Moroccan Journal of Chemistry
ISSN: 2351-812X
http://revues.imist.ma/?journal=morjchem&page=login
ZGou & al. / Mor. J. Chem. 8 N°4 (2020) 841-852

QSAR modeling of antiradical properties of phenolic compounds using DFT calculations

Hind. Lafridi[(a)], Abdellah. Ousaa[(b)], Hsaine. Zgou[(a)]*, Mohammed. Bouachrine[(b)]

[(a)] Material Sciences, Processes, Environment& Modeling, Ibn Zohr University, B.P. 638, Ouarzazate, Morocco
[(b)] Molecular Chemistry and Natural Substances Laboratory, University Moulay Ismail, B.P. 11201 Meknes, Morocco.
[(b)] Materials, Environment & Modeling Laboratory, Moulay Ismail University, B.P. 50040 Meknes, Morocco.

** Corresponding author:*
h.zgou@uiz.ac.ma
Received 23 April 2020,
Revised 03 June 2020,
Accepted 15 Sept 2020

Abstract

This paper describes a quantitative structure – activity relationship study of the antiradical properties of 31 flavonoids belonging to different groups such as: flavonols, flavonones, dihydroflavonols and biflavonones. Using density functional theory (DFT) calculations, some structural characteristics such as frontier molecular orbitals, molecular descriptors, have been studied. To gain insights into the chemical structure and property of the studied compounds, many types of descriptors are generated by using DFT/B3LYP 6-31G(d,p) and other software. Also, The Principle Component Analysis (PCA), Multiple Linear and Nonlinear Regression (MLR and MNLR), and Artificial Neural Network (ANN) have been investigated to select the descriptors, and to generate the correlation models that relate the structural feature to the biological activity. The statistical results of the MLR, MNLR, and ANN indicate that the determination coefficient R^2 were 0.811, 0.646, 0.982, respectively. A good correlation coefficient is obtained, and the antiradical activities of these compounds are well predicted. These models are expected to be useful for screening of polyphenolic antioxidants.

Keywords: Flavonoid derivatives; 2D-QSAR; DFT Study; Artificial neural network, PCA, MLR, MNLR, Domain of Applicability

1. Introduction

Flavonoids are the major classes of phenolic compounds of plants, which are regarded as a kind of natural constituents that has been the subject of considerable scientific interest. Many studies [1-2] have suggested that these compounds have the property of inhibiting autoxidation reactions and scavenging of free radicals exhibit health promoting properties, such as antioxidant, anti-inflammatory, antibacterial, antifungal, and anticancer activities [3-4]. The chemical structure of flavonoids and the relationship between their structure and their activity play an important role in the antiradical potentials of this classes, such as: the ortho-hydroxylation on the ring B, the hydroxyl substitution in ortho-position in the ring B, the number of free hydroxyl groups, the double bond in the ring C (C2-C3) [5-6]. In the present paper, we propose to study, from the theoretical point of view, the antiradical properties of the mentioned phenolic compounds listed in Table 1. It is broadly recognized that the antiradical activity of phenolic compounds is due to the presence of aromatic OH groups [7]. Several research studies have also shown that some CH bonds play an important role in the antiradical properties of many natural product [8]. However, to understand which properties are important to control a specific biological activity of flavonoids, and by which mechanism these compound operate for free radical scavenging for inhibiting oxidation, we have extended this work towards the development of Quantitative Structure- Activity Relationship (QSAR) on the basis of relevant structural descriptors such as reactivity properties and chemical reactivity indexes in density functional theory of these compounds. Praveena et al., indicated that molecular descriptors such as hardness (η), electronegativity(χ) and electophilic (ω) can be used to characterize the intrinsic antiradical properties of phenolic compounds [9]. To improve the estimate quality of the activity of flavonoids, molecular descriptors which reflect others specific interactions are also included. To the best of our knowledge, there is no study describing a comparative DFT and QSAR studies of the antiradical properties of the selected compounds. The computational method DFT-B3LYP, with the basis set 6-31G (d,p) was applied in order to calculate the structure and reactivity properties for the study compounds. We have computed some structural characteristics such as molecular descriptors (energy E_{HOMO}, energy E_{LUMO}, energy gap Δ_{EL-H}, dipole moment μ, total energy ET). The best descriptors were selected to establish the quantitative structure activity relationship (QSAR). The present study was performed using namely, the principal components analysis (PCA), and performed using stepwise and multiple linear regression method (MLR), multiple non-linear regression method (MNLR), creation and training of artificial neural networks (ANN), which would give us a clear idea about the relationship between the antiradical activity and the structural patterns of these compounds. A total of thirty-one compounds were tested for DPPH radical scavenging antiradical activity and their activity data was used to develop QSAR models, Table 1 displays structures as well as the IC_{50} (sample concentration having 50% radical inhibition activity) values of these compounds.

2. Materials and methods

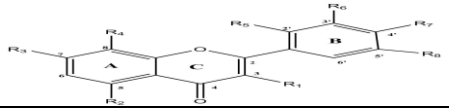
2.1. Data sources

In the present study, we choose 31 substituted Flavonoids group (the largest subset containing 21 flavonols, 6 flavanones, and 4 biflavanones), listed in Table 1, for which their radical scavenging activity tested in a methanolic solution of DPPH are reported in the literature by Stanislaw et al., [10] .The radical scavenging activity was expressed with IC_{50}, defined as the concentration of the samples needed to cause 50% scavenging of DPPH radical, calculated by an equation generated from linear regression [11]. From the values of pIC_{50} presented by these compounds, it can be seen that the compounds have their activities between 2.56 and 4.77(μM). For the proper validation of our data set with a QSAR model, the substituted phenols were divided into training and tests

sets. 25 molecules are considered as training set to build QSAR models while remaining 6 molecules are taken as test set. Structures of flavonoids and antiradical activity values are shown in Table 1.

Table 1. Structures and antiradical activity of tested flavonoids

Structure of flavonols and flavones



N°	Name (IUPAC)	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	pIC_{50}
1	Kaempferol	OH	OH	OH	H	H	H	OH	H	4.75
2	Galangin	OH	OH	OH	H	H	H	H	H	4.75
3	Quercetin	OH	OH	OH	H	H	OH	OH	H	4.74
4	Kaempferol 3,7-O-dirhamnoside	O-rha	OH	O-rha	H	H	H	OH	H	4.63
5	Robinetin	OH	H	OH	H	H	OH	OH	OH	4.70
6	Fisetin	OH	H	OH	H	H	OH	OH	H	4.68
7	3-hydroxyflavone	OH	H	H	H	H	H	H	H	4.60
8	Laricytrin	OH	OH	OH	H	H	OH	OH	O-Me	4.71
9	Laricytrin 3'-O-gucoside	OH	OH	OH	H	H	O-glu	OH	O-Me	4.71
10	Myricetin	OH	OH	OH	H	H	OH	OH	OH	4.64
11	3,5,7,3',4',5'-hexamethoxyflavone	O-Me	O-Me	O-Me	H	H	O-Me	O-Me	O-Me	3.88
12	Quercetin 3-O-glucoside-7-O-rhamnoside	O-glu	OH	O-rha	H	H	OH	OH	H	4.72
13	Rutin	O-rut	OH	OH	H	H	OH	OH	H	4.72
14	Morin	OH	OH	OH	H	OH	H	OH	H	4.77
15	5-hydroxyflavone	H	OH	H	H	H	H	H	H	2.56
16	7-hydroxyflavone	H	H	OH	H	H	H	H	H	3.23
17	Chrysin	H	OH	OH	H	H	H	H	H	2.82
18	8-methoxyflavone	H	H	O-Me	H	H	H	H	H	2.63
19	Apigentin	H	OH	OH	H	H	H	OH	H	2.63
20	Apigentin 8- C-glucoside	H	OH	OH	glu	H	H	OH	H	4.10
21	Apigentin 7- O-glucoside	H	OH	O-glu	H	H	H	OH	H	4.32
Structure of flavanones										
22	Luteolin 7-O-glucoside	H	OH	O-glu	H	H	OH	OH	H	4.73
23	Flavanone	H	H	H	H	H	-	-	-	3.20
24	Naringenin	H	OH	OH	H	OH	-	-	-	3.58
25	Hesperetin	H	OH	OH	OH	O-Me	-	-	-	4.26
26	Fustin (dihydrofisetin)	OH	H	OH	OH	OH	-	-	-	4.75
27	Taxifolin (dihydroquercetin)	OH	OH	OH	OH	OH	-	-	-	4.76
	Structure of biflavanones									
28	GB-1 (biflavanone)	OH	OH	H	-	-	-	-	-	3.76
29	GB-la (biflavanone)	H	OH	H	-	-	-	-	-	3.83
30	GB-2 (biflavanone)	OH	OH	OH	-	-	-	-	-	4.75
31	GB-2a (biflavanone)	H	OH	OH	-	-	-	-	-	3.53

2.2. Methodology

Theory and computational details

Molecular structure of the phenolic compound was designed with the program GAUSSVIEW 3.0. All the computational studies have been carried out using the Density Functional Theory (DFT) methods implemented in the GAUSSIAN 03W package [12]. The B3LYP functional [13] and the 6-31G (d,p) basis set have been used for calculating the geometry optimization. In order to determinate the ability of phenolic compounds to act through the antiradical activity, we calculated as many physicochemical descriptors as possible by using available computational packages, as we do not know beforehand which properties are closely related to the biological activity. A number of quantum-chemical descriptors , the highest occupied molecular orbital energy E_{HOMO} (ev), the lowest unoccupied molecular orbital energy E_{LUMO} (ev), the energy gap $\Delta E_{L\text{-}H}$(ev), the dipole moment μ (Debye), the total energy ET (ev) have been chosen and have been obtained from density functional theory (DFT) calculations. These descriptors give information about the electronic structure of the molecule and they mainly describe the electronic interactions.

Other molecular descriptors have been calculated using the Chemsketch program [14] and Chemoffice programs [15] such as: Density D (g/cm3), the surface tension γ (Dyne/cm), the absolute hardness η (ev), the absolute electronegativity χ (ev), the electrophilicity ω (ev), the critical pressure CP (Bar), the number of rotatable bonds Nrotb, the polar surface area PSA $(A°)^2$, the wiener index WI ÷ Thus, the descriptors generated are shown in Table 2.

$$\eta = (E_{LUMO} - E_{HOMO})/2 \; ; \chi = (E_{LUMO} + E_{HOMO})/2 \; ; \quad \omega = \mu^2 / 2\eta \qquad (1)$$

Main computational Methods

In QSAR modeling of radical scavenging properties of phenolic compounds, we have investigated polynomial and multiple linear relationships between the following descriptors : surface tension ($^{\gamma}$); Density (D); Total energy (ET); HOMO energy (E_{HOMO}); LUMO energy (E_{LUMO}); dipole moment (μ); Gap energy ($\Delta E_{L\text{-}H}$); hardness (η); electronegativity (χ); electrophilicity (ω); Critical Pressure (CP); polar surface area (PSA); Wiener Index (WI) number of rotatable bonds (Nrotb) of investigated compounds and antiradical activity depended. In order to select out the predominant descriptors affecting the antiradical activity, the correlation analysis has been performed by using statistical methods based on the principal component analysis (PCA) [16]. To eliminate independent descriptors that are highly correlated and to minimize the information overlap in the data set, the descriptors with lower inter-correlation (|r|<5) were considered [17]. The descriptors with higher correlation to the pIC_{50} and lower inter-correlation have been selected to carry out the descendant and stepwise multiple linear regression analysis to establish the optimal QSAR equation. The multiple linear regression (MLR) analysis with descendant and stepwise selection has been employed in order to obtain a correlation between the putative descriptors and the experimental radical scavenging activity [17]. MLR and MNLR are generated using software XLSTAT version 2009 [18]. To estimate the quality of the model, various statistical indicators analyses such as the correlation coefficient R, determination coefficient (R^2), adjusted coefficient (R_{adj}^2), Mean squared Error (MSE), significance level (p-value), and Fischer test F have been employed.The artificial neural networks (ANN) [19] analysis has been used to incorporate nonlinear dependencies between the dependent and independent variables without using an explicit mathematical function. The network was built for three layers and organized in a layered topology as follows: (i) the input layer formed by a number of neurons equal to the number of descriptors obtained in the multiple linear regression models (In this work i=14), (ii) the output layer (in this work m=1) consisting of the calculated activity values, and (iii) the hidden layers (between them). The number of artificial neural in the hidden layer was adjusted experimentally.

The ANN analysis has been performed using MATLAB software version 2009a Neural Fitting tool (nftool) toolbox [20]. Testing the stability, predictive capacity and generalization ability of the models are notably important steps in a QSAR study. Internal validation, the leave-on-out (LOO) cross validation has been used for this propose in which one compound is removed from the dataset and rebuilding the model [21]. The square of LOO Q^2 wish is used as a criterion should be characterized by a value of $Q^2_{cv} > 0.5$ for a satisfactory model and for an excellent model when $Q^2_{cv} > 0.9$ [22].An external validation was also used. Thus, dataset of 31 compounds was divided into a training set comprising randomly selected 26 compounds and a test set comprising the rest 6 compounds. QSAR model was built from training set, and it was determined with the test set to confirm its predictive ability. To detect outliers and outsides compounds, the applicability domain of MLR models were defined [23].

Table 2. Values of chemical descriptors.

Cpd.code	γ	D	ET	E_{HOMO}	E_{LUMO}	μ	ΔE_{L-H}	η	χ	ω	CP	Nrotb	PSA	WI
Test Set														
1	98.9	1.688	-27998.81	-5.294	-1.493	3.0786	3.801	1.9006	3.394	10.944	63.796	1	107.22	872
5	114.8	1.799	-30045.64	-5.615	-1.789	3.2191	3.827	1.9133	3.702	13.113	82.196	1	127.45	987
8	100.2	1.73	-33161.94	-5.166	-1.484	1.789	3.682	1.841	3.325	10.180	65.991	2	136.68	1245
12	125.2	1.82	-61236.15	-5.642	-1.717	8.4589	3.925	1.962	3.679	13.284	31.811	6	265.52	6336
16	58.2	1.34	-21858.60	-6.307	-1.711	3.228	4.596	2.298	4.009	18.472	37.777	1	46.53	586
22	79.4	1.555	-44616.88	-5.622	-1.269	4.831	4.352	2.176	3.446	12.918	44.15	4	186.37	2997
Training Set														
2	84.9	1.579	-25952.04	-5.522	-1.624	4.5321	3.898	1.949	3.573	12.440	50.948	1	86.99	747
3	114.8	1.799	-30045.55	-5.242	-1.485	1.7935	3.757	1.879	3.363	10.627	82.196	1	127.45	986
4	103.6	1.7	-57142.56	-5.398	-1.196	8.0759	4.202	2.1008	3.297	11.417	24.654	5	225.06	5622
6	98.9	1.688	-27999.02	-5.351	-1.743	5.3135	3.608	1.804	3.547	11.347	63.796	1	107.22	878
7	62	1.367	-21858.71	-5.787	-1.954	3.0829	3.833	1.916	3.870	14.355	34.643	1	46.53	565
9	106.2	1.761	-49780.28	-5.379	-1.683	4.8193	3.696	1.848	3.531	11.518	41.303	5	215.83	3608
10	133	1.912	-32092.33	-5.237	-1.523	2.2796	3.714	1.857	3.380	10.610	109.876	1	147.68	1104
11	51.4	1.3	-38508.90	-5.431	-1.302	2.7907	4.129	2.065	3.366	11.697	15.451	7	81.68	2078
13	125.2	1.82	-61236.04	-5.424	-1.625	14.3497	3.799	1.899	3.524	11.796	34.561	6	265.52	6416
14	114.8	1.799	-30045.64	-5.587	-1.415	8.7817	4.171	2.085	3.501	12.782	82.196	1	127.45	973
15	58.2	1.34	-21858.36	-6.045	-1.603	3.9971	4.441	2.221	3.824	16.237	37.777	1	46.53	580
17	68.2	1.443	-23905.11	-6.025	-1.530	3.6707	4.495	2.247	3.778	16.039	45.777	1	66.76	670
18	47.7	1.24	-22928.18	-6.247	-1.682	3.7449	4.564	2.282	3.965	17.937	27.585	2	35.53	690
19	79.5	1.548	-25951.90	-5.866	-1.388	4.2981	4.478	2.239	3.627	14.731	56.617	1	86.99	788
20	99	1.686	-42570.61	-5.871	-1.379	3.3085	4.491	2.246	3.626	14.762	45.164	3	177.14	2446
21	71.8	1.487	-42570.15	-5.726	-1.287	5.6285	4.439	2.219	3.506	13.643	36.554	4	166.14	2754
23	46	1.192	-19844.59	-6.330	-1.535	2.5174	4.795	2.397	3.933	18.540	29.861	1	26.3	500
24	72.8	1.485	-25984.68	-5.914	-1.037	3.5481	4.876	2.438	3.475	14.727	52.282	1	86.99	788
25	67.4	1.485	-29101.04	-5.648	-0.968	2.6225	4.680	2.340	3.308	12.805	43.8	2	96.22	1038
26	86.9	1.599	-28031.35	-5.954	-1.631	3.2234	4.323	2.161	3.792	15.543	57.046	1	107.22	878
27	100.4	1.702	-30078.09	-5.544	-1.352	2.9232	4.192	2.096	3.448	12.463	72.431	1	127.45	986
28	93.6	1.65	-51936.32	-5.251	-1.079	1.5668	4.172	2.086	3.165	10.451	44.444	3	177.14	4715
29	85.3	1.589	-49889.83	-5.199	-1.061	1.2062	4.138	2.069	3.130	10.138	41.144	3	156.91	4450
30	101	1.706	-52758.68	-5.233	-1.074	2.9706	4.158	2.079	3.154	10.340	54.788	3	197.37	5046
31	92.1	1.643	-51936.59	-5.184	-1.047	2.583	4.136	2.068	3.116	10.039	50.299	3	177.14	4775

3. Results and Discussions

3.1. Quantitative structure – activity relationship equation

Multiple Linear Regressions (MLR)

Optimal model equations for 25 compounds in training set by multiple linear regressions were obtained with good statistical parameters. Experimental and calculated pIC_{50} values using obtained models and associated 95%

confidence intervals are shown in Table 3. Among the D, E_{LUMO}, W, NRB, PSA, was found to correlate well with pIC_{50} values of analyzed polyphenols. The QSAR models built using descendant multiple linear regression methods are expressed through the following equations:

Descendant MLR:

$pIC_{50} = 14 - 6.61*D + 10^{-4}*ET - 1.73* E_{LUMO} - 0.24*\omega - 0.68*NRB + 0.051*PSA$ (2)

$N_{training}$= 25; R = 0.803; R^2 = 0.811; R_a^2 = 0.748; $Q^{2=}$0.600; MSE= 0.138; F=12.90; p< 0.0001.

N_{test} = 6; R = 0.987; R^2 = 0.975; $Q^{2=}$0.822; MSE test= 0.015.

Where N is the number of compounds in training set; R^2 is the square of correlation coefficient regression; R^2_a is the square of adjusted correlation coefficient; R is the correlation coefficient; F is the Fisher's F using the F statistics; p is the p value using the F statistics; and Q^2 is the square of LOO cross validation coefficient. The above established QSAR model shows significant statistical quality. A higher correlation coefficient and a lower mean squared error indicate that the model is more reliable [24]. The leave on-out-cross-validated correlation coefficient LOO (Q^2 =0.686), greater than 0.50 indicates the reliability of the model and qualifies the model as valid. The plot of the observed versus predicted pIC_{50} values in Eq.2 of Flavonoids compounds are depicted in Figure 1.

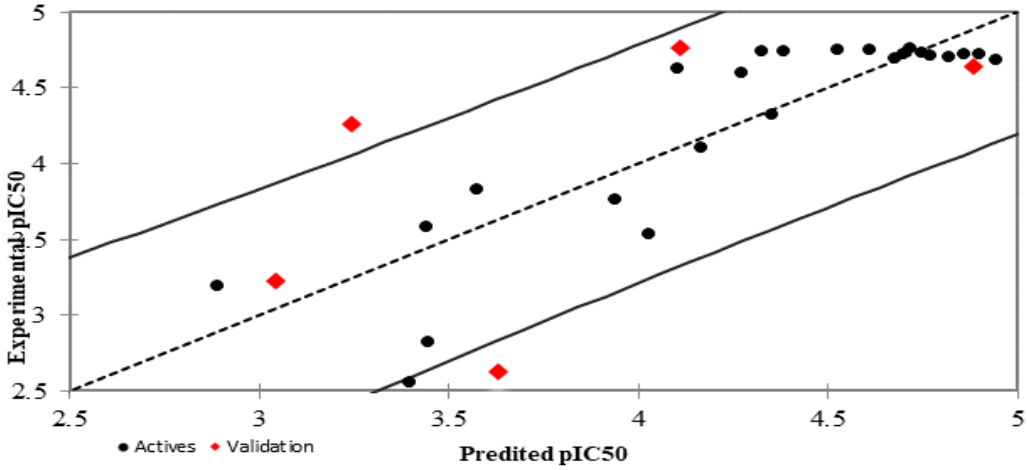


Figure 1. Plot of predicted activities versus experimental ones for the QSAR model number 1 Eq. (2), in wish 25 compounds are the training set (dot) and corresponding 6 compounds are the test set.

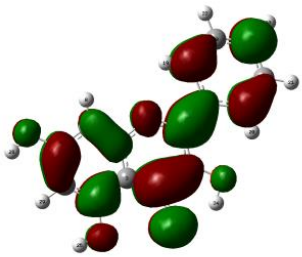
Since the factors (D, E_{LUMO}, W, Nrotb) are correlated negatively with the value of the pIC_{50} in Eq.2 shows that decreasing of the value of this factors can lead to an increase of the ability to scavenge the $ABTS^{\cdot+}$radical cation. The positive correlation of the factors ET and PSA with the value of radical scavenging activity shows that an increase in the values of these descriptors involves the value of the pIC_{50}.Frontier molecular orbitals; highest occupied molecular orbital HOMO and Lowest unoccupied molecular orbital LUMO are important parameters that characterizes the antiradical activity of organic compounds [25]. The energy of the LUMO describes the electron-accepting character of a substance. The lower energy of LUMO implies the compound more easily accepts electron in charge transfer processes that can occur when a compound interacts with a biological target [26].These compounds are electron-acceptors and can easily acceptor electrons to macromolecules such as DNA. From Table 1 and 2, we can observe that the compounds 2, 5, 6, 7, 9,12 and 13 has the lower LUMO energy, the higher the activity of the compound, and thus the stronger DNA-binding affinity is. In order to explain the action mechanism, the spatial LUMO distributions of some compounds as example were drawn as shown in Figure 2, we can see that the atomic contributions of LUMO orbital of the compounds are all distributed on the main conjugative part (composed by three aromatic rings A, B, and C). We can clearly find that, the LUMO's of compounds 2, 7,9,12 wish

is flavonol and flavones derivatives, are spread along the tree aromatics rings (A, B, and C rings). The exceptions are flavanone derivatives, the main atomic contributions for LUMO are localized on the two aromatic rings A and C (example: compound 27).

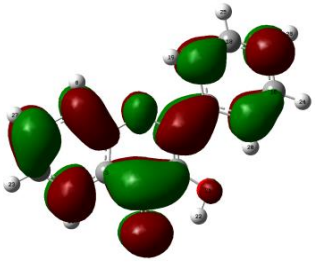
Table 3. Experimental (obs) and calculated (calc.) pIC50 values using models expressed by Eq (3)-(5) with residual (resid.). [a] Experimental values were taken from Stanislaw et al., 2001.

N° of Compd.	pIC_{50obs} [a] µM	Model number 1		Model number 2		Model number 3		CV	
		$pIC_{50calc.}$	Resid.	$pIC_{50calc.}$	Resid.	$pIC_{50calc.}$	Resid.	$pIC_{50calc.}$	Resid.
1*	4.753	4.609	0.144	4.569	0.185	4.750	0.00	4.633	0.117
2	4.745	4.385	0.361	4.287	0.458	4.768	-0.02	4.633	0.117
3	4.736	4.748	-0.012	4.815	-0.079	4.718	0.02	4.633	0.107
4	4.631	4.102	0.530	4.295	0.337	4.630	0.00	4.633	-0.003
5*	4.698	4.676	0.022	4.763	-0.065	4.700	0.00	4.633	0.067
6	4.680	4.944	-0.264	4.712	-0.032	4.680	0.00	4.633	-0.153
7	4.602	4.274	0.328	3.928	0.674	4.600	0.00	3.096	1.504
8*	4.710	4.772	-0.063	4.738	-0.028	4.740	- 0.03	4.633	0.077
9	4.706	4.818	-0.113	4.787	-0.082	4.710	0.00	4.633	0.077
10	4.645	4.883	-0.238	5.064	-0.419	4.640	0.00	4.633	0.007
11	3.883	0.184	3.699	3.579	0.304	3.880	0.00	3.880	0.000
12*	4.721	4.700	0.021	4.731	-0.010	4.720	0.00	4.633	0.087
13	4.721	4.900	-0.179	4.825	-0.103	4.720	0.00	4.633	0.087
14	4.767	4.111	0.656	4.508	0.259	4.770	0.00	4.633	0.137
15	2.561	3.395	-0.834	3.424	-0.864	2.560	0.00	3.096	-0.536
16*	3.230	3.043	0.186	3.309	-0.080	3.230	0.00	3.096	0.134
17	2.824	2.824	-0.623	3.583	-0.759	2.689	0.14	3.096	-0.276
18	2.628	2.628	0.194	3.141	-0.513	2.630	0.00	3.096	-0.466
19	2.628	2.628	-1.005	3.797	-1.170	2.672	0.04	3.096	-0.466
20	4.105	4.166	-0.061	4.053	0.052	4.345	0.24	4.100	-0.000
21	4.324	4.353	-0.029	3.709	0.615	4.286	0.04	4.633	-0.313
22*	4.725	4.859	-0.134	3.904	0.821	4.730	0.00	4.633	0.097
23	3.198	2.888	0.309	2.878	0.320	3.200	0.00	3.096	0.104
24	3.582	3.441	0.141	3.381	0.201	3.580	0.00	3.580	-0.000
25	4.260	3.245	1.015	3.526	0.733	4.717	0.46	4.260	0.000
26	4.746	4.327	0.419	4.011	0.735	4.750	0.00	4.633	0.117
27	4.759	4.717	0.043	4.306	0.454	4.770	0.01	4.633	0.127
28	3.760	3.940	-0.180	4.221	-0.460	3.760	0.00	4.633	-0.873
29	3.832	3.576	0.256	4.128	-0.297	3.830	0.00	3.830	-0.000
30	4.750	4.527	0.223	4.338	0.412	4.750	0.00	4.633	0.117
31	3.531	4.030	-0.499	4.234	-0.703	3.707	0.18	4.633	0.117

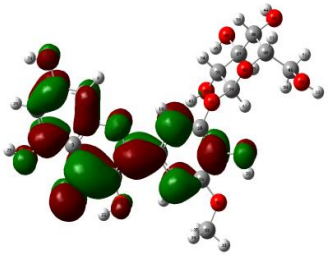
That is to say, the presence of a free hydroxyl in C-3 position had a high ability to scavenge DPPH radicals. Also, the presence of a free hydroxyl group at the position C-4' is essential for the antiradical activity of this group of flavonoids. Also, the antiradical effectiveness is strengthened by double bond in the C-ring (C2-C3).



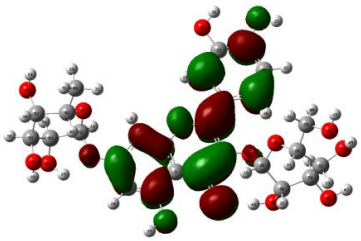
Compound 2 (pIC_{50}= 4.75)



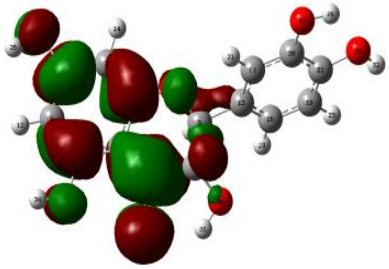
Compound 7 (pIC_{50}= 4.68)



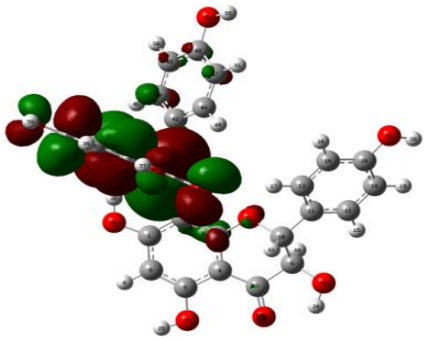
Compound 9 (pIC_{50}= 4.71)



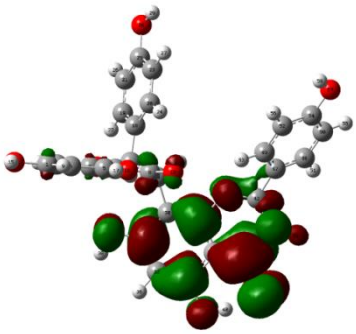
Compound 12 (pIC_{50}= 4.72)



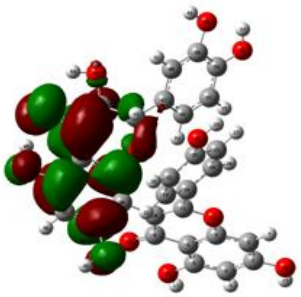
Compound 27 (pIC_{50}= 4.76)



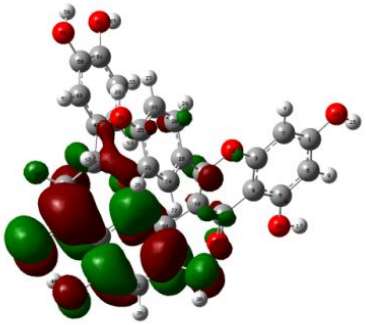
Compound 28 (pIC_{50}= 3.76)



Compound 29 (pIC_{50}= 3.83)



Compound 30 (pIC_{50}=4.75)



Compound 31 (pIC_{50}= 3.53)

Figure 2. LOMO plots of some flavonoid compounds.

In order to increase the probability of good characterization of studies compounds, stepwise regression model was used to generate predictive models of quantitative structure–activity relationships (QSAR) between a set of molecular descriptors obtained and the observed activity. The correlations coefficients obtained with the stepwise regression (R = 0.900), show that the selected descriptors by stepwise regression are pertinent and that the model proposed to predict activity is relevant. The leave on-out-cross-validated correlation coefficient LOO (Q^2 =0.686), greater than 0.50 indicates the reliability of the model and qualifies the model as valid. Stepwise regression resulted in the following statistically significant model using density (D) and hardness (η) as descriptors:

Stepwise MLR:

pIC50 = 4. 14+1, 92*D-1. 48*η (3)

$N_{training}$ =25; R = 0.900; R^2 = 0.646; R_a^2 = 0.612; S=0.274; Q^2= 0.500; MSE= 0.218; F= 20.68; p-value < 0.0001.
N_{test} =6; R = 0.834; R^2 = 0.69; Q^2(LOO)=0.623; MSE= 0.119. The deviations of regression of model number 2 are also listed in Table 3. The density is related to the size of the molecules. A positive coefficient (+1.92) in Eq.3 reflects the types of atoms and how tightly they are packed in a molecule. The density can be related to transport and melt behavior. A negative coefficient of the hardness (-1. 48) in Eq.3 reflects stereochemical hindrance between substituents and free radical. The obtained QSAR shows that the lower the hardness value the higher the activity. The plot of the observed versus predicted pIC_{50} values in Eq.3 of the selected compounds are depicted in Figure 3.

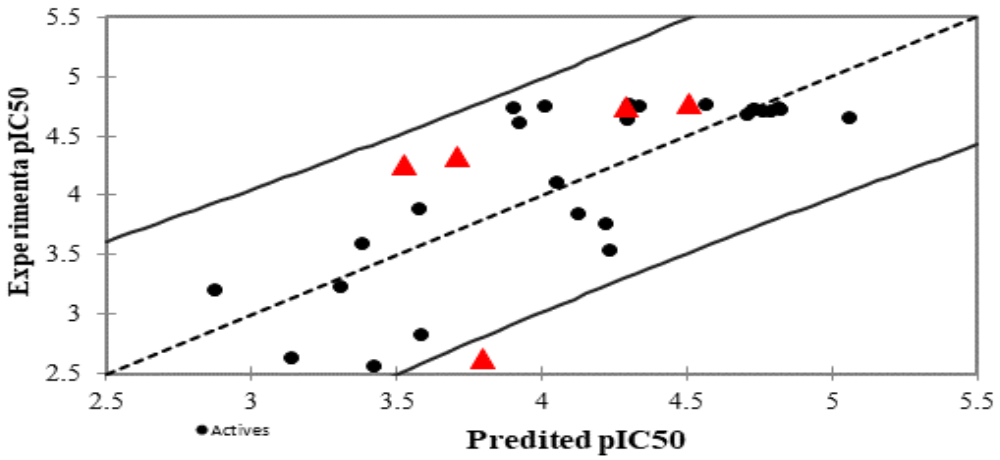


Figure 3. Plot of predicted activities versus experimental ones for the QSAR model number 3 Eq. (3), in wish 25 compounds are the training set (dot) and corresponding 6 compounds are the test set.

Artificial Neural Networks (ANN)

In order to increase the probability of good characterization of studies compounds, neural networks (ANN) was used to generate predictive models of quantitative structure–activity relationships (QSAR) between a set of molecular descriptors obtained from the MLR and observed activity. The obtained correlation coefficients value with the Neural network is 0.991 for this data set of flavonoids, show that the selected descriptors by MLR are pertinent and that the model proposed to predict activity is relevant. The cross validated squared correlation coefficient Q^2=0.918 and mean squared error 0.0128 suggest the good internal consistency of the biological activity.
$N_{training}$=25; R = 0.991; R^2 = 0.982; MSE= 0.0128; Q^2 (LOO) = 0.918.
N_{test}=6; R = 0.990; R^2 = 0.999; MSE= 0.000; Q^2 (LOO) = 0.999
The correlation of the predicted and observed activities and the residual graph of absolute numbers are shown in Figure 4. The parameters of the performance of the generated models are shown in Table 4. It can be seen that the ANN is statistically better than the MLR model and has a better predictive ability and good internal stability. Also the ANN model has the highest cross validation coefficient (Q^2=0.918). However, both the results obtained by the MLR should be regarded as satisfactory for predicting antiradical scavenging activity using the proposed descriptors.

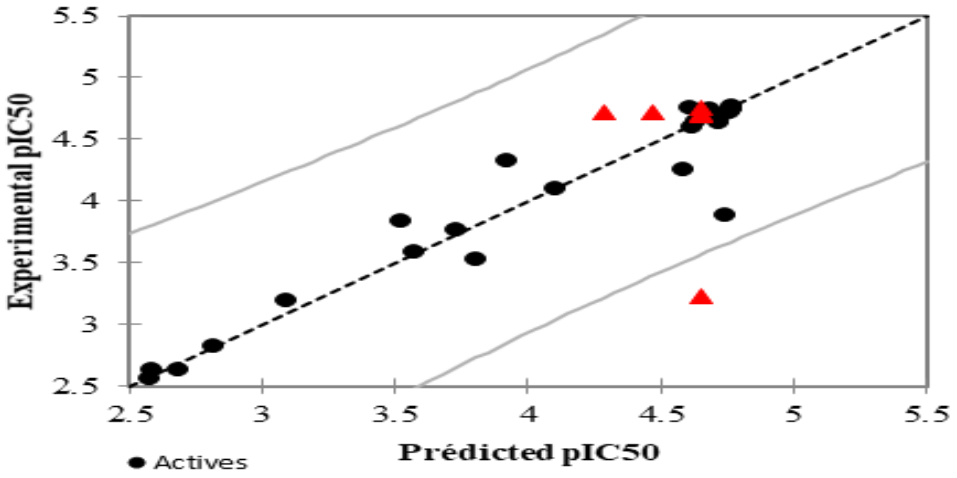


Figure 4. Plot of predicted activities versus experimental ones for the ANN QSAR model, in wish 25 compounds are the training set (dot) and corresponding 6 compounds are the test set.

Table 4. Performance comparison between models obtained by MLR and ANN

Model	Training set				Test set			
	R	R^2	Q^2	MSE	R	R^2	Q^2	MSE
MLR descendant	0.803	0.811	0.600	0.138	0.987	0.975	0.822	0.015
MLR stepwise	0.900	0.646	0.500	0.218	0.834	0.69	0.623	0.119
ANN	0.991	0.982	0.918	0.0128	0.990	0.999	0.999	0.000

External validation: Domain of applicability

To estimate the reliability of any QSTR model and its ability to predict new compounds, the domain of applicability must be essentially defined. The predicted compounds that fall within this domain may be considered as reliable [27]. The applicability domain was discussed with the Williams graph in Fig.5, which the standardized residuals and the leverage values (h_i) are plotted. It is based on the calculation of the leverage h_i for each molecule, for which QSAR model is used to predict its activity:

$$h_i = x_i (X^T X)^{-1} x_i^T \qquad (i = 1, \ldots, n) \qquad (4)$$

Where x_i is the row vector of the descriptors of compound i and X is the variable matrix deduced from the training set variable values. The index T refers to the matrix/vector transposed. The critical leverage h^* is, generally, fixed at 3(k+1)/N, where N is the number of training molecules, and k is the number of model descriptors. If the leverage value h of molecule is higher than the critical value (h^*) i.e., $h > h^*$, the prediction of the compound can be considered as not reliable. The Williams plot for the presented MLR model is shown in Figure 5. From this plot, the leverage values (h_i) of any molecule in the training and test sets are less than the critical value ($h^* = 0.28$) excepting the compound 3 is outliers. Also, the standardized residuals of all molecules in the training and test sets are less than three standard deviation units (±2σ). Thus, the predicted activity by the developed MLR model is reliable.

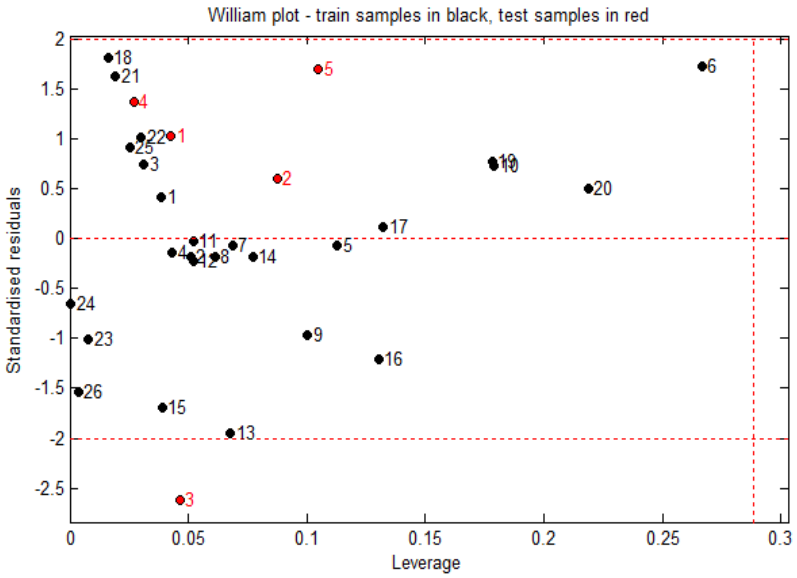


Figure 5. Williams plot for the presented MLR model.

4. Conclusion

In this study, two different modelling methods, MLR and ANN were used in the construction of a QSAR model to predict the antiradical activity of some flavonoids and the resulting models were compared. It was shown the artificial neural network results have substantially better predictive capability than the Multiple linear regression, yields a regression model with improved predictive power, we have established a relationship between several descriptors and the antiradical activity in satisfactory manners. The good results obtained with the cross-validation CV, shows that the model proposed in this paper is able to predict activity with a great performance, and that the selected descriptors are pertinent. The most significant descriptors associated with the antiradical activity of these compounds were D, E_{LUMO}, W, NRB, PSA, and η.Thus, grace to QSAR studies, especially with the ANN that has allowed us to improve the correlation between the observed biological activity and that predicted, we can enjoy the performance of the predictive power of this model to explore and propose new molecules could be active.

Acknowledgements-The authors are grateful to the "Association Marocaine des Chimistes Théoriciens (AMCT)" for help on computation software.

References

[1] C. Rice Evans, *Antioxidant Food Supplements in Human Health*.,(1999) 239-253.
[2] S. V. Jovanic, S. Steenken, M. Tosic , B. Marjanovitz, M. G. Simic, *J. Am. Chem. Soc*., 11 (1994), 4846-4851.
[3] W. Wang, P.C. VanAlstyne, K. A. Irons, S. Chen, J.W. Stewart, D. F. Birt, *Nutr. Cancer*., (2004) 48-106.
[4] J. S. Choi, Y. J. Choi, S. H. Park, Y. H. J. Kang, *Nutr*., (2004) 134-1013.
[5] W. Bors, W. Heller, C. Michel, M. Saran, *Method enzymol*., 186 (1990) 343-355.
[6] W. Bors, C. Michel, K. Stettmaier, *Biofactors.,6 (*1997) 399-402.
[7] M. Leopoldini, T. Marino, N. Russo, *The Journal of Physical Chemistry A*., 108(22) 2004 4916-4922.
[8] K. Hassanzadeh, K. Akhtari, H. Hassanzadeh, *Food Chemistry*.,164 (2014) 251-258
[9] R. Praveena, K. Sadasivam, V. Deepha, R. Sivakumar, *J. Mol. Struct*., 1061 (2014) 114-123.
[10] S. Burda, W. Oleszek, *J. Agric. Food Chem.*,49 (2001) 2774.
[11] R. Pellegrini, N. Proteggente, A. Pannala, A. Yang, C. Rice-Evans, *Free Rad. Biol. Med*., 26(1999)1231-1238

[12] M. J. Frish, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, et al., *Gaussian 03 Revision C.02, Gaussian, Inc., Wallingford C*, (2004).
[13] A. D. Becke, *Density-functional thermochemistry. III, J. Chem. Phys*., 98 (1993) 5648-5652.
[14] Advanced Chemistry Development Inc., Toronto, Canada (2015) http://www.acdlabs.com.
[15] Chem bio Office, PerkinElmer Informatics., (2010) http://www.cambridgesoft.com.
[16] I. T. Jolliffe, Principal Component Analysis, New-York, *NY: Springer, 2ème edition*, (2002).
[17] K. Roy, S. Kar, and P. Ambure, *Chemometrics and Intelligent Laboratory Systems.*, 145 (2015) 22-29.
[18] XLSTAT 2009 Add - in software, XLSTAT Company, www.xlstat.com.
[19] H. Demuth, M. Hugan, M. Beal, *Neural Network Toolbox: For Use with MATLAB, User Guid's, Version 9.*, (2011).
[20] MATLAB (R2016a), and statistics Toolbox Release 2016a, The Math Works, Inc... Natick. Massachussetts., United States, (2016).
[21] T. D. Mark, T. Puzyn, J. Leszczynski, M. T. Cronin, *Recent Advances in QSAR Studies., Springer, edition* (2010).
[22] H. Martens, T. Næs, Multivariate calibration, John Wiley& Sons, Chichester, (1989).
[23] L. Eriksson, J. Jaworska, A. Worth, M. T. D. Cronin, R. M. Mc Dowell, and P. Gramatica, *Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification and Regression Based QSARs, Environmental Health Perspectives.*, 111 (2003) 1361-1375.
[24] H. Lafridi, H. Zaki, H. Zgou, M. Bouachrine, *Der Pharma Chemica.*, 10 (2018) 33-42.
[25] Y. Z. Zheng, G. Deng, Q. Liang, et al. *Scientific Reports*., 7 (2017) 7543.
[26] K. M. Holder and A. B. F da Silva, *Int. J. Quantum Chem.*, 95 (2003) 126.
[27] A. Ousaa, B. Elidrissi, M. Ghamali, S. Chtita, A. Aouidate, M. Bouachrine and T. Lakhlifi, *J. Mater. Environ. Sci*., 9 (2018) 256-265.