2001**, *9*, 3035-3040, Doi.org/10.1016/S0968-0896(01)00211-5.

[44] Baumann, K., Chance Correlation in Variable Subset Regression: Influence of the Objective Function, the Selection Mechanism, and Ensemble Averaging. *QSAR Comb. Sci.,* **2005**, *24*, 1033-1046, DOI: 10.1002/qsar.200530134.

[45] Asadollahi, T.; Dadfarnia, S.; Shabani, A. M.; Ghasemi, J. B.; Sarkhosh, M., QSAR Models for $CXCR_2$ receptor antagonists based on the genetic algorithm for data preprocessing prior to application of the PLS linear regression method and design of the new compounds using in silico virtual screening. *Molecules*, **2011**, *16*, 1928-1955; DOI: 10.3390/molecules16031928.

[46] Netzeva, T. I; Worth, A; Aldenberg, T; Benigni, R; Cronin, M. T; Gramatica, P; Jaworska, J. S; Kahn, S; Klopman, G; Marchant, C. A; Myatt, G; Nikolova-Jeliazkova, N; Patlewicz, G. Y; Perkins, R; Roberts, D; Schultz, T; Stanton, D. W; van de Sandt, J. J; Tong, W; Veith, G.; Yang, C., Current status of methods for defining the applicability domain of (Quantitative) structure-activity relationships. The report and recommendations of ECVAM Workshop 52. *Altern Lab Anim,* **2005**, *33*, 155-173.

[47] Eriksson, L; Jaworska, J; Worth, A. P; Cronin, M. T. D; McDowell, R. M.; Gramatica, P., Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-baes QSARs. *Environ. Health Persp.,* **2003**, *111*, 1361-1375.

[48] Nandi, S; Monesi, A; Drgan, V; Merzel, F.; Novič, M., Quantitative structure-activation barrier relationship modeling for Diels-Alder ligations utilizing quantum chemical structural descriptors. *Chem. Cent. J.,* **2013**, *7*, 1-13, DOI.org/10.1186/1752-153X-7-171.

[49] Gramatica, P; Giani, E.; Papa, E., Statistical external validation and consensus modeling: A QSPR case study for KOC prediction. *J. Mol. Graphics Modell.,* **2007**, *25*, 755-66. DOI: 10.1016/j.jmgm.2006.06.005.

[50] Gramatica, P., Chemiometric methods and theoretical molecular descriptors in predictive QSAR modeling of the environmental behavior of organic pollutants. In T. Puzyn *et al.* (Eds.), *Recent Advances in QSAR Studies*, **2010**, 327-366, DOI: 10.1007/978-1-4020-9783-6_12.

[51] Sharma, B. K.; Singh, P., Chemometric descriptor based QSAR rationales for the MMP-13 inhibition activity of non-zinc-chelating compounds. *Med chem.,* **2013**, *3*, 168-178. Doi: 10.4172/2161-0444.1000134.

[52] Kiralj, R.; Ferreira, M. M. C., Basic validation procedures for regression models in QSAR and QSPR studies: Theory and application. *J. Braz. Chem. Soc.,* **2009**, *20*, 770-787.

[53] Roy, P. P. J.; Leonard, T.; Roy, K., Exploring the impact of size of training sets for the development of predictive QSAR models. *Chemometrics and Intelligent Laboratory Systems.* **2008**, *90*, 31-42, DOI: 10.1016/j.chemolab.2007.07.004.