

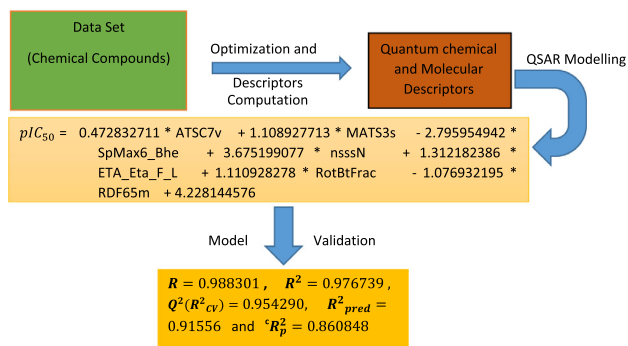


Original Article

Evaluation of the antioxidant properties of curcumin derivatives by genetic function algorithm

Ikechukwu Ogadimma Alisi ^{a,*}, Adamu Uzairu ^b, Stephen Eyije Abechi ^b, Sulaiman Ola Idris ^b^a Department of Applied Chemistry, Federal University Dutsinma, Katsina State, Nigeria^b Department of Chemistry, Ahmadu Bello University Zaria, Kaduna State, Nigeria

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 17 November 2017

Revised 24 February 2018

Accepted 7 March 2018

Available online 28 March 2018

Keywords:

Antioxidants

Curcumins

Descriptors

Free radicals, GFA, model validation

QSAR

ABSTRACT

The prevalence of degenerative diseases in recent time has triggered extensive research on their control. This condition could be prevented if the body has an efficient antioxidant mechanism to scavenge the free radicals which are their main causes. Curcumin and its derivatives are widely employed as antioxidants. The free radical scavenging activities of curcumin and its derivatives have been explored in this research by the application of quantitative structure activity relationship (QSAR). The entire data set was optimized at the density functional theory (DFT) level using the Becke's three-parameter Lee-Yang-Parr hybrid functional (B3LYP) in combination with the 6-311G* basis set. The training set was subjected to QSAR studies by genetic function algorithm (GFA). Five predictive QSAR models were developed and statistically subjected to both internal and external validations. Also the applicability domain of the developed model was accessed by the leverage approach. Furthermore, the variation inflation factor, (VIF), mean effect (MF) and the degree of contribution (DC) of each descriptor in the resulting model were calculated. The developed models met all the standard requirements for acceptability upon validation with highly impressive results ($R = 0.965$, $R^2 = 0.931$, $Q^2(R^2_{CV}) = 0.887$, $R^2_{pred} = 0.844$, $R_p^2 = 0.842$, $s = 0.226$, $rmsep = 0.362$). Based on the results of this research, the most crucial descriptor that influence the free radical scavenge of the curcumins is the nsssN (count of atom-type E-state: >N-) descriptor with DC and MF values of 12.980 and 0.965 respectively.

© 2018 Production and hosting by Elsevier B.V. on behalf of Cairo University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Curcumin [(1E,6E)-1,7-bis(4-hydroxy-3-methoxyphenyl)hepta-1,6-diene-3,5-dione] is a naturally occurring phenolic compound which is responsible for the yellowish orange colour present in

Peer review under responsibility of Cairo University.

* Corresponding author.

E-mail addresses: ikeogadialisi@gmail.com, ialisi@fudutsinma.edu.ng (I.O. Alisi).<https://doi.org/10.1016/j.jare.2018.03.003>

2090-1232/© 2018 Production and hosting by Elsevier B.V. on behalf of Cairo University.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

turmeric (*Curcuma longa* L.) [1,2]. Turmeric is a herbaceous plant of the Zingiberaceae family. It is a spice that has long been used to enhance the flavour of foods in the form of “curry leaf or powder”. The broad range of biological and pharmacological activities of curcumin and its derivatives have been widely explored and reported. These include antimetastatic activities by differentially decreasing the extracellular matrix (ECM) degradation enzyme secretion from invasive cells [3], antibacterial activities [4], anticancer activities [5] antitumor activities [6] antimalarial activities [7] and antioxidant activities [8–11].

Antioxidants are substances that employ various mechanisms to scavenge free radicals by inhibiting their formation or interrupting their propagation [12]. Thus, through various mechanisms antioxidants have the ability to inhibit the adverse effects of oxidative stress.

Free radicals are atoms or molecules that contain one or more unpaired electrons in their orbitals [13]. The high reactivity of free radicals is attributed to the presence of these unpaired electrons. Free radicals produced in the human system include reactive oxygen species (ROS) such as hydroxyl radical $\cdot\text{OH}$, superoxide anion radical O_2^- and hydroperoxyl radical $\text{HOO}\cdot$. Also produced are reactive nitrogen species (RNS) such as nitric oxide radical $\text{NO}\cdot$ and nitrogen dioxide radical NO_2 . Low concentrations of these radicals are essential for cell physiological processes. When the level of free radicals generated become higher than they can be scavenged, excess free radicals are produced which give rise to a condition termed “oxidative stress”. Oxidative stress is responsible for degenerative diseases in the human system such as cancer, cardiovascular diseases and immune system decline [13]. Under normal conditions, the human system maintains a balance between the level of these free radicals and antioxidants.

Various methods have been adopted to evaluate the antioxidant activities of various substances. These methods include the 2,2-diphenyl-1-picrylhydrazyl (DPPH) free radical scavenging assay [14]; the superoxide anion scavenging activity [14]; the oxygen radical absorbance capacity by fluorescence (ORAC-FL) method [15]; and the 2,2'-azino bis (3-ethylbenzothiazoline-6-sulfonate) (ABTS) cation radical assay [16]. The DPPH free radical scavenging assay is a widely used method that depends on the hydrogen donating ability of the tested compound in which the stable DPPH free radical is converted to 2,2'-diphenyl-1-picrylhydrazine [17]. This reaction which is accompanied by a change in colour from deep-violet to light-yellow is the preferred method in this research.

The development of predictive Quantitative Structure Activity Relationship (QSAR) models for chemical compounds by computational methods, has received great attention in recent time [18]. QSAR is a method widely employed in the correlation of the biological and pharmacological activities of compounds with their molecular structures [19]. It provides the basis for understanding the influence of the chemical structure of compounds on their biological activities, thus facilitating the link for rational design of new compounds with improved biological activities [20]. This method has been applied for modelling the antioxidant activities of compounds [19].

In this research, the antioxidant activities of the curcumin derivatives based on the DPPH assay were investigated. A data set of 47 curcumin derivatives was optimized and submitted for the generation of quantum chemical and molecular descriptors. The optimized structures were employed in the generation of QSAR models by Genetic Function Algorithm (GFA). The data set was divided into training and test sets. The training set was employed in model development, while the test set was used to validate the developed models. Various validation tests were conducted. These include: Internal validations, external validations and y-randomization tests. The assessment of the applicability domain of the model was executed by the leverage approach. To investigate

the strength of the descriptors in the developed model, various parameters such as variation inflation factor (VIF), mean effect and degree of contribution of the descriptors were calculated.

Computational methods

Data set collection and optimization

The data set of 47 curcumin antioxidants and their corresponding experimental DPPH IC_{50} activities in μM were obtained from literature [8–11]. The ChemBioDraw Ultra (version 12.0) [21], was employed in drawing the molecular structures. These structures were subjected to energy minimization and subsequently optimized using Spartan 14v112 program package [22]. The density functional theory (DFT) level was employed [23], using Becke's three-parameter Lee-Yang-Parr hybrid functional (B3LYP) in combination with the 6-311G* basis set without symmetry constraints [24,25]. This optimization condition has been recognised to give a reliable estimate of the antioxidant properties of molecules. Also, due to the presence of polarization functions, it has been observed to give a better description of the electronic interactions outside the nucleus [26]. Full optimization of the geometries and energies for all of the studied molecules was carried out in the gas phase.

Descriptors calculation

The optimized molecules were converted to standard database format (sdf) files and submitted for the generation of molecular descriptors using “PaDel-Descriptor (version 2.20)” program package [27]. These descriptors were combined to the quantum chemical descriptors obtained during optimization of the molecules.

Data pre-treatment, normalization and division

The resulting data after optimization were subjected to pre-treatment using “Data Pre-Treatment GUI 1.2” program [28]. Data normalization was achieved by scaling between the intervals 0–1 [29]. The entire data set was divided into training and test sets by the application of Kennard Stone algorithm using the program “Dataset Division GUI 1.2” [30].

Development of the QSAR model

The training set was employed in the development of the QSAR model by genetic function approximation (GFA) where the molecular descriptors (independent variables) and the pIC_{50} values (dependent variables) were subjected to multivariate analysis using the material studio program package. The GFA was performed by using 50,000 crossovers, a smoothness value of 1.00 and an initial of five and a maximum of ten terms per equation. By employing GFA the Friedman lack-of-fit (LOF) value was calculated. LOF which measures the fitness of the model was calculated using Eq. (1).

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

where

SSE is the sum of squares of errors.

c is the number of basis functions terms in the model, ignoring the constant term.

d is a user-defined smoothing parameter which was set to 0.5.

p is the total number of descriptors contained in all model terms outside the constant term.

M is the number of samples in the training set [31].

Internal validation of the developed models

The leave-one-out (LOO) cross-validation method was employed to internally validate the developed models. This

method involves the elimination of one compound from the data set and building the model using the rest of the compounds. The resulting model thus formed is employed to predict the activity of the eliminated compound. This procedure is repeated until all the compounds have been eliminated [32].

The internal validation parameters calculated include:

The Cross-validated squared correlation coefficient, $R_{cv}^2(Q^2)$ which was calculated using Eq. (2).

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

Y_{obs} = Observed activity of the training set compounds.

Y_{pred} = Predicted activity of the training set compounds.

\bar{Y} = Mean observed activity of the training set compounds.

The adjusted R^2 (R_a^2) overcomes the drawbacks associated with R^2 . Thus it is a modification of R^2 [33]. The R_a^2 values were calculated using Eq. (3).

$$R_a^2 = \frac{(n-1)R^2 - p}{n - p - 1} \quad (3)$$

where p is the number of predictor variables used to develop the model.

The variance ratio, F value was also calculated using Eq. (4):

$$F = \frac{\sum (Y_{cal} - \bar{Y})^2 / p}{\sum (Y_{obs} - Y_{cal})^2 / (N - p - 1)} \quad (4)$$

This parameter represents the ratio of regression mean square to deviations mean square. It is employed to judge the overall significance of the regression coefficients.

For the calculation of the Standard Error of estimate (s), Eq. (5) was employed.

$$s = \sqrt{\frac{RSS}{n - p'}} \quad (5)$$

where RSS is the sum of squares of the residuals between the experimental and predicted activities for the training set. p' is the number of model variables plus one. n is the number of objects used to calculate the model [34].

Randomization test

The robustness of the models were checked using the y-randomization test. It was applied by permuting the activity values with respect to the descriptor matrix. The R_p^2 parameter gives the deviation in the values of the squared mean correlation coefficient of the randomized model (R_r^2) from the squared correlation coefficient of the non-random model (R^2) as presented in Eq. (6) [35].

$$R_p^2 = R^2 \times \sqrt{R^2 - R_r^2} \quad (6)$$

For randomized models, the average value of R_r^2 is zero which will make the value of R_p^2 to be equal to the value of R^2 in an ideal situation (Eq. (6)). In 2010, Todeschini [36] suggested a correction for R_p^2 as presented in Eq. (7).

$$cR_p^2 = R \times \sqrt{R^2 - R_r^2} \quad (7)$$

The program package “MLR Y-Randomization Test 1.2” was employed in the computation of the y-randomization test parameters [37].

External model validation

The developed models were subjected to external validation in order to ascertain their predictive capacity. Among the calculated external validation parameters was the predicted squared correlation coefficient, R^2 (R^2_{pred}) value (Eq. (8)). This parameter was calculated from the predicted activity of all the test set compounds.

$$R^2_{pred} = 1 - \frac{\sum (Y_{pred(Test)} - Y_{(Test)})^2}{\sum (Y_{(Test)} - \bar{Y}_{(Training)})^2} \quad (8)$$

where $Y_{pred(Test)}$ is the predicted activity values of the test set compounds, and $Y_{(Test)}$ indicates their observed activity values. $\bar{Y}_{(Training)}$ is the mean activity value of the training set. From Eq. (7), the computed R^2_{pred} value is controlled by $\sum (Y_{(Test)} - \bar{Y}_{(Training)})^2$. This may result in considerable difference between the observed and predicted results even though the overall intercorrelation may be quite encouraging.

For a better measure of external predictivity of the developed model, a modified R^2 denoted by r_m^2 as defined in Eq. (9), is thus introduced.

$$r_m^2 = r^2 \left(1 - \sqrt{r^2 - r_0^2} \right) \quad (9)$$

where r_0^2 is the squared correlation coefficients of linear relations between the observed and predicted results when zero is the intercept, while, r^2 is the squared correlation coefficients of linear relations between the observed and predicted results when the intercept is not set to zero. When the axes are interchanged, the parameter r_m^2 is obtained as defined by Eq. (10).

$$r_m^2 = r^2 \times \left(1 - \sqrt{r^2 - r_0^2} \right) \quad (10)$$

The program pack “DTC-MLR Plus Validation GUI 1.2” was employed in the calculation of the external validation results [38].

Estimation of the variation inflation factor (VIF)

The multi-collinearity, among the descriptors in the developed model were investigated by computing their variation inflation factors (VIF) as presented in Eq. (11).

$$VIF = \frac{1}{1 - r^2} \quad (11)$$

where r is the correlation coefficient of multiple regressions of one descriptor with the other descriptors in the model.

Estimation of the mean effect and degree of contribution of the descriptors

The mean effect (MF) of each descriptor in the developed model was calculated using Eq. (12).

$$MF_j = \frac{\beta_j \sum_{i=1}^n d_{ij}}{\sum_j \beta_j \sum_1^n d_{ij}} \quad (12)$$

where MF_j represents the mean effect for the considered descriptor j . β_j is the coefficient of the descriptor j . d_{ij} is the value of the target descriptors for each molecule. m is the number of descriptors in the model. The relative significance and contribution of a given descriptor compared with the other descriptors in the model is described by the magnitude of MF, while the sign of its MF indicates the variation direction with respect to a given descriptor for the considered molecules. Also the degree of contribution (DC) was calculated for each descriptor in the developed model.

Applicability domain investigation

The applicability domain of a QSAR model is the response and chemical structure space in which the model makes predictions with a given reliability. Predictions outside the applicability domain of the developed model are considered unreliable.

The leverage approach was employed in the assessment of the applicability domain of the developed QSAR model. The leverage value of each compound in the dataset X , was calculated by obtaining the leverage (hat) matrix (H) as defined by Eq. (13).

$$H = X(X^T X)^{-1} X^T \quad (13)$$

where X is the two-dimensional $n \times k$ descriptor matrix of the training set compounds with n compounds and k descriptors, while X^T is the transpose of X .

The leverage h_i of the i th compound is the i th diagonal element of H as defined in Eq. (14).

$$h_i = x_i(X^T X)^{-1} x_i^T \quad (i = 1, \dots, m) \quad (14)$$

The leverage threshold, h^* , is the limit of normal values for X outliers Eq. (15).

$$h^* = \frac{3(k+1)}{n} \quad (15)$$

The standard residuals for each compound in the data set were also calculated (Eq. (16)).

$$\text{Standard Residual} = \frac{\text{Residual}}{\text{RMSE}} \quad (16)$$

where $RMSE$ is the root mean square error. Furthermore, the Williams plot which is a plot of standard residuals versus leverage values, (Williams plot) is used to detect the response outliers and structurally influential chemicals in the model [39]. Response outliers are those compounds with standard residuals greater than 2.5 standard deviation units. While Structural outliers are those with $h > h^*$, [40].

Results and discussion

Descriptors calculation, data pre-treatment and division

Table 1 gives the chemical name of the entire data set together with their IC_{50} and pIC_{50} values. The optimized structures of the entire data set are presented in Fig. S1 of the supplementary data. Also, the bond lengths, bond angles and dihedral angles of representative members of the data set with impressive antioxidant activities were calculated (Table S1). A total of 1907 descriptors were generated of which 32 of them are quantum chemical descriptors obtained from the DFT calculation, while the other 1875 are molecular descriptors. These descriptors include constitutional, topological, radial distribution function (RDF), 3D-Morse, and Geometrical descriptors. The application of data pre-treatment resulted in 1044 descriptors. Pre-treatment ensures that descriptors with constant values and pairs of variables with correlation coefficients greater than 0.9 are removed. Data division produced 37 training set compounds and 10 test set compounds.

Model development and validation

Five QSAR models were developed as presented in Table 2. The descriptors in these models can broadly be categorized into Auto-correlation, Burden Modified Eigenvalues, Electrotopological State Atom Type, Extended Topochemical Atom, PaDEL Rotatable Bonds Count, Topological Distance Matrix and Radial Distribution Function Descriptors as presented in Table S2 of the supplementary data. Also the developed models were employed in predicting the antiox-

idant activities of the training set and test set compounds as presented in Tables S3 and S4 respectively of the supplementary data.

The summary of the internal validation results for the developed models are presented in Table 3. All the five models satisfied the necessary internal validation requirements for acceptability with R^2 values well above the threshold value of 0.6. This parameter measures the variation between the calculated data and the observed data. Thus it measures the fitting power of the model. The computed R^2 values were very close to unity which represents a perfect fit. Results of other validation parameters were also quite encouraging. From literature 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 [41]. From Table 3, the differences between R^2 and R_a^2 for models 1, 2, 3, 4 and 5 are 0.015, 0.016, 0.016, 0.017 and 0.017 respectively. Thus the number of descriptors in the developed models are within the acceptable range. Based on the results in Table 3, model 3 recorded the highest values for R^2 and R_a^2 of 0.932 and 0.916 respectively. Also this model has the lowest standard error value of 0.223, while model 1 has the highest Q^2 value of 0.892.

The y-randomization results for all the models are presented in Table 4. For the acceptance of a Y-randomization test, the results must satisfy the condition: $R \geq 0.8$, $R^2 \geq 0.6$, $Q^2 > 0.5$, ${}^cR_p^2 \geq 0.5$ [35]. The five models satisfied this condition appreciably with model 4 having the highest ${}^cR_p^2$ value of 0.842, while model 5 has the lowest value of 0.826. The y-randomization test dictates that the predictive power of a model is poor when the observations are not sufficiently independent of each other [42]. This is actually reflected in the value of ${}^cR_p^2$ which must satisfy the condition: ${}^cR_p^2 \geq 0.5$. Thus the generated results were not the mere outcome of chance. Judging from the results of internal validation and y-randomization tests 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 developed models passed all the Golbraikh and Tropsha criteria for model acceptability which dictates that: $R_{pred}^2 > 0.5$, $r^2 > 0.6$, $r_m^2 \geq 0.5$, $\Delta r_m^2 < 0.2$, $|r_0^2 - r_0'^2| < 0.3$, $(r^2 - r_0^2)/r^2 < 0.1$ and $0.85 \leq k \leq 1.15$, or $(r^2 - r_0'^2)/r^2 < 0.1$ and $0.85 \leq k' \leq 1.15$ [29]. Also the results of the external validation were all within the recommended threshold values for the various validation parameters as shown in Table 5. Thus all the five models can safely be employed in predicting the activities of new set of curcumin antioxidants based on their highly encouraging external validation results.

In terms of the external validation results, model 1 has the highest R_{pred}^2 value of 0.853 and lowest rmsep value of 0.352. These results are closely followed by the results generated for model 4. Model 4 has R_{pred}^2 value of 0.844, rmsep value of 0.362, the lowest delta r_m^2 value of 0.025 and a higher number of seven descriptors in the developed model in comparison to model 1. In addition, model 4 has the highest values for r^2 (0.864), r_0^2 (0.861) and Reverse r_0^2 (0.857). Based on the results of internal and external validation, model 4 is thus recognized as the best of the five models. This model 4 is represented as:

$$pIC_{50} = 0.473 * ATSC7v + 1.109 * MATS3s - 2.796 * SpMax6_Bhe + 3.675 * nssn + 1.312 * ETA_Eta_F.L + 1.111 * RotBtFrac - 1.077 * RDF65m + 4.228$$

$$R = 0.965, R^2 = 0.931, Q^2(R_{cv}) = 0.887, R_{pred}^2 = 0.844, {}^cR_p^2 = 0.842, s = 0.226, \text{rmsep} = 0.362$$

Table 1
Chemical name of curcumin derivatives data set and their antioxidant activities.

Comp no	Compound	IC ₅₀	pIC ₅₀		
			Observed	Predicted	Residual
M01 ^a	(1E,6E)-1,7-bis(4-hydroxy-3-methoxyphenyl)hepta-1,6-diene-3,5-dione	11.048	4.957	4.316	0.641
M02	(1E,6E)-1,7-bis(3,4-dihydroxyphenyl)hepta-1,6-diene-3,5-dione	2.290	5.640	5.407	0.233
M03	(1E,6E)-1,7-bis(4-hydroxy-3,5-dimethoxyphenyl)hepta-1,6-diene-3,5-dione	9.696	5.013	4.984	0.030
M04	(1E,4E)-1,5-bis(4-hydroxy-3-methoxyphenyl)penta-1,4-dien-3-one	14.898	4.827	4.883	-0.057
M05	(1E,4E)-1,5-bis(3,4-dihydroxyphenyl)penta-1,4-dien-3-one	2.873	5.542	5.660	-0.119
M06	(1E,4E)-1,5-bis(4-hydroxy-3,5-dimethoxyphenyl)penta-1,4-dien-3-one	14.710	4.832	4.771	0.061
M07	(2E,5E)-2,5-bis(4-hydroxy-3-methoxybenzylidene)cyclopentanone	35.873	4.445	4.867	-0.422
M08	(2E,5E)-2,5-bis(3,4-dihydroxybenzylidene)cyclopentanone	3.088	5.510	5.644	-0.134
M09	(2E,5E)-2,5-bis(4-hydroxy-3,5-dimethoxybenzylidene)cyclopentanone	6.517	5.186	5.215	-0.029
M10	(2E,6E)-2,6-bis(4-hydroxy-3-methoxybenzylidene)cyclohexanone	25.220	4.598	4.278	0.321
M11 ^a	(2E,6E)-2,6-bis(3,4-dihydroxybenzylidene)cyclopentanone	4.436	5.353	5.265	0.088
M12	(2E,6E)-2,6-bis(4-hydroxy-3,5-dimethoxybenzylidene)cyclohexanone	22.884	4.640	4.711	-0.071
M13	(1E,4E)-1,5-bis(3,4-dimethoxyphenyl)penta-1,4-dien-3-one	32.612	4.487	4.763	-0.277
M14	(1E,4E)-1,5-bis(3-hydroxy-4-methoxyphenyl)penta-1,4-dien-3-one	16.347	4.787	4.936	-0.149
M15 ^a	(1E,4E)-1,5-bis(4-hydroxy-3-methoxyphenyl)penta-1,4-dien-3-one	3.016	5.521	4.884	0.636
M16	(1E,4E)-1-(3,4-dimethylphenyl)-5-(4-hydroxy-3-methoxyphenyl)penta-1,4-dien-3-one	12.785	4.893	4.577	0.316
M17	(1E,4E)-1-(3,4-dimethoxyphenyl)-5-(4-hydroxy-3-methoxyphenyl)penta-1,4-dien-3-one	6.709	5.173	4.786	0.388
M18	(1E,4E)-1-(3-hydroxy-4-methoxyphenyl)-5-(3,4,5-trimethoxyphenyl)penta-1,4-dien-3-one	12.734	4.895	4.848	0.047
M19	(1E,4E)-1-(4-hydroxy-3-methoxyphenyl)-5-(3-hydroxy-4-methoxyphenyl)penta-1,4-dien-3-one	15.120	4.820	4.895	-0.075
M20	(1E,4E)-1-(4-hydroxy-3,5-dimethoxyphenyl)-5-(4-hydroxy-3-methoxyphenyl)penta-1,4-dien-3-one	10.210	4.991	4.846	0.145
M21	(1E,4E)-1-(3-ethoxy-4-hydroxyphenyl)-5-(4-hydroxy-3-methoxyphenyl)penta-1,4-dien-3-one	10.746	4.969	4.801	0.168
M22 ^a	(1E,4E)-1-(3,4-dimethylphenyl)-5-(2-hydroxy-4-methoxyphenyl)penta-1,4-dien-3-one	62.582	4.204	4.173	0.031
M23 ^a	(1E,4E)-1-(3,4-dimethoxyphenyl)-5-(2-hydroxy-4-methoxyphenyl)penta-1,4-dien-3-one	32.046	4.494	4.408	0.086
M24	(1E,4E)-1-(2-hydroxy-4-methoxyphenyl)-5-(3,4,5-trimethoxyphenyl)penta-1,4-dien-3-one	35.047	4.455	4.803	-0.348
M25	(1E,4E)-1-(3,4-dimethylphenyl)-5-(4-hydroxy-3,5-dimethoxyphenyl)penta-1,4-dien-3-one	11.018	4.958	5.062	-0.105
M26	(1E,4E)-1-(3,4-dimethoxyphenyl)-5-(4-hydroxy-3,5-dimethoxyphenyl)penta-1,4-dien-3-one	5.004	5.301	5.320	-0.019
M27 ^a	(1E,4E)-1-(4-hydroxy-3,5-dimethoxyphenyl)-5-(3,4,5-trimethoxyphenyl)penta-1,4-dien-3-one	11.248	4.949	5.227	-0.279
M28	(1E,6E)-1-(3-((dimethylamino)methyl)-4-hydroxyphenyl)-7-(4-hydroxy-3-methoxyphenyl)hepta-1,6-diene-3,5-one	7.356	5.133	5.362	-0.228
M29	(1E,4E)-1,5-bis(3-((dimethylamino)methyl)-4-hydroxyphenyl)penta-1,4-dien-3-one	0.647	6.189	6.260	-0.070
M30	(2E,5E)-2,5-bis(3-((dimethylamino)methyl)-4-hydroxybenzylidene)cyclopentanone	0.935	6.029	5.948	0.081
M31	(2E,6E)-2,6-bis(3-((dimethylamino)methyl)-4-hydroxybenzylidene)cyclohexanone	0.967	6.014	5.753	0.262
M32	(2E,6E)-2,6-bis(3-((dimethylamino)methyl)-4-hydroxy-5-methoxybenzylidene)cyclohexanone	2.307	5.637	5.678	-0.041
M33	(2E,6E)-2-(3-((dimethylamino)-5-((dimethylamino)methyl)-4-hydroxybenzylidene)-6-(3-((dimethylamino)-4-hydroxybenzylidene)cyclohexanone	0.927	6.033	6.111	-0.079
M34	(E)-2-benzylidene-6-cinnamoylcyclohexanone	904.90	3.043	3.158	-0.115
M35	(E)-2-(4-hydroxybenzylidene)-6-((E)-3-(4-hydroxyphenyl)acryloyl)cyclohexanone	898.87	3.046	3.384	-0.338
M36 ^a	(E)-2-(4-methoxybenzylidene)-6-((E)-3-(4-methoxyphenyl)acryloyl)cyclohexanone	1532.2	2.815	3.028	-0.213
M37	(E)-2-(4-hydroxy-3-methoxybenzylidene)-6-((E)-3-(4-hydroxy-3-methoxyphenyl)acryloyl)cyclohexanone	294.08	3.532	3.657	-0.126
M38	(E)-2-(4-chlorobenzylidene)-6-((E)-3-(4-chlorophenyl)acryloyl)cyclohexanone	273.56	3.563	3.462	0.101
M39 ^a	(E)-2-(4-methylbenzylidene)-6-((E)-3-(p-tolyl)acryloyl)cyclohexanone	468.46	3.329	3.069	0.260
M40	(E)-2-benzylidene-5-cinnamoylcyclopentanone	21.166	4.674	4.365	0.310
M41 ^a	(E)-2-(4-hydroxybenzylidene)-5-((E)-3-(4-hydroxyphenyl)acryloyl)cyclopentanone	20.062	4.698	4.465	0.233
M42 ^a	(E)-2-(4-methoxybenzylidene)-5-((E)-3-(4-methoxyphenyl)acryloyl)cyclopentanone	123.23	3.909	3.425	0.484
M43	(E)-2-(4-hydroxy-3-methoxybenzylidene)-5-((E)-3-(4-hydroxy-3-methoxyphenyl)acryloyl)cyclopentanone	27.610	4.559	4.419	0.140
M44	(E)-2-(3,4-dimethoxybenzylidene)-5-((E)-3-(3,4-dimethoxyphenyl)acryloyl)cyclopentanone	12.674	4.897	4.529	0.368
M45	(E)-2-(4-chlorobenzylidene)-5-((E)-3-(4-chlorophenyl)acryloyl)cyclopentanone	33.414	4.476	4.632	-0.156
M46	(E)-2-(4-methylbenzylidene)-5-((E)-3-(p-tolyl)acryloyl)cyclopentanone	168.52	3.773	3.765	0.008
M47	(E)-2-(4-nitrobenzylidene)-5-((E)-3-(4-nitrophenyl)acryloyl)cyclopentanone	141.25	3.850	3.871	-0.022

^a Test Set.

Table 2
Developed models for curcumin antioxidant derivatives by genetic function approximation.

S/No	Equation
1	$pIC_{50} = 1.018 * MATS3s - 2.724 * SpMax6_Bhe + 3.412 * nssnN + 1.399 * ETA_Eta_F_L + 1.198 * RotBtFrac - 1.087 * RDF65m + 4.420$
2	$pIC_{50} = 1.493 * MATS3s - 2.669 * SpMax6_Bhe + 2.902 * nssnN + 1.285 * RotBtFrac + 1.374 * SpMAD_D - 1.216 * RDF65m + 4.187$
3	$pIC_{50} = 0.893 * MATS3s + 0.575 * GATS4s - 2.812 * SpMax6_Bhe + 3.321 * nssnN + 1.373 * ETA_Eta_F_L + 1.736 * RotBtFrac - 1.126 * RDF65m + 3.950$
4	$pIC_{50} = 0.473 * ATSC7v + 1.109 * MATS3s - 2.796 * SpMax6_Bhe + 3.675 * nssnN + 1.312 * ETA_Eta_F_L + 1.111 * RotBtFrac - 1.077 * RDF65m + 4.228$
5	$pIC_{50} = 1.011 * MATS3s - 2.760 * SpMax6_Bhe + 3.424 * nssnN + 1.248 * ETA_Eta_F_L + 1.270 * RotBtFrac - 1.137 * RDF65m + 0.310 * RDF135m + 4.356$

Thus the predicted activities and residual values presented in Table 1 are generated from the results of model 4. Also the plots of predicted activities against experimental activities for the training and test sets as presented in Figs. 1 and 2 respectively are generated from the results of model 4.

Results of applicability domain

Applicability domain results for training set and test set compounds are presented in Tables S5 and S6 respectively of the supplementary data. Also the William's plot (plot of standard residuals

Table 3
Summary of internal validation results for curcumin antioxidant derivatives.

Validation parameters	Model 1	Model 2	Model 3	Model 4	Model 5
Friedman LOF	0.104	0.109	0.112	0.115	0.115
R-squared	0.925	0.921	0.932	0.931	0.931
Adjusted R-squared	0.909	0.905	0.916	0.914	0.914
Cross validated R-squared	0.892	0.884	0.891	0.887	0.886
Significant Regression	Yes	Yes	Yes	Yes	Yes
Significance-of-regression F-value	61.260	58.010	57.190	55.840	55.570
Critical SOR F-value (95%)	2.434	2.434	2.354	2.354	2.354
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	30.000	30.000	29.000	29.000	29.000
Min expt. error for non-significant LOF (95%)	0.193	0.197	0.185	0.187	0.187
Standard Error of Estimate	0.233	0.239	0.224	0.226	0.227

*The criteria for model acceptability is: $R^2 \geq 0.6$ [35].

Table 4
Results of y-randomization for curcumin antioxidant derivatives.

Parameters	Model 1	Model 2	Model 3	Model 4	Model 5
R	0.962	0.960	0.966	0.965	0.965
R²	0.925	0.921	0.932	0.931	0.931
Q²	0.892	0.884	0.891	0.887	0.886
<i>Random Model Parameters</i>					
Average r	0.398	0.392	0.438	0.412	0.445
Average r²	0.164	0.165	0.202	0.180	0.206
Average Q²	-0.305	-0.312	-0.358	-0.41	-0.325
cR_p²	0.842	0.840	0.831	0.842	0.826

*Model acceptability criteria: $R \geq 0.8$, $R^2 \geq 0.6$, $Q^2 > 0.5$, $cR_p^2 \geq 0.5$ [35].

Table 5
External validation results for curcumin antioxidant derivatives.

Validation Parameters	Model 1	Model 2	Model 3	Model 4	Model 5
r²	0.853	0.841	0.840	0.864	0.836
r₀²	0.853	0.832	0.838	0.861	0.834
Reverse r₀²	0.829	0.753	0.788	0.857	0.819
r_m²	0.851	0.760	0.802	0.817	0.800
Reverse r_m²	0.720	0.591	0.648	0.792	0.729
Average r_m²	0.786	0.675	0.725	0.805	0.765
Delta r_m²	0.131	0.169	0.154	0.025	0.071
r² - r₀²/r²	0.000	0.011	0.002	0.003	0.002
r² - r₀²/r²	0.028	0.105	0.062	0.008	0.020
k	1.035	1.034	1.038	1.045	1.034
k'	0.961	0.962	0.958	0.953	0.962
 r₀² - r_m² 	0.024	0.079	0.050	0.004	0.015
rmsep	0.352	0.369	0.371	0.362	0.367
R_{pred}²	0.853	0.838	0.836	0.844	0.839

The acceptable threshold values for the given parameters are as follows: $R_{pred}^2 > 0.5$, $r^2 > 0.6$, $r_m^2 \geq 0.5$, Delta $r_m^2 < 0.2$, $|r_0^2 - r_m^2| < 0.3$, $(r^2 - r_0^2)/r^2 < 0.1$ and $0.85 \leq k \leq 1.15$, or $(r^2 - r_0^2)/r^2 < 0.1$ and $0.85 \leq k' \leq 1.15$ [29].

against leverages) for Curcumin training and test sets are presented in Fig. 3. The computed threshold leverage (h^*) for the curcumin antioxidants is 0.649. From Fig. 3, no response outliers were observed for both training and test set compounds, since the standard residuals of all the tested compounds fell within ± 2.5 standard deviation units. Also, among the training set compounds, no structural outliers were observed as their leverage values were all below the threshold value. For the test set compounds, five structural outliers namely, compound No. 11, 22, 36, 39 and 41 were observed. These compounds are thus outside the applicability domain of the developed curcumin antioxidants model.

Interpretation and significance of the descriptors in the developed QSAR model

The results of Coefficient, Standard Error, Mean Effect, Variation Inflation Factor and Degree of Contribution of the Descriptors in the developed curcumin antioxidants QSAR model are presented in Table 6. The VIF results presented in Table 6 were within the acceptable range of 1–5, which means that the developed model is acceptable [43]. Recall that there is no inter-correlation among the descriptors if the calculated VIF result is equal to 1. If the value falls within the range 1 – 5, then the model is acceptable. Also a recheck is recommended if the computed VIF result is larger than 10 [43].

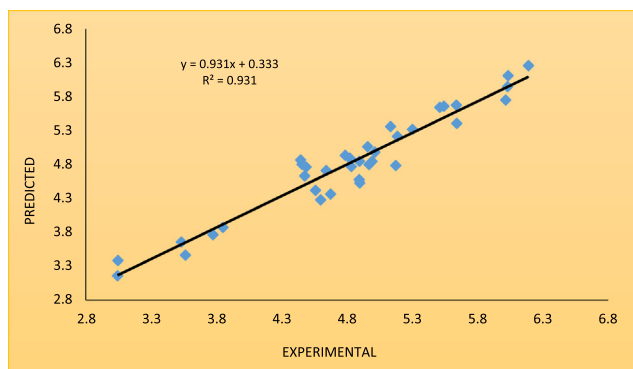


Fig. 1. Plot of experimental activities against predicted activities for training set of curcumin antioxidants.

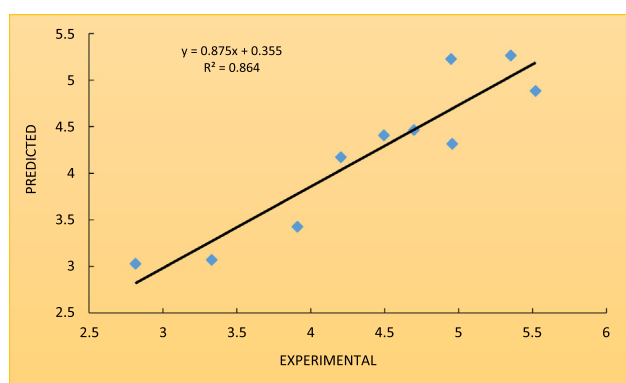


Fig. 2. Plot of experimental activities against predicted activities for test set of curcumin antioxidants.

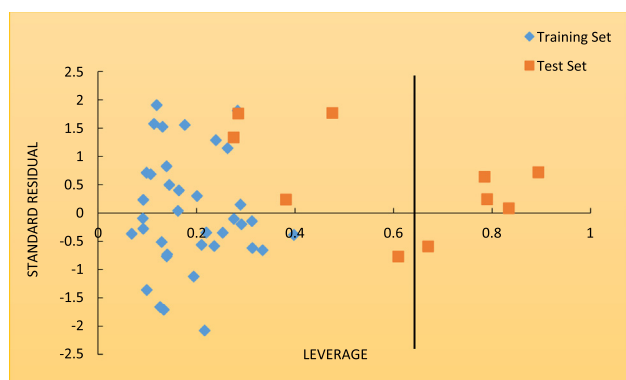


Fig. 3. William's plot for curcumin antioxidants.

ATSC7v (Centered Broto-Moreau autocorrelation - lag 7/ weighted by van der Waals volume) and **MATS3s** (Moran autocorrelation - lag 3/weighted by 1-state). These are 2D autocorrelation descriptors weighted by van der Waals volume and 1-state respectively. These two descriptors are positively correlated with the antioxidant activities of the curcumins with coefficients of 0.473 and 1.109 respectively.

SpMax6_Bhe Largest absolute eigenvalue of Burden modified matrix - n 6/weighted by relative Sanderson electronegativities. From the results presented in Table 6, this 2D descriptor has the lowest contribution towards influencing the antioxidant activities of the curcumin derivatives based on its value for DC, MF and coefficient of -9.086 , -0.734 and -2.796 respectively.

nsssN (Count of atom-type E-state: >N-). This descriptor dictates the number of nitrogen atoms attached to the curcumin antioxidant moiety. As presented in Table 6, the DC, MF and coefficient results for this descriptor are 12.976, 0.965 and 3.675 respectively. These results are by far higher than those recorded by the other descriptors. This is an indication of the strong contribution and relative significance of this descriptor in influencing the antioxidant activities of the curcumins. In addition, this descriptor has a very strong positive correlation with the antioxidant activities of the curcumin derivatives. Thus by increasing the number of nitrogen atoms attached to the curcumin moiety at the E-state, the antioxidant activities of the curcumins increases.

ETA_Eta_F_L (Local functionality contribution EtaF local). This descriptor is also positively correlated with antioxidant activities of the curcumins.

RotBtFrac (Fraction of rotatable bonds, including terminal bonds). This is the fraction of bonds which allow free rotation around themselves. They can also be regarded as the fraction of single bonds, not in a ring, bound to a nonterminal heavy atom. This descriptor is positively correlated with the activities of the curcumin antioxidants with DC, MF and coefficient values of 5.710, 0.292 and 1.111 respectively. The high DC value implies that this descriptor also has a strong influence on the antioxidant activities of the curcumins. Thus increasing the number of rotatable bonds, including terminal bonds in curcumin antioxidants appreciably improves their antioxidant activities.

RDF65m (Radial distribution function - 065/weighted by relative mass). This is a 3D descriptor in which the associated weighing scheme is the relative mass. The negative DC and MF values of -4.903 and -0.283 are in very good agreement with the negative coefficient result of -1.077 for this descriptors. Thus this descriptor is strongly negatively correlated with the antioxidant activities of the curcumins.

Conclusions

The free radical scavenging activities of the curcumin derivatives were investigated by QSAR studies which culminated in the design of five predictive models with highly impressive results upon internal and external validations. The degree of contribution,

Table 6

Specifications of coefficient, standard error, mean effect, variation inflation factor and degree of contribution of the descriptors for curcumin antioxidants.

Descriptor	Coefficient	Standard Error	P-Value	DC	MF	VIF
ATSC7v	0.473	0.289	0.11205	1.639	0.124	2.295
MATS3s	1.109	0.184	1.45E-06	6.033	0.291	1.299
SpMax6_Bhe	-2.796	0.308	5.54E-10	-9.086	-0.734	3.775
nsssN	3.675	0.283	1.32E-13	12.98	0.965	3.844
ETA_Eta_F_L	1.312	0.288	8.54E-05	4.563	0.345	3.611
RotBtFrac	1.111	0.195	3.54E-06	5.710	0.292	2.099
RDF65m	-1.077	0.220	3.32E-05	-4.903	-0.283	1.929

variation inflation factor and mean effect of each descriptor in the developed model were all calculated. Also, the leverage approach was employed in accessing the applicability domain of the model. These results indicate that the main descriptors that influence the free radical scavenging activities of the curcumin antioxidants are the **nsssN** (Count of atom-type E-State: >N-); **MATS3s** (Moran autocorrelation - lag 3/weighted by I-state) and **RotBtFrac** (Fraction of rotatable bonds, including terminal bonds) descriptors. Thus, these descriptors must be considered in the design of potent antioxidants with improved activities based on the curcumin moiety.

Conflict of interest

The authors have declared no conflict of interest.

Compliance with Ethics Requirements

This article does not contain any studies with human or animal subjects.

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.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.jare.2018.03.003>.

References

- [1] Wichitnithad W, Jongaroonngamsang N, Pummungura S, Rojsitthisak P. A simple isocratic HPLC method for the simultaneous determination of curcuminoids in commercial turmeric extracts. *Phytochem Anal* 2009;20:314–9.
- [2] Bayomi SM, El-Kashef HA, El-Ashmawy MB, Nasr NA, El-Sherbeny MA, Badria FA, et al. Synthesis and biological evaluation of new curcumin derivatives as antioxidant and antitumor agents. *Med Chem Res* 2013;22:1147–62.
- [3] Yodkeereea S, Chaiwangyena W, Garbisab S, Limtrakul P. Curcumin, demethoxycurcumin and bisdemethoxycurcumin differentially inhibit cancer cell invasion through the down-regulation of MMPs and uPA. *J Nutr Biochem* 2009;20:87–95.
- [4] Hamed OA, Mehdawi N, Taha AA, Hamed EM, Al-Nuri MA, Hussein AS. Synthesis and antibacterial activity of novel curcumin derivatives containing heterocyclic moiety. *Iran J Pharm Res* 2013;12(1):47–56.
- [5] Kumar D, Mishra PK, Anand AK, Agrawal PK, Mohapatra R. Isolation, synthesis and pharmacological evaluation of some novel curcumin derivatives as anticancer agents. *J Med Plants Res* 2012;6(14):2880–4.
- [6] Li Q, Chen J, Luo S, Xu J, Huang Q, Liu T. Synthesis and assessment of the antioxidant and antitumor properties of asymmetric curcumin analogues. *Eur J Med Chem* 2015;93:461–9.
- [7] Neto Z, Machado M, Lindeza A, do Rosário V, Gazarini ML, Lopes D. Treatment of *Plasmodium chabaudi* parasites with curcumin in combination with antimalarial drugs: drug interactions and implications on the ubiquitin/proteasome system. *J Parasitol Res* 2013;429736:1–11.
- [8] Shang YJ, Jin XL, Shang XL, Tang JJ, Liu GY, Dai F, et al. Antioxidant capacity of curcumin-directed analogues: structure–activity relationship and influence of microenvironment. *Food Chem* 2010;119:1435–42.
- [9] Fang X, Fang L, Gou S, Cheng L. Design and synthesis of dimethylaminomethyl-substituted curcumin derivatives/analogues: potent antitumor and antioxidant activity, improved stability and aqueous solubility compared with curcumin. *Bioorg Med Chem Lett* 2013;23:1297–301.
- [10] Bhullar KS, Jha A, Youssef D, Rupasinghe HV. Curcumin and its carbocyclic analogs: structure-activity in relation to antioxidant and selected biological properties. *Molecules* 2013;18:5389–404.
- [11] Li Q, Chen J, Luo S, Xu J, Huang Q, Tianyu L. Synthesis and assessment of the antioxidant and antitumor properties of asymmetric curcumin analogues. *Eur J Med Chem* 2015;93:461–9.
- [12] Brewer MS. Natural antioxidants: sources, compounds, mechanisms of action, and potential applications. *Compr Rev Food Sci Food Saf* 2011;10:221–47.
- [13] Birben E, Sahiner UM, Sackesen C, Erzurum S, Kalayci O. Oxidative stress and antioxidant defense. *World Allergy Organ J* 2012;5(1):9–19.
- [14] Taha M, Ismail NH, Jamil W, Rashwan H, Kashif SM, Sain AA, et al. Synthesis of novel derivatives of 4-methylbenzimidazole and evaluation of their biological activities. *Eur J Med Chem* 2014;84:731–8.
- [15] Luo X, Wang C, Liu Y, Huang Z. New multifunctional melatonin-derived benzylpyridinium bromides with potent cholinergic, antioxidant, and neuroprotective properties as innovative drugs for Alzheimer's disease. *Eur J Med Chem* 2015;103:302–11.
- [16] Kurt BZ, Gazioglu I, Sonmez F, Kucukislamoglu M. Synthesis, antioxidant and anticholinesterase activities of novel coumarylthiazole derivatives. *Bioorg Chem* 2015;59:80–90.
- [17] Shekhar TC, Anju G. Antioxidant activity by DPPH radical scavenging method of *Ageratum conyzoides* Linn. Leaves. *Am J Ethnomed* 2014;1(4):244–9.
- [18] Ogadimma AI, Adamu U. Quantitative structure activity relationship analysis of selected chalcone derivatives as *Mycobacterium tuberculosis* inhibitors. *Open Access Lib J* 2016;3:e2432. doi: <https://doi.org/10.4236/oalib.1102432>.
- [19] Mitra I, Saha A, Roy K. Quantitative structure–activity relationship modeling of antioxidant activities of hydroxybenzalacetones using quantum chemical, physicochemical and spatial descriptors. *Chem Biol Drug Des* 2009;73:526–36.
- [20] Yehye WA, Rahman NA, Saad O, Ariffin A, Hamid SBA, Alhadi AA, et al. Rational design and synthesis of new, high efficiency, multipotent schiff base-1,2,4-triazole antioxidants bearing butylated hydroxytoluene moieties. *Molecules* 2016;21:847. doi: <https://doi.org/10.3390/molecules21070847>.
- [21] Li Z, Wan H, Shi Y, Ouyang P. Personal experience with four kinds of chemical structure drawing software: review on ChemDraw, ChemWindow, ISIS/Draw, and ChemSketch. *J Chem Inform Comput Sci* 2004;44(5):1886–90.
- [22] Hehre WJ, Huang WW. Chemistry with computation: an introduction to SPARTAN. Irvine: Wavefunction, Inc.; 1995. ISBN: 9780964349520.
- [23] Hohenberg P, Kohn W. Inhomogeneous electron gas. *Phys Rev* 1964;136(3B):864–71.
- [24] Lee C, Yang W, Parr RG. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B Condens Matter* 1988;37(2):785–9.
- [25] Becke AD. Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys* 1993;98(7):5648–52.
- [26] Mikulski D, Eder K, Molski M. Quantum-chemical study on relationship between structure and antioxidant properties of hepatoprotective compounds occurring in cynara scolymus and silybum marianum. *J Theor Comput Chem* 2014;13(1):1–24.
- [27] Yap CW. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem* 2011;32(7):1466–74.
- [28] Ambure P, Aher RB, Gajewicz A, Puzyn T. "NanoBRIDGES" software: open access tools to perform QSAR and nano-QSAR modeling. *Chemom Intell Lab Syst* 2015;147:1–13.
- [29] Golbraikh A, Tropsha A. Beware of q²! *J Mol Graph Model* 2002;20(4):269–76.
- [30] Todd MM, Harten P, Douglas MY, Muratov EN, Golbraikh A, Zhu H, et al. Does rational selection of training and test sets improve the outcome of QSAR modeling? *J Chem Inform Model* 2012;52(10):2570–8.
- [31] Mandal AS, Roy K. Predictive QSAR modeling of HIV reverse transcriptase inhibitor TIBO derivatives. *Eur J Med Chem* 2009;44:1509–24.
- [32] Wold S. Cross-validation estimation of the number of components in factor and principal components models. *Technometrics* 1978;20:397–405.
- [33] Rudra ND, Kunal R. Development of classification and regression models for *Vibrio fischeri* toxicity of ionic liquids: green solvents for the future. *Toxicol Res* 2012;1:186–95.
- [34] Tropsha A, Gramatica P, Gombar VK. 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] Mitra I, Saha A, Roy K. Chemometric modeling of free radical scavenging activity of flavone derivatives. *Eur J Med Chem* 2010;45:5071–9.
- [36] Todeschini R. Milano Chemometrics. Italy (personal communication); 2010.
- [37] Pravin A. Drug Theoretics and Cheminformatics (DTC) laboratory. Jadavpur University; 2013.
- [38] Tropsha A. Best practices for QSAR model development, validation, and exploitation. *Mol Inform* 2010;29(6–7):476–88.
- [39] Sharma BK, Singh P. Chemometric descriptor based QSAR rationales for the MMP-13 inhibition activity of non-zinc-chelating compounds. *Med Chem* 2013;3:168–78.
- [40] Saaidpour S. Quantitative modeling for prediction of critical temperature of refrigerant compounds. *Phys Chem Res* 2016;4(1):61–71.
- [41] Leach AR. Molecular modelling: principles and applications. Harlow, England: Pearson Education Ltd.; 2001.
- [42] Ravichandran V, Harish R, Abhishek J, Shalini S, Christopher PV, Ram KA. Validation of QSAR models –strategies and importance. *Int J Drug Des Discovery* 2011;2(3):511–9.
- [43] Baumann K. Chance correlation in variable subset regression: influence of the objective function, the selection mechanism, and ensemble averaging. *QSAR Comb Sci* 2005;24:1033–46.