RECORDSOF CHEMICAL SCIENCES

Design and Validation of Predictive Antioxidant Models for Chalcones

Ikechukwu Ogadimma Alisi [1], Adamu Uzairu [2], and Stephen Eyije Abechi [2]*

Accepted: 17th Sept, 2019
Published: 10th Nov, 2019

1. Department of Applied Chemistry, Federal University Dutsinma, Katsina State, Nigeria.

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

**Corresponding Author:*
Ikechukwu Ogadimma Alisi
ikeogadialisi@gmail.com, ialisi@fudutsinma.edu.ng , +2347064888876

FRsCS Vol.1 No. 2
Official Journal of Dept. of Applied Chemistry, Federal University of Dutsin-Ma, Katsina State.
http://journal.fudutsinma.edu.ng/index.php/rcs/index
ISSN: 2705-2354 Print
2705-2362 Electronic

Abstract

The Chalcones are open-chain flavonoids with diverse biological activities. In recent time, quantitative structure activity relationship (QSAR) is employed in the elucidation of the structural requirements of various biological activities. In this research, QSAR was employed to investigate the structural requirements of free radical scavenging properties of selected chalcones and their derivatives. The data set was optimized at the density functional theory (DFT) level. The optimized structures were employed to develop chalcone antioxidant models by genetic function algorithm (GFA). The range of applicability of these models were assessed by leverage approach. Five predictive models were developed for the chalcone antioxidants with highly encouraging results upon validation [R= 0.988, R^2= 0.977, Q^2 (R^2_{CV})= 0.954, R^2_{pred} = 0.916, $^cR_p^2$= 0.861]. The electrotopological state atom type and the extended topochemical atom descriptors were found to be significant in the determination of the free radical scavenging properties of the chalcone antioxidants. This research is a gateway towards the design of new set of chalcone antioxidants with potent free radical scavenging properties.

Keywords: Antioxidants; Chalcones; Descriptors; Model validation; QSAR

Introduction

Compounds that have the ability to delay autoxidation by preventing the formation of free radicals or disrupting their propagation are regarded as antioxidants (Brewer, 2011). On the other hand, molecules that contain unpaired electrons in their atomic or molecular orbitals are recognised as free radicals (Zhang et al., 2014). The metabolic processes that take place in the human system result in the generation of free radicals, which are also scavenged by an efficient network of antioxidants. When the balance between the free radicals and the antioxidants is disrupted, oxidative damage to cell structures results. This is responsible for many degenerative diseases such as cancer, cardiovascular disease, cataracts, immune system decline, and brain dysfunction (Sinha et al., 2009). Antioxidants and their derivatives such as the Chalcones, scavenge free radicals such as peroxide, hydroperoxide or lipid peroxyl, and consequently inhibit the oxidative mechanisms that lead to degenerative disease. According to Todorova et al in 2011, Chalcones ((1,3-diphenylprop-2-en-1-ones) are open-chain flavonoids that consist of two aromatic rings joined by a three-carbon α, β-unsaturated carbonyl linkage. The synthesis of many biologically important heterocycles such as flavones employ the chalcones as starting materials (Patel et al., 2013). In this research, various chalcones and their derivatives whose antioxidant activities were tested using the stable 1,1-diphenyl-2-

picrylhydrazyl, inhibition assay, were subjected to quantitative structure activity relationship (QSAR) studies. The QSAR method is a statistical technique that is widely employed in correlating the properties of molecular structures with their biological activities (Mitra et al., 2010; Ravichandran et al., 2011). In recent time, various attempts have been made by researchers towards the elucidation of the structural requirements of the antioxidant activities of various compounds by employing QSAR technique (Wang et al., 2015; Alisi et al., 2018; Ogadimma et al., 2018). In 2014, Pachori et al., employed QSAR technique to computationally design and identify novel methoxy substituted chalcones. QSAR has been employed in the chemometric modelling of free radical scavenging activity of flavone derivatives (Mitra et al., 2010). Also, Xue et al., in 2012, employed QSAR to investigate the molecular structure and radical scavenging activity of newly synthesized hydroxychalcones. Other QSAR studies involving chalcone derivatives include the application of QSAR in the analysis of selected chalcone derivatives as *Mycobacterium tuberculosis* inhibitors (Ogadimma and Adamu, 2016).

Materials and methods

2.1. Data set, descriptors development and pre-treatment

The data set of sixty substituted Chalcone derivatives and their respective antioxidant activities were generated from literature (Beom-Tae *et al.*, 2008; Shenvi *et al.*, 2013; Lahsasni *et al.*, 2014). The free radical scavenging activity of these molecules were investigated using the 1,1-diphenyl-2-picrylhydrazil (DPPH) free radical inhibition assay. The DPPH free radical inhibition assay is preferred to other methods such as nitric oxide radical ($NO^{\bullet}$), superoxide anion radical ($O_2^{\bullet-}$), Hydroxyl radical ($OH^{\bullet}$), hydrogen peroxide (H_2O_2) or 2,2'-azinobis (3-ethylbenzothiazoline-6-sulfonate) cationic radical ($ABTS^{\bullet+}$) scavenging assays, due to its simplicity and rapidity. Also, it is unaffected by sample polarity (Koleva et al., 2001).

The antioxidant activities of these molecules are represented by their IC_{50} values in $\mu g/ml$. The IC_{50} values were subsequently transformation to their corresponding pIC_{50} values according to equation (1).

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

ChemBioDraw version 12.0 program (Li et al., 2004), was employed in drawing the chemical structures of the compounds. The molecular geometries were first minimized and subsequently optimized at the DFT level using Becke's three-parameter Lee-Yang-Parr hybrid functional (B3LYP), and the 6-311G* basis set. This was accomplished with the aid of the Spartan 14 program (Shao *et al.*, 2006). The low energy conformers were subequently submitted for the generation of molecular descriptors using the software "PaDel-Descriptor version 2.20" (Yap, 2011).

Data pre-treatment was accomplished using "*Data Pre-Treatment GUI 1.2*" tool that uses V-WSP algorithm (Ballabio et al., 2014; Ambure et al., 2015). The resulting data was normalized and subsequently split into training and test sets by employing Kennard Stone algorithm technique using the program "Dataset Division GUI 1.2" (Todd et al., 2012).

2.2. Model development and validation

Material studio program was employed in the development of QSAR models from the training set compounds. The descriptors (independent variables) and pIC_{50} values (dependent variables) were subjected to multivariate analysis by Genetic Function Approximation (GFA). The model fitness was determined by calculating the Friedman lack-of-fit (LOF) (Ogadimma *et al.*, 2018).

The validation of the developed models internally was achieved by leave- one- out (LOO) cross- validation technique. This resulted in the computation of various parameters such as the correlation coefficient, R; the Cross-validated squared correlation coefficient, $R_{cv}^2(Q^2)$; the adjusted R^2 (R_α^2); the variance ratio, F and the standard error of estimate (s) (Rudra and Kunal, 2012).

The robustness of the developed QSAR model was checked by Y-randomization test using the relation presented in equation 2.

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

Where, R^2 = squared correlation coefficient of the non-random model.

R_r^2 = squared mean correlation coefficient of the randomized model.

$^{c}R_p^2$ = correction in squared mean correlation coefficient deviation.

Various parameters were calculated in order to validate the models externally. These include:

The predictive R^2 (R^2pred) of the test set compounds which indicates the extent of predictivity of the model (equation 3).

$$R^2_{pred} = 1 - \frac{\Sigma\left(Y_{pred(Test)} - Y_{(Test)}\right)^2}{\Sigma\left(Y_{(Test)} - \bar{Y}_{(Training)}\right)^2}$$

(3)

Where,

$Y_{pred(Test)}$ = Predicted activity values of the test set compounds.

$Y_{(Test)}$ = Observed activity values of the test set compounds.

$\bar{Y}_{(Training)}$ = Mean activity value of the training set.

Other external validation parameters calculated include: the Root Mean Square Error in Prediction (RMSEP), the modified R^2 (r^2_m), $r^2 - r_0^2/r^2$, $r^2 - r_0'^2/r^2$, k *and* k'.

Where,

r_0^2 = Squared correlation coefficient of plots between observed and predicted activities with intercept set to zero and slope equal to k. Interchanged of the axes gives slope equal to k'.

r^2 = Squared correlation coefficient of plots between observed and predicted activities with intercept not set to zero (Roy and Roy, 2008; Rudra and Kunal, 2012).

2.3. Investigation of the mean effect and degree of contribution

The mean effect (MF) of each descriptor in the model was estimated using equation (4). The MF value for a given descriptor in a model indicates the relative significance and contribution of that descriptor in comparison to the other descriptors (Alisi *et al.*, 2019).

$$MF_j = \frac{\beta_j \Sigma_{i=1}^{i=n} d_{ij}}{\Sigma_j^m \beta_j \Sigma_1^n d_{ij}}$$

(4)

Where,

MF_j = Descriptor j mean effect value.

β_j = Coefficient of descriptor j.

d_{ij} = Descriptors value for each molecule.

m = Number of descriptors in the model.

The Degree of Contribution (DC) (standard regression coefficient) for each descriptor was also calculated.

2.4. Applicability domain investigation

The applicability domain of a given antioxidant model is the chemical space where the model can reliably make predictions (Netzeva et al., 2005). The leverage approach was employed in assessing the applicability domain of the model (Gramatica et al., 2007). This was accomplished by computing the leverage matrix (H) for the dataset X (equation 5).

$$H = X(X^TX)^{-1}X^T$$

(5)

Where,

X = Training set two-dimensional $n \times k$ descriptor matrix.

n = number of training set molecules.

k = number of descriptors in the model.

X^T = Transpose of X.

The leverage of the ith molecule, h_i, was also calculated (equation 6).

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

(6)

The cut-off leverage, h*, was estimated using equation (7) (Gramatica, 2010).

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

(7)

The ratio of the residual to the Root Mean Square Error ($RMSE$) was employed in estimating the standard residual of each molecule.

3. Results and discussion

3.1. Results of data set generation, descriptors development and pre-treatment

The data set and their corresponding activities (observed, predicted and residual) are presented in Table 1. Upon geometry optimization and subsequent minimization a total of 1907 descriptors were produced. These descriptors fall in the categories of electronic, spatial, structural, thermodynamic and topological.

After data processing the entire descriptors were reduced to 1028 descriptors. This procedure eliminated highly correlated descriptors which result in poor models. The normalized data was also obtained, and this ensures that no particular descriptor dominates the model due to larger or smaller pre-scaled (Brignole et al., 2013). Also, data division resulted in 48 molecular compounds in the training set and 12 compounds in the test set.

3.2. Model development and validation

A total of five models were developed for the chalcone antioxidant derivatives from the training set by GFA as presented in Table 2. The summary of internal validation results for these models is presented in Table 3. All the five developed models have results that exceed the threshold value of 0.5 for R^2, R_α^2 and R^2_{cv}. R_α^2 is a modification of R^2 in which its result is unaffected by increase in the number of descriptor terms in the model unlike R^2, except if such an increase improves the robustness of the resulting model. (Rudra and Kunal, 2012).

Table 1 Chalcone antioxidants data set and their activities.

No	Compounds	IC_{50}	pIC_{50}		
			Observed	Predicted	Residual
	(E)-1-(3-nitrophenyl)-3-(2,4,5-trimethoxyphenyl)prop-2-en-1-one	4.309	5.366	5.282	0.084
	(E)-1-(4-chlorophenyl)-3-(2,4,5-trimethoxyphenyl)prop-2-en-1-one	4.593	5.338	5.332	0.006
	(E)-1-(4-methoxyphenyl)-3-(2,4,5-trimethoxyphenyl)prop-2-en-1-one	4.394	5.357	5.348	0.009
	(E)-1-(p-tolyl)-3-(2,4,5-trimethoxyphenyl)prop-2-en-1-one	4.325	5.364	5.305	0.059
	(E)-1-phenyl-3-(2,4,5-trimethoxyphenyl)prop-2-en-1-one	4.506	5.346	5.379	-0.033
	(E)-1-(2,4-dihydroxyphenyl)-3-(2,4,5-trimethoxyphenyl)prop-2-en-1-one	4.368	5.36	5.435	-0.076
	(E)-1-(4-fluorophenyl)-3-(2,4,5-trimethoxyphenyl)prop-2-en-1-one	4.710	5.327	5.309	0.018
	(E)-1-(4-morpholinophenyl)-3-(2,4,5-trimethoxyphenyl)prop-2-en-1-one	4.269	5.37	5.308	0.062
	(E)-1-(4-(1H-imidazol-2-yl)phenyl)-3-(2,4,5-trimethoxyphenyl)prop-2-en-1-one	4.757	5.323	5.408	-0.086
	(E)-1-(4-chloro-3-nitrophenyl)-3-(2,4,5-trimethoxyphenyl)prop-2-en-1-one	4.472	5.349	5.276	0.074
	(E)-1-(4-nitrophenyl)-3-(2,4,5-trimethoxyphenyl)prop-2-en-1-one	4.754	5.323	5.186	0.137
	(E)-1-(3-(trifluoromethyl)phenyl)-3-(2,4,5-trimethoxyphenyl)prop-2-en-1-one	5.153	5.288	5.329	-0.041
	(E)-1-(5-chloro-2-hydroxyphenyl)-3-(2,4,5-trimethoxyphenyl)prop-2-en-1-one	5.073	5.295	5.36	-0.066
	(E)-1-(4-hdroxy-3-methoxyphenyl)-3-(2,4,5-trimethoxyphenyl)prop-2-en-1-one	4.310	5.366	5.305	0.06
	(E)-1-(2,5-dihydroxyphenyl)-3-(2,4,5-trimethoxyphenyl)prop-2-en-1-one	2.653	5.576	5.404	0.172
	(E)-3-(2,4,5-trimethoxyphenyl)-1-(2,4,6-trimethoxyphenyl)prop-2-en-1-one	5.023	5.299	5.446	-0.147
	(E)-1-(4-bromophenyl)-3-(2,4,5-trimethoxyphenyl)prop-2-en-1-one	4.318	5.365	5.397	-0.033
	(E)-1-(3,5-dibenzylphenyl)-3-(2,4,5-trimethoxyphenyl)prop-2-en -1-one	5.002	5.301	5.158	0.143
	(E)-1-(4-hydroxyphenyl)-3-(2,4,5-trimethoxyphenyl)prop-2-en-1-one	4.293	5.367	5.313	0.055
	(E)-2-(3-(2,4,5-trimethoxyphenyl)acryloyl)-1,4-phenylene diacetate	4.650	5.333	5.427	-0.095
	7-hydroxy-2-(2,4,5-trimethoxyphenyl)-4H-chromen-4-one	4.808	5.318	5.364	-0.046
	6-hydroxy-2-(2,4,5-trimethoxyphenyl)-4H-chromen-4-one	4.487	5.348	5.382	-0.034
	6-chloro-2-(2,4,5-trimethoxyphenyl)-4H-chromen-4-one	4.973	5.303	5.33	-0.027
	6-chloro-3-hydroxy-2-(2,4,5-trimethoxyphenyl)-4H-chromen-4-one	4.680	5.33	5.37	-0.04
	3,7-dihydroxy-2-(2,4,5-trimethoxyphenyl)-4H-chromen-4-one	4.496	5.347	5.335	0.012
	6-hydroxy-2-(2,4,5-trimethoxyphenyl)chromen-4-one	4.566	5.34	5.379	-0.038
	(E)-3-(2,3-dihydroxyphenyl)-1-phenylprop-2-en-1-one	23.000	4.638	4.685	-0.047
	(E)-3-(2,3-dihydroxyphenyl)-1-(p-tolyl)prop-2-en-1-one	14.000	4.854	4.644	0.21
	(E)-3-(2,3-dihydroxyphenyl)-1-(4-methoxyphenyl)prop-2-en-1-one	30.000	4.523	4.677	-0.154
	(E)-3-(2,3-dihydroxyphenyl)-1-(4-hydroxyphenyl)prop-2-en-1-one	26.000	4.585	4.606	-0.021
	(E)-1-(4-chlorophenyl)-3-(2,3-dihydroxyphenyl)prop-2-en-1-one	27.000	4.569	4.544	0.024
	(E)-3-(2,5-dihydroxyphenyl)-1-phenylprop-2-en-1-one	19.000	4.721	4.959	-0.238
	(E)-3-(2,5-dihydroxyphenyl)-1-(p-tolyl)prop-2-en-1-one	16.000	4.796	4.891	-0.095
	(E)-3-(2,5-dihydroxyphenyl)-1-(4-methoxyphenyl)prop-2-en-1-one	20.000	4.699	4.909	-0.21
	(E)-3-(2,5-dihydroxyphenyl)-1-(4-hydroxyphenyl)prop-2-en-1-one	19.000	4.721	4.86	-0.139
	(E)-1-(4-chlorophenyl)-3-(2,5-dihydroxyphenyl)prop-2-en-1-one	11.000	4.959	4.84	0.119

(E)-3-(3,4-dihydroxyphenyl)-1-phenylprop-2-en-1-one	38.000	4.42	4.505	-0.085
(E)-3-(3,4-dihydroxyphenyl)-1-(p-tolyl)prop-2-en-1-one	40.000	4.398	4.475	-0.077
(E)-3-(3,4-dihydroxyphenyl)-1-(4-methoxyphenyl)prop-2-en-1-one	39.000	4.409	4.462	-0.053
(E)-3-(3,4-dihydroxyphenyl)-1-(4-hydroxyphenyl)prop-2-en-1-one	34.000	4.469	4.387	0.082
(E)-1-(4-chlorophenyl)-3-(3,4-dihydroxyphenyl)prop-2-en-1-one	68.000	4.167	4.282	-0.115
(E)-1-(3-hydroxyphenyl)-3-(3-nitrophenyl)prop-2-en-1-one	6.936	5.159	5.149	0.01
(E)-3-(3,5-dimethoxyphenyl)-1-(3-hydroxyphenyl)prop-2-en-1-one	6.349	5.197	5.073	0.124
(E)-3-(4-bromophenyl)-1-(3-hydroxyphenyl)prop-2-en-1-one	0.378	6.423	6.508	-0.085
(E)-1-(4-bromophenyl)-3-(3-hydroxyphenyl)prop-2-en-1-one	0.040	7.398	7.313	0.085
(E)-3-(3-hydroxyphenyl)-1-(4-nitrophenyl)prop-2-en-1-one	20.560	4.687	4.862	-0.175
(E)-3-(3-(3,5-dimethoxyphenyl)acryloyl)phenyl palmitate	12.635	4.898	5.07	-0.171
(E)-3-(3-(4-bromophenyl)acryloyl)phenyl palmitate	10.830	4.965	5.103	-0.137
(E)-3-(3-(4-nitrophenyl)-3-oxoprop-1-en-1-yl)phenyl palmitate	6.165	5.21	5.089	0.121
(E)-3-(3-(3,5-dimethoxyphenyl)acryloyl)phenyl stearate	6.325	5.199	5.292	-0.093
(E)-3-(3-(4-bromophenyl)acryloyl)phenyl stearate	7.916	5.101	5.066	0.036
(E)-3-(3-(4-bromophenyl)-3-oxoprop-1-en-1-yl)phenyl stearate	4.993	5.302	5.296	0.006
(E)-3-(3-(4-nitrophenyl)-3-oxoprop-1-en-1-yl)phenyl stearate	6.513	5.186	5.093	0.093
2,3-dibromo-3-(3,5-dimethoxyphenyl)-1-(3-hydroxyphenyl)propan-1-one	225.446	3.647	3.721	-0.074
2,3-dibromo-3-(4-bromophenyl)-1-(3-hydroxyphenyl)propan-1-one	124.118	3.906	3.881	0.025
2,3-dibromo-1-(4-bromophenyl)-3-(3-hydroxyphenyl)propan-1-one	129.148	3.889	3.887	0.002
3-(3,5-dimethoxyphenyl)-1-(3-hydroxyphenyl)-2,3-di(piperidin-1-yl)propan-1-one	5.119	5.291	5.004	0.287
3-(3,5-dimethoxyphenyl)-1-(3-hydroxyphenyl)-2,3-dimorpholino propan-1-one	9.883	5.005	4.985	0.02
3-(4-bromophenyl)-1-(3-hydroxyphenyl)-2,3-di(piperidin-1-yl)propan-1-one	19.349	4.713	4.561	0.152
1-(4-bromophenyl)-3-(3-hydroxyphenyl)-2,3-dimorpholinopropan-1-one	14.066	4.852	4.928	-0.076

*Test Set

Table 2 Developed models for chalcone antioxidants

Model No	Equation
1	pIC_{50} = - 3.558 * AATSC8m - 2.296 * AATSC5v + 0.511 * MATS8m - 1.804 * GATS5v + 0.450 * SpMax6_Bhs - 0.247 * SHBint3 - 2.361 * SsBr - 11.183 * maxsssCH + 12.877 * ETA_dAlpha_A + 0.654 * RDF35u + 7.426
2	pIC_{50} = - 2.955 * AATSC8m - 0.464 * AATSC5v - 0.463 * MATS2c - 0.611 * GATS4v - 0.409 * SHBint3 - 1.726 * SsBr - 11.375 * maxsssCH + 11.622 * ETA_dAlpha_A + 0.590 * RDF55s + 6.654
3	pIC_{50} = - 3.471 * AATSC8m - 2.395 * AATSC5v + 0.472 * MATS8m - 1.972 * GATS5v + 0.361 * SpMax7_Bhs - 0.251 * SHBint3 - 2.349 * SsBr - 10.839 * maxsssCH + 12.504 * ETA_dAlpha_A + 0.704 * RDF35u + 7.574
4	pIC_{50} = - 2.978 * AATSC8m - 2.341 * AATSC5v + 0.398 * MATS1e - 1.748 * GATS5v - 2.224 * SsBr + 0.398 * minssCH2 - 11.002 * maxsssCH + 12.439 * ETA_dAlpha_A + 0.699 * RDF55s + 7.446
5	pIC_{50} = - 3.587 * AATSC8m - 1.994 * AATSC5v + 0.934 * MATS8m - 1.663 * GATS5v - 0.363 * GATS6e - 2.040 * SsBr - 0.296 * minHBint3 - 10.097 * maxsssCH + 11.296 * ETA_dAlpha_A + 0.617 * RDF30v + 7.484

From table 3, model 1 has the highest values for R^2, $R_\alpha{}^2$ and $R^2{}_{cv}$ of 0.977, 0.971 and 0.956

Table 4 Results of y-randomization for chalcone antioxidants

Parameters	Model 1	Model 2	Model 3	Model 4	Model 5
R	0.989	0.985	0.988	0.984	0.988
R^2	0.977	0.97	0.977	0.969	0.977
Q^2	0.956	0.939	0.954	0.936	0.956
Random Model Parameters					
Average r	0.428	0.440	0.467	0.433	0.408
Average r^2	0.204	0.208	0.236	0.200	0.173
Average Q^2	-1.134	-0.501	-0.686	-0.470	-0.591
cR_p^2	0.881	0.868	0.861	0.871	0.889

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

Table 5 Results of external validation for chalcone antioxidants

Validation Parameters	Model 1	Model 2	Model 3	Model 4	Model 5
r^2	0.760	0.729	0.789	0.688	0.710
r_0^2	0.753	0.726	0.785	0.688	0.701
Reverse r_0^2	0.602	0.666	0.681	0.555	0.467
r^2_m	0.696	0.691	0.736	0.684	0.643
Reverse r^2_m	0.458	0.546	0.530	0.437	0.360
Average r^2_m	0.577	0.619	0.633	0.560	0.502
Delta r^2_m	0.238	0.145	0.206	0.247	0.283
$r^2 - r_0^2/r^2$	0.009	0.004	0.006	6E-05	0.012
$r^2 - r_0'^2/r^2$	0.208	0.086	0.137	0.194	0.341
k	0.986	0.963	0.991	0.977	0.99
k'	1.013	1.037	1.008	1.022	1.008
$\lvert r_0^2 - r_0'^2 \rvert$	0.151	0.060	0.104	0.133	0.234
rmsep	0.179	0.250	0.160	0.216	0.187
R^2_{pred}	0.894	0.792	0.916	0.846	0.883

*The acceptable threshold values for the given parameters are as follows:

$R^2_{pred} > 0.5$, $r^2 > 0.6$, $r^2_m \geq 0.5$, Delta $\lvert r_0^2 - r_0'^2 \rvert < 0.3$, $(r^2 - r_0^2)/r^2 < 0.1$ *and* $0.85 \leq k \leq 1.15$, *or* $(r^2 - r_0'^2)/r^2 < 0.1$ *and* $0.85 \leq k' \leq 1.15$ (Golbraikh, and Tropsha, 2002).

Table 6 Specification results of the observed descriptors in model 3

Descriptor	Description	Coefficient	Standard Error	P-Value	DC	MF
AATSC8m	Average centered Broto-Moreau autocorrelation – lag 8 / weighted by mass	-3.471	0.223	8.11E-18	-15.560	0.480
AATSC5v	Average centered Broto-Moreau autocorrelation – lag 5 / weighted by van der Waals volumes	-2.395	0.294	8.72E-10	-8.155	0.331
MATS8m	Moran autocorrelation – lag 8 / weighted by mass	0.472	0.13	0.00084	3.635	-0.065
GATS5v	Geary autocorrelation – lag 5 / weighted by van der Waals volumes	-1.972	0.286	3.83E-08	-6.902	0.273
SpMax7_Bhs	Largest absolute eigenvalue of Burden modified matrix – n 7 / weighted by relative I-state	0.361	0.09	0.000267	4.030	-0.050
SHBint3	Sum of E-State descriptors of strength for potential hydrogen bonds of path length 3	-0.251	0.063	0.00028	-4.014	0.035
SsBr	Sum of atom-type E-State: -Br	-2.349	0.213	2.99E-13	-11.030	0.325
maxsssCH	Maximum atom-type E-State: >CH-	-10.839	0.405	7.87E-26	-26.790	1.498
ETA_dAlpha_A	A measure of count of non-hydrogen heteroatoms	12.504	0.61	8.73E-22	20.486	-1.728
RDF35u	Radial distribution function – 035 / unweighted	0.704	0.096	9.82E-09	7.347	-0.097

3.2. Model development and validation

A total of five models were developed for the chalcone antioxidant derivatives from the training set by GFA as presented in Table 2. The summary of internal validation results for these models is presented in Table 3. All the five developed models have results that exceed the threshold value of 0.5 for R^2, $R_\alpha{}^2$ and $R^2{}_{cv}$. $R_\alpha{}^2$ is a modification of R^2 in which its result is unaffected by increase in the number of descriptor terms in the model unlike R^2, except if such an increase improves the robustness of the resulting model. (Rudra and Kunal, 2012).. Thus in terms of internal model validation results, model No 1 should be recognised as the best of the five models. Never the less, this should not be accepted as the best of the five models until other validation criteria such as external validation and y-randomization test results are compared. It is worthy to note that the ability of the developed model to make good prediction of the test set activities depends on the impute parameters employed during model development. The y-randomization test results for the five developed models are presented in Table 4. From this Table, model 1 has the highest R value of 0.989. This parameter R measures how closely the observed data tracks the fitted regression line (Indrani *et al*., 2011). We observe that the five models met all the criteria for model acceptability and statistical robustness as judged by the highly encouraging results of y-randomization tests with values of $^{c}R_p^2$ well above the threshold value of 0.5 after several trials. These encouraging results for y-randomization suggest that the developed models are robust and were not obtained as a mere outcome of chance. Model 5 with $^{c}R_p^2$ value of 0.890 is recognised as the most robust of the five developed models. The external validation results for the developed models are presented in Table 5. External validation was employed in order to determine the predictive capacity of the developed model as judged by its ability to predict the activity values of the test set. These developed models passed all the Golbraikh and Tropsha criteria for model acceptability. All the external validation results were above threshold values for the various parameters as presented in Table 5.

Model 3 has the highest $r^2{}_m$ value of 0.736. The $r^2{}_m$ value determines how closely the predicted activity data fits the corresponding observed activity range (Ravichandran et al., 2011). Also model 3 has the lowest rmsep value of 0.160 and the highest $R^2{}_{pred}$ value of 0.916. Since a good model is expected to have a small value for rmsep and a high value for $R^2{}_{pred}$ and model 3 has the best external validation results in terms of these parameters, it is therefore recognised as the best of the five developed models. Recall that the choice of the best model is based on (i) meeting the requirements for internal validation and (ii) possessing the best external validation results. Based on this result, the predicted activities for the entire data set as represented in table 1 are those generated from model 3. Also a plot of the predicted activities against the experimental values are presented in Figure 1 for the training set, while that of the test set is presented in Figure 2. These plots indicate that the experimental activities are in very good agreement with and predicted activities with residual values having little deviations from zero (Table 1). The predicted activities of the training set compounds were also generated during model development. The training and test set predicted activities for the five generated models are presented in tables S1 and S2 of the supplementary data. Table 1 gives the predicted activities together with the residual and standardized residual values generated using model 3.

3.4. Results of applicability domain: A plot of standardized residuals against the corresponding leverage values gave the Williams plot. The Williams plot was employed in the estimation of the applicability domain for the entire data set as presented in Figure 3. In the Williams plot, the applicability domain was established inside a squared area in the range of ± 3 bound for residuals and a leverage threshold, h^* value of 0.688. No response outliers were detected while six structural outliers (two training and four test set compounds) were observed (Fig. 3). Predictions for these six molecules are unreliable since they are outside the applicability domain of the developed model for chalcone antioxidants (Alisi *et al*., 2018). Observe that majority of the data set are found within the applicability domain of the developed model. This result is a confirmation of the strong predictive ability of this

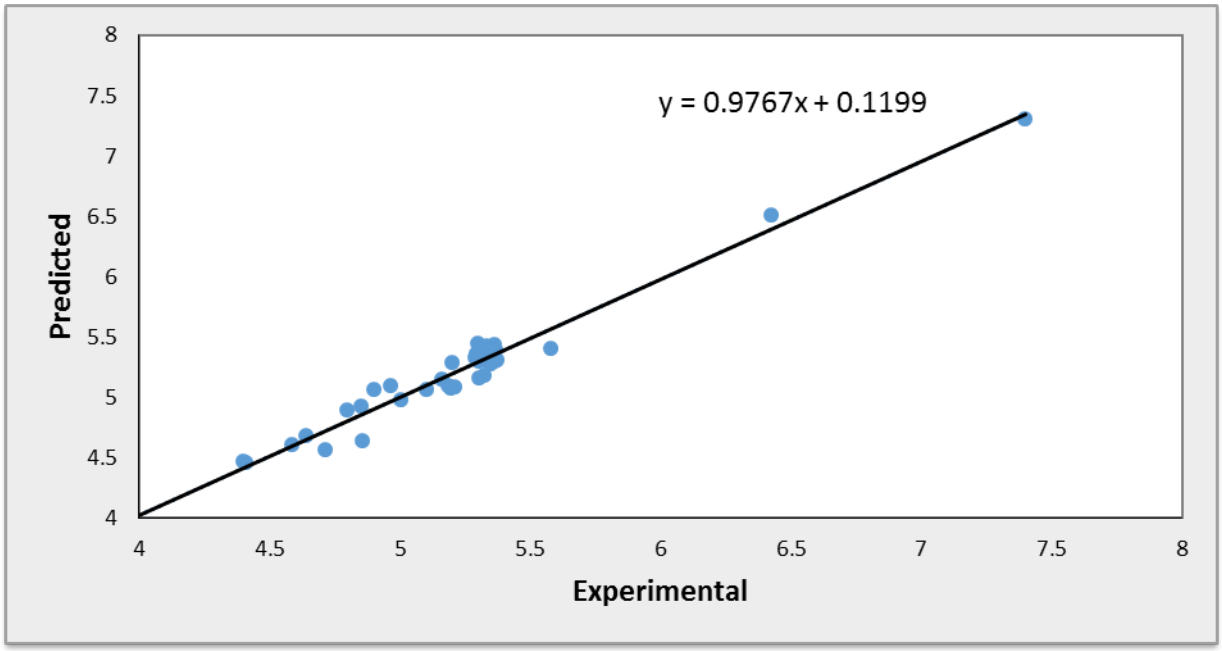
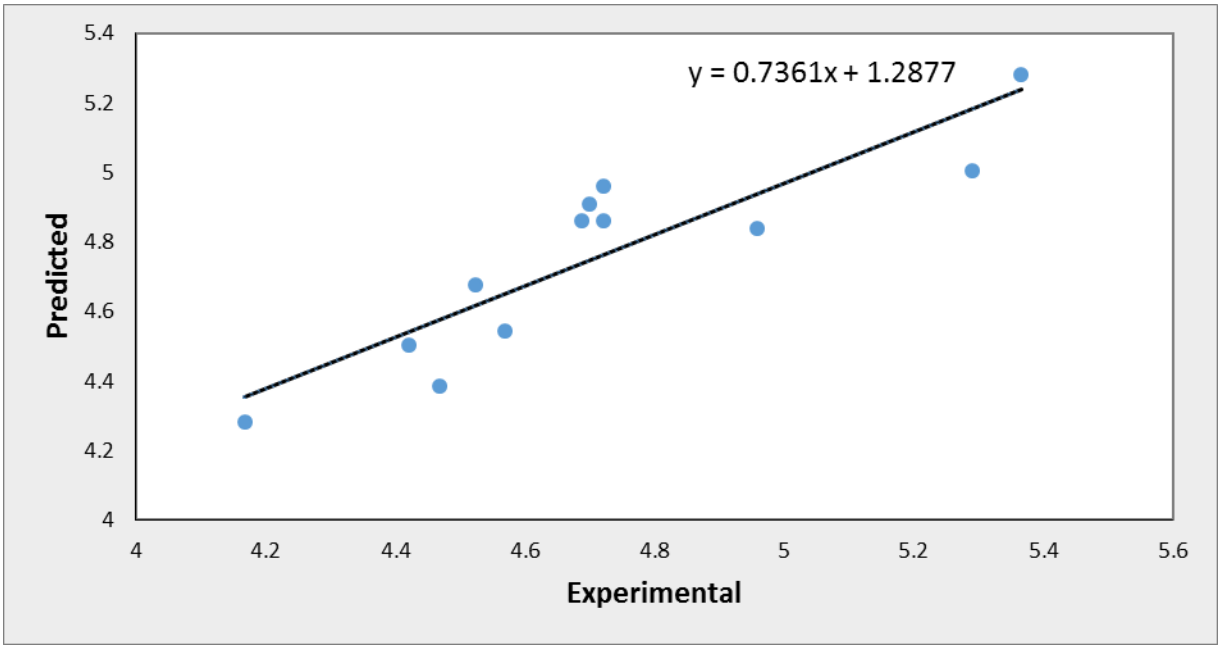


Fig. 1. Plot of predicted pIC_{50} against experimental pIC_{50} values of chalcone antioxidants training set.

Fig. 2. Plot of predicted pIC_{50} against experimental pIC_{50} values of chalcone antioxidants test set.

model. Response outliers are molecules with standard residuals greater than ±2.5 standard deviation units while, structural outliers are those molecules with $h \geq h^*$ ((Sharma and Singh, 2013; Saaidpour, 2016).



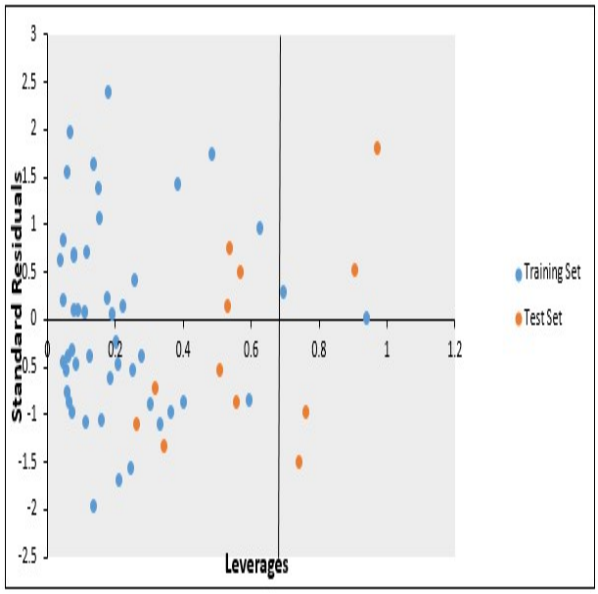


Fig. 3. Williams plot of standard residuals against leverage values for the chalcone antioxidants data set

3.5. Interpretation and significance of the descriptors in the developed QSAR model

The results for the standard error, P-value and mean effect (MF) computations for the various descriptors in the developed model are presented in table 6. Also the degree of contribution (DC) for each descriptor in the developed model was calculated and the results also given in table 5.

AATSC8m (Average centered Broto-Moreau autocorrelation – lag 8 / weighted by mass); ***AATSC5v*** (Average centered Broto-Moreau autocorrelation – lag 5 / weighted by van der Waals volumes); ***MATS8m*** (Moran autocorrelation – lag 8 / weighted by mass) and ***GATS5v*** (Geary autocorrelation – lag 5 / weighted by van der Waals volumes): These are 2D autocorrelation descriptors whose functions are applied to molecular graphs in order to measure the distribution of atomic properties (mass and van der Waals volume) on the molecule topology (Consonni and Todeschini, 2010). A weighting scheme in terms of a physicochemical property (m for mass, v for van der Waals volume) is incorporated in each descriptor. Also a number n which indicates the number of consecutively connected edges considered in the computation of the unit fragment is embedded in the nomenclature of the descriptor. From the developed QSAR model in this research, the descriptors *AATSC8m, AATSC5v* and *GATS5v* are negatively correlated with the antioxidant activities of the chalcones while *MATS8m* is positively correlated.

SpMax7_Bhs: (Largest absolute eigenvalue of burden modified matrix – n 7 / weighted by relative I-state). This is a 2D burden modified eigenvalues descriptor which is related to the molecular weight of the antioxidant. This descriptor is positively correlated with the antioxidant activities of the chalcone derivatives. This implies that an overall increase in the molecular weight of the compound improves the antioxidant activity of chalcones.

SHBint3: (Sum of E-State descriptors of strength for potential hydrogen bonds of path length 3); ***SsBr***: (Sum of atom-type E-State: -Br) and ***maxsssCH***: Maximum atom-type E-State: >CH-. These are 2D electrotopological state atom type descriptors whose indices are numerical values computed for each atom in a molecule and which encode information about both the topological environment of that atom and the electronic interactions due to all other atoms in the molecule. The topological relationship is based on the graph distance to each other atom. The electronic aspect is based on an intrinsic state plus perturbation due to intrinsic state differences between atoms in the molecule. The ***SHBint3*** descriptor defines the E-state descriptors of potential internal H-bond strength that describe the H-bond in the antioxidant molecule in spatial distance. It is associated with the Formation of 5-membered ring for potential internal H bond. The role of the number of H-bonds suggested is not an essential structural requirement for the improvement of antioxidant activity for the chalcone series based on the negative correlation of the ***SHBint3*** descriptor in the developed **QSAR** model. The ***SsBr*** descriptor indicates the sum of the atom level E-state values for all the bromine atoms in the molecule. It is negatively correlated with antioxidant activities of the chalcones. ***maxsssCH***: This descriptor is also negatively correlated with antioxidant activities of the chalcone series. This negative correlation is supported by negative sign of the DC value of -26.793 which is the highest in comparison with the other descriptors. This highest value for DC in conjunction with the high MF value of 1.498 are indications of the strength of this descriptor in the determination of antioxidant activities of the chalcones.

ETA_dAlpha_A: This is an extended topochemical atom descriptor that signifies a measure of count of non-hydrogen heteroatoms in the molecule. It has an encouraging P-value of 8.73E-22 at the 95% confidence level and it is positively correlated with the antioxidant activities of the chalcones. In comparison to the other

descriptors in the developed model, this descriptor has high DC and MF values of 20.486 and -1.728 respectively. If we recall that for a given descriptor, the higher the value of the MF, the higher the relative significance and contribution of that descriptor in the developed model. These features depict the magnitude, importance and strength of this descriptor in influencing the free radical scavenging activities of the chalcones.

RDF35u (Radial distribution function – 035 / unweighted): This is a 3D descriptor based on the radial distribution function and signifies the positive correlation of the antioxidant radical scavenging activity of the chalcones with the 3D molecular distribution of the unweighted scheme calculated at a radius of 3.5Å from the geometrical centers of each molecule. Thus the ability of a descriptor in a model to influence the activity of a compound is determined by its sign, magnitude, degree of contribution and mean effect values.

Conclusion

The antioxidant radical scavenging activities of the selected chalcones have been successfully investigated by the application of QSAR studies at the DFT level of theory. Five models were developed and subjected to various statistical validation tests. Degree of contribution and mean effect values for descriptors in the developed model were also computed. All the five developed models satisfied the various validation standards for model acceptability. Based on the results of the various validation tests conducted, model 3 was found to be the best of the five models. This model indictates that the descriptors which are relevant in the determination of the radical scavenging activities of chalcones are the autocorrelation, burden modified eigenvalues, electrotopological state atom type, topochemical atom and radial distribution function descriptors. The highly encouraging validation results obtained are indications of the good predictive ability and acceptability of the developed models. Thus this model can be applied in the design of new set of antioxidants employed in combating the dangerous effects of free radicals in the human system. We also recognise that the ability of a descriptor in a model to influence the activity of a compound is determined by its sign, magnitude, degree of contribution and mean effect values.

Conflict of Interest

The authors have declared no conflict of interest

References

Alisi, I., Uzairu, A., Abechi, S., Idris, S., 2019. Development of Predictive Antioxidant Models for 1,3,4-Oxadiazoles by Quantitative Structure Activity Relationship. JOTCSA, 6(2):103–14. https://dx.doi.org/10.18596/jotcsa.460207.

Alisi, I.O., Uzairu, A., Abechi, S.E., Idris, S.O., 2018. Free Radical Scavenging Activity Evaluation of Hydrazones by Quantitative Structure Activity Relationship. *J. Mex. Chem. Soc. 62*(*1*). DOI: http://dx.doi.org/10.29356/jmcs.v62i1.585

Alisi, I.O., Uzairu, A., Abechi, S.E., Idris, S.O., 2018. quantitative structure activity relationship analysis of coumarins as free radicalscavengers by genetic function algorithm. *Phys. Chem. Res.* 6(1), 208-222. DOI: 10.22036/pcr.2017.95755.1409

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

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

Beom-Tae, K., Kwang-Joong, O., Jae-Chul, C., Ki-Jun, H., 2008. Synthesis of Dihydroxylated Chalcone Derivatives with Diverse Substitution Patterns and their Radical Scavenging Ability toward DPPH Free Radicals. *Bull. Korean Chem. Soc.* 29(6), 1125-1130.

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

Brignole, M., Auricchio, A., Baron-Esquivias, G., Bordachar, P., Boriani, G., Breithardt, O., et al. 2013. ESC Guidelines on cardiac pacing and cardiac resynchronization therapy. *Eur Heart J.* 34(29), 2281-329. DOI: 10.1016/j.rec.2013.11.003

Consonni, V., Todeschini, R., 2010. Molecular descriptors. In: Recent advances in QSAR

studies methods and applications. Ed: Puzyn, T.; Leszczynski, J.; Cronin, M.T.D. Springer Science Business Media, New York.

Golbraikh, A., Tropsha, A., Beware of q2! *J Mol Graph Model*, **2002**, 20(4), 269–276. Doi.org/10.1016/S1093-3263(01)00123-1

Gramatica, P. 2010. 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. Springer, Dordrecht, Heidelberg, London; p. 327.

Gramatica, P., Giani, E., Papa, E., 2007. Statistical external validation and consensus modeling: A QSPR case study for KOC prediction. *J. Mol. Graphics Modell.* 25(6), 755-66. DOI: 10.1016/j.jmgm.2006.06.005

Indrani, M., Achintya, S., Kunal, R., 2011. Chemometric QSAR Modeling and In Silico Design of Antioxidant NO Donor Phenols. *Sci. Pharm.* 79, 31–57. DOI: 10.3797/scipharm.1011-02

Koleva, I.I., van Beek, T.A., Linssen, J.P.H., de Groot, A., Evstatieva, L.N., 2001. Screening of plant extracts for antioxidant activity: a comparative study on three testing methods. *Phytochem Anal.* 13, 8–17. DOI: 10.1002/pca.611

Lahsasni, S. A., Al Korbi, F. H., Aljaber, N. A., 2014. Synthesis, characterization and evaluation of antioxidant activities of some novel chalcones analogues. *Chem. Cent. J.* 8(32), 1-10. DOI: 10.1186/1752-153X-8-32

Li, Z., Wan, H., Shi, Y., Ouyang, P., 2004. Personal experience with four kinds of chemical structure drawing software: review on ChemDraw, ChemWindow, ISIS/Draw, and ChemSketch. *J Chem Inf Comput Sci.* 44(5), 1886–1890. DOI:10.1021/ci049794h

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

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., 2005. 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.* 33(2), 155–173.

Ogadimma, A.I., Adamu, U., 2016. Quantitative Structure Activity Relationship Analysis of Selected Chalcone Derivatives as *Mycobacterium tuberculosis* Inhibitors. *Open Access Library Journal*, 3, e2432. http://dx.doi.org/10.4236/oalib.1102432

Ogadimma, A.I., Uzairu, A., Eyije, A.S., Ola, I.S., 2018. Evaluation of the Antioxidant Properties of Curcumin Derivatives by Genetic Function Algorithm. *J. Adv. Res.* doi: https://doi.org/10.1016/j.jare.2018.03.003

Pachori, K., Sharma, S., Veerkar, P., Yadav, M., 2014. Inverse QSAR approach and molecular docking studies to design novel methoxy substituted chalcones and their computational anticancer activity evaluation. *Res. J. Chem. Sci.* 4(3), 76-80.

Patel, U. H., Gandhi, S. A., Barot, V. M., Patel, M. C., 2013. Synthesis, spectroscopic investigations, quantum chemical studies (*Ab-initio* and DFT) and antimicrobial activities of 3-(3-Chloro-4,5-dimethoxy-phenyl)-1-(4, 5-dimethoxy-2-methyl-Phenyl) prop-2-en-1-one. *Cryst. Struct. Theory Appl.* 2, 167-175. Doi: 10.4236/csta.2013.24023

Ravichandran, V., Harish, R., Abhishek, J., Shalini, S., Christapher P.V., Ram, K.A., 2011. Validation of QSAR models -strategies and importance. *Int. J. Drug Des. Discovery*, 2(3): 511-519.

Roy, P.P., Roy, K., 2008. On some aspects of variable selection for partial least squares regression models. *QSAR Comb. Sci.* 27, 302-313. https://doi.org/10.1002/qsar.200710043

Rudra, N.D., Kunal, R., 2012. Development of Classification and Regression Models for *Vibrio fischeri* Toxicity of Ionic Liquids: Green Solvents for the Future. *Toxicol. Res.* 1, 186-195. DOI: 10.1039/C2TX20020A

Saaidpour, S., 2016. Quantitative modelling for prediction of critical temperature of refrigerant compounds. *Phys. Chem. Res.* 4(1), 61-71. DOI: 10.22036/pcr.2016.11759

Shao, Y., Molnar, L.F., Jung, Y., Kussmann, J., Ochsenfeld, C., Brown, S.T., Gilbert, A.T.B., *et*

al., 2006. Advances in methods and algorithms in modern quantum chemistry program package. *Phys. Chem. Chem. Phys*., 8(27), 3172-3191. Doi.org/10.1039/B517914A

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

Shenvi, S., Kumar, K., Hatti, K. S., Rijesh, K., Diwakar, L., Reddy, G. C., 2013. Synthesis, anticancer and antioxidant activities of 2,4,5-trimethoxy chalcones and analogues from asaronaldehyde: Structure-activity relationship. *Eur. J. Med. Chem*. 62, 435-442. Doi: 10.1016/j.ejmech.2013.01.018

Sinha, R.J., Singh, R., Mehrotra, S., Singh, R.K., 2009. Implications of free radicals and antioxidant levels in carcinoma of the breast: a never ending battle for survival. *Indian J Cancer*, 46(2), 146-150. DOI:10.4103/0019-509X.49153

Todd, M.M., Harten, P., Douglas, M.Y., Muratov, E.N., Golbraikh, A., Zhu, H., *et al*. 2012. Does rational selection of training and test sets improve the outcome of QSAR modeling? *J. Chem. Inf. Model*. 52(10), 2570–2578. **DOI:** 10.1021/ci300338w

Todorova, I.T., Batovska, D.I., Stamboliyska, B. A., Parushev, S.P., 2011. Evaluation of the radical scavenging activity of a series of synthetic hydroxychalcones towards the DPPH radical. *J. Serb. Chem. Soc*. 76 (4), 491–497. Doi: 10.2298/JSC100517043T

Wang, G., Xue, Y., An, L., Zheng, Y., Dou, Y., Zhang, L., Liu, Y., 2015. Theoretical study on the structural and antioxidant properties of some recently synthesised 2,4,5-trimethoxy chalcones. *Food Chemistry*, 171, 89–97. DOI: 10.1016/j.foodchem.2014.08.106

Xue, Y., Zheng, Y., An, L., Zhang, L., Qian, Y., Yu, D., Gong, X., Liu, Y., 2012. A theoretical study of the structure–radical scavenging activity of hydroxychalcones. *Comput. Theor. Chem.*, 982, 74–83. https://doi.org/10.1016/j.comptc.2011.12.020

Yap, C.W., 2011. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J Comput Chem*. 32(7), 1466–1474. Doi: 10.1002/jcc.21707

Zhang, K., Ding, W., Sun, J., Zhang, B., Lu, F., Lai, R., Zou, Y., Yedid, G., 2014. Antioxidant and antitumor activities of 4-arylcoumarins and 4-aryl-3,4-dihydrocoumarins. *Biochimie*, 107, 203-210. DOI: 10.1016/j.biochi.2014.03.014.