A Quantitative Structure–Activity Relationship (QSAR) Study of the Antioxidant Activity of Flavonoids

ARTICLE *in* QSAR & COMBINATORIAL SCIENCE · JULY 2005

Impact Factor: 1.55 · DOI: 10.1002/qsar.200430013

4 AUTHORS, INCLUDING:

Bakhtiyor Rasulev
North Dakota State University
76 PUBLICATIONS **815** CITATIONS

QSAR & Combinatorial Science QCS

A Quantitative Structure-Activity Relationship (QSAR) Study of the Antioxidant Activity of Flavonoids

Bakhtiyor F. Rasulev[a,b],* Nasrulla D. Abdullaev[b], Vladimir N. Syrov[b] and Jerzy Leszczynski[a]

[a] Computational Center for Molecular Structure and Interactions, Jackson State University, 1325, J. R. Lynch Street, P. O. Box 17910, Jackson, Mississippi 39217-0510 USA, E-mail: rasulev@ccmsi.us
[b] Institute of Chemistry of Plant Substances AS RUz, Kh. Abdullaev str., 77, Tashkent, 700170, Uzbekistan

Keywords: antioxidant activity, flavonoids, inhibition of lipids peroxidation, Genetic algorithm, QSAR, quantum-chemical descriptors

Received: December 30, 2004; Accepted: May 27, 2005

Abstract

A Quantitative Structure-Activity Relationship (QSAR) study has been carried out for 27 flavonoids belonging to four different groups (isoflavons, flavons, flavonols, flavanons) to correlate and predict the inhibition of lipids peroxidation effects (antioxidant activity). The genetic algorithm (GA) and multiple linear regression analysis (MLRA) were used to select the descriptors and to generate the correlation models that relate the structural features to the biological activities. The obtained equations consist of one to four descriptors calculated from the characteristics of the molecular structures with use of *DRAGON* software and quantum-chemical methods. A number of molecular descriptors was obtained from the density functional theory (*DFT*) B3LYP/6-31G(d, p) level optimized geometries (quantum-chemical descriptors). The results of the GA-MLRA analysis show that the position of the OH groups, the magnitude of dipole moment and the shape of the molecule play an important role in inhibition of lipids peroxidation by flavonoids. The significant QSAR models were obtained with r value of 0.935 and 0.933 for basic models. The q^2 (cross validation r^2) values and scrambling/randomization experiments also confirms the statistical significance of our models. These models are expected to be useful for screening of flavonoid antioxidants.

1 Introduction

Flavonoids occur in most plant species. Structural diversity of flavonoids allow them to exhibit antineoplastic, antihepatites, antibacterial, anti-inflammatory, antimutagenic, antiallergic, anti-thrombotic, antiviral, and vasodilatory activity [1–3]. The potent antioxidant activity of flavonoids, their ability to scavenge hydroxyl radicals, superoxide anions and lipid peroxy radicals, could be the most important function of flavonoids and underlies many of the above processes in the body [4]. Oxidative damage is implicated in most disease processes, and epidemiological, clinical, and laboratory research on flavonoids and other antioxidants suggest their possible applications in the prevention and treatment of a number of diseases. Such flavons and flavonols as Galangin and its derivatives, quercetin, morin, myricetin, oligomeric proanthocyanidins, and flavanons have been shown to inhibit the oxidation of lipids and can be utilized in preventative and treatment protocols for inflammatory conditions, cancer, asthma, liver disease, cardiovascular disease and macular degeneration [5, 6].

By trapping free radicals, flavonoids are capable of inhibiting the Lipids PerOxidation (LPO) induced by various factors [6]. Due to this feature the flavonoids are potential antioxidant preparates.

Search and design of new highly effective, low-toxic inhibitors of the LPO with a wide spectrum of applications is of both basic and commercial interest. Such inhibitors could be designed using flavonoids fragments. And good understanding of their structure and function is a prerequisite for a rational design.

Lipid Oxidation

Lipids in biological systems can undergo oxidation, leading to deterioration. In foods, these reactions can lead to rancidity, loss of nutritional value from the destruction of vitamins (A, D, and E) and essential fatty acids, and the possible formation of toxic compounds and colored products.

The important lipids involved in oxidation are the unsaturated fatty acid moieties such as oleic, linoleic, and linolenic. The rate of oxidation of these fatty acids increases with the degree of unsaturation. The overall mechanism of lipid oxidation consists of three phases: (1) initiation, the formation of free radicals; (2) propagation, the free-radical

chain reactions; and (3) termination, the formation of non-radical products.

Initiation of autoxidation occurs when hydrogen atom at α-methylene group in double bonds of unsaturated fatty is removed to form an alkyl radical (R·)

Initiation:

$$RH \rightarrow R\cdot + \cdot H$$

Propagation:

$$R + O_2 \rightarrow \cdot + ROO\cdot$$

$$ROO\cdot + RH \rightarrow R\cdot + ROOH$$

$$ROOH \rightarrow RO\cdot + HO\cdot$$

$$RO\cdot + RH \rightarrow ROH + R\cdot$$

Termination:

$$R\cdot + R\cdot \rightarrow RR$$

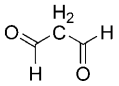
$$2\ RO\cdot \rightarrow ROOR$$

$$R\cdot + ROO\cdot \rightarrow ROOR$$

$$R + RO\cdot \rightarrow ROR$$

$$ROO\cdot + ROO\cdot \rightarrow ROOR + O_2$$

RH represents any unsaturated fatty acid; R·is a free radical formed by removing a labile hydrogen from a carbon atom adjacent to a double bond; and ROOH is a hydroperoxide, one of the major initial oxidation products that decompose to form compounds responsible for off-flavors and odors. Such secondary products include hexanal, pentanal, and malonaldehyde (malondialdehyde), Scheme 1.



Scheme 1. Malondialdehyde

Our experimental data is based on the measurement of the formation of the aforementioned product (malonaldehyde) which directly correlates with lipid oxidation.

Quantitative Structure-Activity Relationship (QSAR) [7–12] is a powerful method for the design of bioactive compounds and the prediction of corresponding activities that correlate with physical and chemical properties. Usually there are two major approaches to analyze QSAR data: i) the property (or activity) of a series of compounds is expressed as a multiple linear regression of descriptors and ii) the non-linear regression method represents the property (or activity) within an artificial neural network (ANN). ANN is an information-processing paradigm inspired by the densely interconnected, parallel structure concerning governing the way that the mammalian brain processes information. Also we can refer to approach (i) the several mixed methods, for example, multiple linear regression combined with variable selection genetic algorithms based on evolution theory.

In previous structure-relationship studies of flavonoids the authors have considered selected classes of flavonoids and have indicated the importance of the OH groups [21, 33–35]. For example, in paper [21] authors studied bond dissociation energies and ionization potentials for 6 substituted phenolic compounds and 20 other compounds related to phenolic antioxidants (flavonoids, aminophenols, Vitamin E and etc). They have shown that such parameter as bond dissociation energy correlates with antioxidant activity of phenolic compounds. In work [33] authors studied 13 polyphenolic antioxidants and have shown that single-electron oxidation and LogP correlates with cytotoxicity ($r^2 \approx 0.72-081$). More recent work concerning structure-radical scavenging activity of 29 flavonoids [34], where authors used only indicator descriptors and have shown that radical scavenger potential depends on number and position of hydroxyl groups in flavonoid skeleton.

However, till now there are no universal models for broad types of flavonoid compounds. In the published papers either a limited variations of the compounds and parameters or indicator descriptors [34] (or experimental parameters as descriptors [35]) are used. Additionally, most of works was performed with use in vitro activity for algae or just by measuring of the scavenging free radical ability at process of interaction with radical agent 1,1-diphenyl-2-picryl-hyrdazyl (DPPH·) [34] or by measuring of oxidation potentials [35], in which other factors that can occur in cells not taken into account.

In this study we have used inhibition of Lipid peroxidation, which was obtained by examination of the liver and blood of rats as an activity parameter. We considered four types of flavonoid compounds together with a wide spectrum of calculated theoretical descriptors, including quantum-chemical descriptors obtained from high-level *ab initio* calculations.

The study was based on a variable selection Genetic Algorithm (GA) [13, 14]. The GA algorithm could be a useful technique for searching the large probability space with a large number of descriptors for a small number of molecules. The purpose of this research was to determine predictive QSAR models [15–18] by analyzing a set of 27 molecules. If the models are reasonable, it is possible to predict the biological activity of non-tested molecules. Finally, the successful models of QSAR certainly decrease the number of compounds to be synthesized, by making it possible to select for the synthetic work only the most promising compounds.

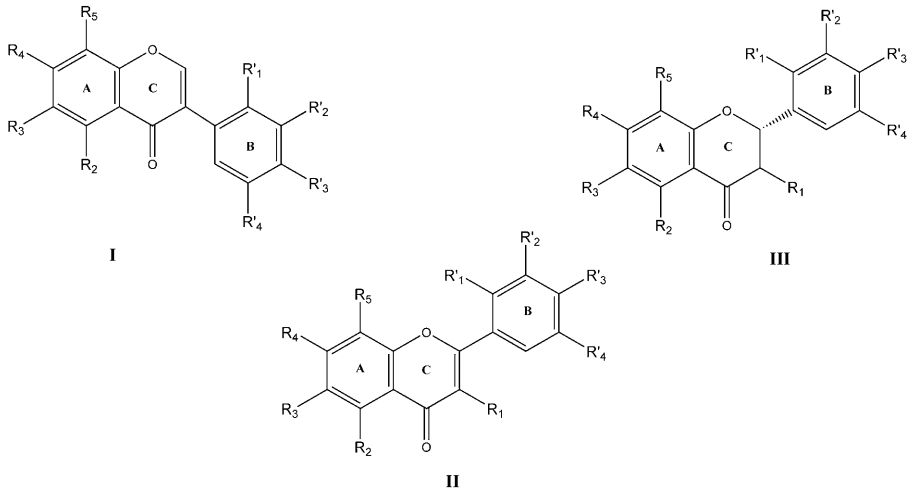


Figure 1. Molecular structures and numbering of different substituents attached to the core structure of flavonoids considered in the present study.

2 Objects

The structural components common to flavonoids include two benzene rings on either side of a 3-carbon ring. Multiple combinations of hydroxyl groups, sugars, oxygens, and methyl groups attached to these structures create the various classes of flavonoids: flavanols, flavanones, flavones, flavan-3-ols (catechins), anthocyanins, and isoflavones. A list of the studied compounds and 2D structures are shown in Table 1 and Figure 1, respectively.

An additional class of flavonoids, which is considered in this work, has recently been identified from hops. These flavonoids are substituted with prenyl groups [19], and to date Glabrol (compound 25) is one of the most potent species and is intermediate between Galangin, Myricetin and Morin.

3 Methods

In this study, quantitative structure-activity relationship (QSAR) analyses and quantum-chemical *DFT* calculations were performed on a series of flavonoids.

3.1 Computations. Data Construction

A series of 27 flavonoids was considered in this work (Table 1). Full geometry optimizations were performed by *DFT* method using the B3LYP functional and the 6-31G(d, p) basis set. The Gaussian 98 program was used in our study [20].

In our previous work the activity of the studied compounds was expressed in terms of the inhibition of lipids peroxidation on the homogenate of the liver and blood of rats [24]. Activities were expressed in percents of inhibition relative to control.

Generation and selection of conformations. In the first step of this study, a set of representative conformers was found that possess one low-energy conformation for each molecule.

Using the conformational search script in the HyperChem 6.01 software package we derived conformations with minimal energy [25]. The conformation search was carried out in the context of steric constraints (e.g., the distances between nonbonded atoms, ring-enclosure limits, torsional resolution) between the large parts of molecule.

The specially constrained script (for generating of conformations and energy minimizing), implemented in HyperChem 6.01, was used to minimize the energy of the initial structures of all selected compounds and to explore their conformational space by means of the AM1 semiempirical method [26]. Energy minimization was carried out using 200 steps of the steepest descent followed by conjugate gradient algorithms until the root mean square deviation (rmsd) of the changes between iterations was less than 0.01 Å.

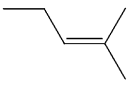
To successfully derive robust QSARs, the molecular conformation included in the model must be energetically reasonable, i.e. there must be a systematic way to screen and exclude high energy structures. Therefore, conformational analysis for each of the selected compounds was carried out using a systematic nested rotation procedure. The structures derived from the conformational searches were ordered according to the energy.

Finally, we selected the most stable conformations which were then optimized at the B3LYP/6-31G(d, p) level. The *DFT* method has shown a good correlation with experimental data concerning studies of the activity of phenolic antioxidants [21]. The authors concluded that the *DFT* method at the B3LYP/6-31G(d), B3LYP/6-311G(d) and B3LYP/6-311 + G(2d,2p) levels accurately predict the elec-

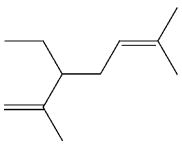
Table 1. Substitution pattern of the series of flavonoids examined for their antioxidant activity (Inhibition of Lipids peroxidation).

N	Compound	Activity, % I_{LPO}	R1	R2	R3	R4	R5	R1′	R2′	R3′	R4′
	I – Isoflavons										
1	Orobol	0.67		OH	H	OH	H	H	OH	OH	H
2	Formononetin	−0.23		H	H	OH	H	H	H	OCH_3	H
3	Genistin	−0.15		OH	H	O-β-D-Glc	H	H	H	OH	H
4	Ononin	−0.22		H	H	O-β-D-Glc	H	H	H	OCH_3	H
5	Biochanin A	−0.09		OH	H	H	H	H	H	OCH_3	H
	II – Flavons, Flavonols										
6	Luteolin	0.64	H	OH	H	OH	H	H	OH	OH	H
7	Chrysoeriol	0.11	H	OH	H	OH	H	H	OH	OCH_3	H
8	Cynaroside	0.38	H	OH	H	O-β-D-Glc	H	H	OH	OH	H
9	Thermopsoside	−0.07	H	OH	H	O-β-D-Glc	H	H	OH	OCH_3	H
10	Hispidulin	0.04	H	OH	OCH_3	OH	H	H	H	OH	H
11	Galangin	0.86	OH	OH	H	OH	H	H	H	H	H
12	Quercetin	0.78	OH	OH	H	OH	H	H	OH	OH	H
13	Rutin	0.24	O-GR[1]	OH	H	OH	H	H	OH	OH	H
14	Morin	0.81	OH	OH	H	OH	H	OH	H	OH	H
15	Myricetin	0.83	OH	OH	H	OH	H	H	OH	OH	OH
16	Myricetin-3-glucoside	0.40	O-β-D-Glc	OH	H	OH	H	H	OH	OH	OH
17	Isorhamnetin	0.81	OH	OH	H	OH	H	H	OCH_3	OH	H
18	Limocitrin	0.70	OH	OH	H	OH	OCH_3	H	OCH_3	OH	H
19	Haplogenin-7-glucosid	0.27	OH	OH	H	O-β-D-Glc	OH	H	OCH_3	OH	H
20	Haplosid	−0.05	OH	OH	H	O-β-DAcGR[2]	OCH_3	H	OCH_3	OH	H
	III – Flavanons										
21	Pinocembrin	0.01	H	OH	H	OH	H	H	H	H	H
22	Glabranin	0.01	H	OH	H	OH	Pr[3]	H	H	H	H
23	Isoglabranin	0.11	H	OH	Pr[3]	OH	H	H	H	H	H
24	Isobavachin	0.03	H	H	H	OH	Pr[3]	H	H	OH	H
25	Glabrol	0.81	H	H	H	OH	Pr[3]	H	Pr[3]	OH	H
26	Vexibinol	0.34	H	OH	H	OH	Lav[4]	OH	H	OH	H
27	Lehmannine	0.73	H	H	H	OH	Lav[4]	OH	H	OH	H

[1] O-GR = O-β-D-Glc-α-L-Rha, where Glc = glucose, Rha = rhamnose
[2] O-β-DAcGR = O-β-D-(6-OAc)-Glc-α-L-Rha
[3] Pr =



[4] Lav =



tronic parameters, including bond dissociation enthalpy and ionization potential. In our recent studies of heterocyclic systems with the hydroxyl and keto groups, the B3LYP/6-31G(d) and B3LYP/6-31G(d, p) levels have also shown good prediction ability [22, 23].

Quantum-chemical molecular descriptors were computed for the corresponding geometry optimized structures and molecular parameters calculated at *DFT* level.

3.2 Molecular descriptors

It is necessary to construct numerical descriptors of a set of molecules in order to build QSAR models. A descriptor can represent a quantitative property that depends on the structure of the molecule. An advantage of the exclusive use of theoretical descriptors is that they are free of the uncertainty of experimental measurements and can be calculated for compounds not yet synthesized.

Molecular descriptors (i.e., features) were computed mainly using the *DRAGON* software of *Todeschini* and *Consonni* [27] for a set of 27 flavonoid compounds with antioxidant activity. The input files for descriptor calculation, containing information on atom and bond types, connectivity, partial charges and atomic spatial coordinates, relative to the minimum energy conformation of the molecule, were obtained after full geometry optimization by

Table 2. List of quantum chemical parameters (descriptors) obtained from low-energy conformations.

N	Compound	Quantum-chemical descriptors (*DFT* B3LYP/6-31G(d, p))				
		E_{HOMO}, eV	E_{LUMO}, eV	H – L gap	μ, (D)	E_{total}, Hartree
1	Orobol	– 5.3715	– 1.0699	4.3016	2.22	– 1028.96
2	Formononetin	– 5.5930	– 1.3363	4.2567	4.79	– 917.83
3	Genistin	– 5.5810	– 1.1891	4.3920	4.65	– 1564.50
4	Ononin	– 5.6287	– 1.4326	4.1960	3.31	– 1528.59
5	Biochanin A	– 5.4850	– 1.0666	4.4184	4.45	– 993.04
6	Luteolin	– 5.7737	– 1.4193	4.3544	2.99	– 1028.96
7	Chrysoeriol	– 5.6689	– 1.3648	4.3041	7.18	– 1068.27
8	Cynaroside	– 5.7353	– 1.4710	4.2643	6.16	– 1639.72
9	Thermopsoside	– 5.6553	– 1.4168	4.2385	6.39	– 1679.03
10	Hispidulin	– 5.7862	– 1.3809	4.4053	7.37	– 1068.26
11	Galangin	– 5.5666	– 1.6380	3.9286	4.39	– 953.75
12	Quercetin	– 5.2836	– 1.5344	3.7493	2.41	– 1104.19
13	Rutin	– 5.6907	– 1.5368	4.1539	7.59	– 2250.48
14	Morin	– 5.3405	– 1.5983	3.7422	4.20	– 1104.19
15	Myricetin	– 5.2722	– 1.5281	3.7441	1.70	– 1179.41
16	Myricetin-3-glucoside	– 5.7593	– 1.8043	3.9550	9.06	– 1790.16
17	Isorhamnetin	– 5.2153	– 1.4930	3.7223	1.91	– 1143.50
18	Limocitrin	– 5.2205	– 1.5175	3.7030	6.18	– 1258.02
19	Haplogenin-7-glucosid	– 5.1326	– 1.5412	3.5914	8.73	– 1829.47
20	Haplosid	– 5.6099	– 1.3126	4.2973	9.50	– 2481.71
21	Pinocembrin	– 6.1084	– 1.0644	5.0439	3.67	– 879.73
22	Glabranin	– 5.8257	– 0.9330	4.8926	5.08	– 1075.09
23	Isoglabranin	– 5.7682	– 0.8976	4.8706	3.57	– 1075.09
24	Isobavachin	– 5.9554	– 1.1044	4.8510	5.53	– 1075.10
25	Glabrol	– 5.8403	– 1.0557	4.7846	5.03	– 1270.45
26	Vexibinol	– 5.9051	– 1.1102	4.7949	5.70	– 1420.88
27	Lehmannine	– 5.8888	– 1.3733	4.5155	3.70	– 1345.64

the *DFT* quantum-chemical method. A total of 1050 molecular descriptors of different kinds (2D and 3D) were used to describe chemical diversity of the compounds. The descriptor typology is: a) constitutional (atom and group fragments), b) functional groups, c) atom centered fragments, d) empirical, e) topological, f) walk counts, g) various autocorrelations from the molecular graph, h) Randic molecular profiles from the geometry matrix, i) geometrical, j) WHIMs, k) GETAWAYs descriptors and various indicator descriptors. The meaning of these molecular descriptors and the calculation procedure are summarized in the manual to the *DRAGON* software and are explained in detail, with related literature references, in the *Handbook of Molecular Descriptors* by *Todeschini* and *Consonni* [28].

In addition, due to the importance of quantum-chemical descriptors for QSAR, several parameters calculated by the *DFT* B3LYP/6-31G(d, p) method were added as molecular descriptors. These additional descriptors include energies of highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO), HOMO-LUMO gap, dipole moment (μ), charges on atoms (q_i) and total energy (E) (see Table 2).

3.3 Chemometric Methods

A reliable equation for structure and activity relationship should possess a high correlation coefficient (R), low standard deviation (s), and least variables. To this end, feature selection was made objectively to eliminate those descriptors that provide minimal or redundant information.

The correlation between biological activity and structural properties was obtained by using the variable selection GA and MLRA methods. GA can automatically select the optimum number of descriptors in regression analysis as well as construct Multiple Linear Regression models through the use of linear, higher order polynomials, splines and gaussians. We applied the Genetic Algorithm to select from all the calculated descriptors only the best combinations of those and the most relevant for obtaining models with highest predictive power for antioxidant activity. Genetic Algorithms have been employed in recent years as a powerful tool to optimize many problems in drug design [29, 30]. As mentioned before, this approach allows selection of the models with the following characteristics: high correlation coefficient R, low standard deviation S and the least number of descriptors involved. The GA technique began with a population of 100 random models and 1000 iterations to evolution, mutation – 35%, fitness/scoring function – correlation coefficient (r). For GA analysis and

for deriving the QSARs, we have used the BuildQSAR program [32]. A final set of QSARs was validated by applying the "leave-one-out" technique with its predicting ability being evaluated and confirmed by cross validation coefficient Q^2 based on predictive error sum of squares (SPRESS).

For strict validation purposes the scrambling/randomization tests was carried out by repetitive randomization of the response values (antioxidant activity). In each cycle of the test the response vector is randomly rearranged and R^2 and leave-one-out cross-validation squared Q^2 coefficient is recorded for each cycle. In each test the process was repeated 100 times.

Table 3. Correlation matrix for the descriptors selected by GA-MLRA analysis and used in models.

	PJI3	**μ, D**	$\mathbf{I_{Glc}}$	$\mathbf{I_{OH}}$
PJI3	1.000	0.074	0.001	0.245
μ, D		1.000	0.315	0.000
$\mathbf{I_{Glc}}$			1.000	0.054
$\mathbf{I_{OH}}$				1.000

PJI3 – Petijean shape index
μ – Dipole moment
$\mathbf{I_{Glc}}$ – Indicator descriptor, number of Glycoside-like fragments
$\mathbf{I_{OH}}$ – Indicator descriptor, number of OH groups at positions C-3 (C-ring) and the 3′,4′ position in the B-ring

4 Results and Discussion

A series of flavonoids with different substitution patterns was tested to define the molecular features required for the high antioxidant activity of these compounds. A set of 27 flavonoid derivatives was taken from the work of one of the authors (*V. N. S)* [24]. The structures of the flavonoids used in the analysis, as well as the experimental values of the related inhibition of lipids peroxidation, are shown in Table 1. The flavonoids studied exhibit vastly different antioxidant activities. As evident from the table, among explored compounds both antioxidants and prooxidants are present, where prooxidants have shown some influence to increasing of lipid peroxidation.

Finally, we generated total of 100 QSAR equations that consist of one to four descriptors among the QSAR random models. As a rule of thumb, for reliable results the data set should be approximately 5 times larger than the number of selected descriptors. For variable selection a maximum number of four variables were allowed to be selected to reduce the risk of chance correlation. The results of the best QSAR model using four descriptors are given in Equations (1) and (2), and a QSAR model using three descriptors is provided in Eqn. (5). The regression models are all significant at the F-value using the Fisher F-statistics. As a result of the highest F-value and the highest R, an optimum fit was found for one to four descriptors. The values in parenthesis following the regression coefficients are the 95% confidence limits. These equations produce the best description for the activity of the flavonoids. While constructing the models, great care was taken in order to avoid inclusion of highly collinear descriptors. All collinear descriptors were eliminated from further consideration. The correlation matrix for the descriptors used in this study is given in Table 3. The Table includes only those descriptors that have comprised the models selected by the variable selection Genetic Algorithm method. In final models we left only those descriptors that have cross-correlation coefficients less than 0.35.

As mentioned above, together with the topological and quantum-chemical descriptors, we employed indicator variables to mark the presence (I = 1) or absence (I = 0) of phenolic groups at any position as well as the presence or absence of O-Glucose groups and the corresponding analogs (for example, O-β-D-(6-OAc)-Glc-α-L-Rha) or OCH_3 groups at the position of the R_3' substituent in the B-ring. A GA-MLRA analysis confirms that the OH substitution pattern is responsible for the variation of the antioxidant activity ($\mathbf{I_{LPO}}$) of the studied flavonoids.

One of the obtained models (Eqn.1) consists of four descriptors: a topological descriptor (**PJI3**), an electronic descriptor (dipole moment **μ**), and two indicator descriptors ($\mathbf{I_{OH}}$ and $\mathbf{I_{Glc}}$).

$$\textbf{Activity } (\mathbf{I_{LPO}}) = 0.526(\pm 0.074)\mathbf{I_{OH}} + 0.287(\pm 0.352)\mathbf{PJI3} - 0.029(\pm 0.018)\mu - 0.262(\pm 0.078)\mathbf{I_{Glc}} + + 0.028(\pm 0.313), \quad (1)$$

$N = 27$, $R = 0.935$, $Q^2 = 0.808$, $s = 0.147$, $F = 38.16$, $SPRESS = 0.181$

In Eqn. (1) the indicator descriptor I_{OH} denotes the presence of the 3′,4′-dihydroxy structure at the B-ring or the presence of the OH group at the C-3 atom. Indicator descriptor I_{Glc} denotes the presence of O-glucose groups and/or the presence of OCH_3 groups at the position of the R_3' substituent in the B-ring.

Congruity between the observed and the GA-MLRA-predicted I_{LPO} for this set is shown in Table 4. The residual of those values is 0.002 – 0.308, and the mean residual is 0.100.

By investigation of each descriptor in Eqn. (1) we conclude that the descriptor PJI3 quite weakly describes the properties of the prooxidant compounds 2, 3, 4, 5, 9 and 20. Based on this conclusion, we have built a model which excludes prooxidant compounds:

$$\textbf{Activity } (\mathbf{I_{LPO}}) = 0.429(\pm 0.184)\mathbf{I_{OH}} + 0.761(\pm 0.646)\mathbf{PJI3} - 0.035(\pm 0.015)\mu - 0.128(\pm 0.106)\mathbf{I_{Glc}} + 0.282(\pm 0.299), \quad (2)$$

$N = 21$, $R = 0.938$, $Q^2 = 0.846$, $s = 0.125$, $F = 39.52$, $SPRESS = 0.182$

From the obtained model one can see that the importance of the PJI3 descriptor has increased significantly, while the coefficients of the other descriptors have changed slightly. The residual of the observed and predicted values in this case is 0.005 – 0.300, and the mean residual is 0.081.

The PJI3 descriptor is the Petijean shape index. It is a topological anisometric descriptor [28], also called graph-theoretical shape coefficient I_2, and is defined as:

$$I_2 = \frac{D - R}{R} 0 \leq I_2 \leq 1,$$

where R and D are the *topological radius* and the *topological diameter*, respectively, obtained from the *distance matrix* representing the considered *molecular graph*.

If one considers topological descriptor PJI3 alone and derives correlation between PJI3 and the observed activity for the two cases - to the considered set of all compounds and to the set which excludes prooxidant compounds then the correlation coefficient changes significantly from r = 0.418 (poor correlation) to r = 0.776 (good correlation):

$$\textbf{Activity } (\mathbf{I_{LPO}}) = 2.207(\pm 0.857)\mathbf{PJI3} - 1.389(\pm 0.723), \quad (3)$$

N = 21, R = 0.776, Q^2 = 0.530, s = 0.209, F = 28.85, SPRESS = 0.367

It can be noted that for the most active compounds the PJ13 descriptor has a high value (PJI3 ≥ 0.9).

A study of indicator descriptor $\mathbf{I_{OH}}$ also showed its high importance in the activity of flavonoids. The simultaneous presence of OH group at the C-3 and 3′,4′-dihydroxyl group within the compound defines the main contribution to the antioxidant activity. The following equation summaries this conclusion:

$$\textbf{Activity } (\mathbf{I_{LPO}}) = 0.617(\pm 0.181)\mathbf{I_{OH}} - 0.041(\pm 0.139), \quad (4)$$

N = 27, R = 0.813, Q^2 = 0.612, s = 0.226, F = 48.80, SPRESS = 0.241

Analyzing the Eqn. 4 one can conclude that all active compounds include the OH group at C-3 and the 3′,4′ positions in the B-ring. These OH groups, most likely, participate in electron delocalization, which allow to improve the stability of the flavonoids phenoxyl radicals.

From all calculated quantum-chemical descriptors only the dipole moment have been selected by GA-MLR analysis.

The magnitude of the dipole moments provides an essential contribution to activity, i.e., it inversely correlates with the activity and the size of the molecule. Actually, bulky groups dramatically increase the size of the molecule and the values of dipole moment; the OCH_3 group also slightly increases the dipole moment. A designed model, where only three descriptors are taken in account,

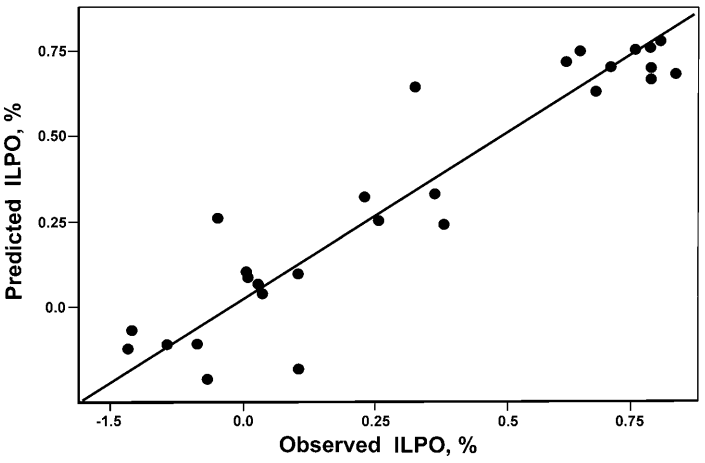


Figure 2. A correlation plot of observed and GA-MLR-predicted $\mathbf{I_{LPO}}$ values for a set of 27 flavonoids, Eqn.5 (linear correlation).

$\mathbf{I_{OH}}$, the dipole moment, and $\mathbf{I_{Glc}}$ for all 27 compounds, is shown in Eqn. (5), Figure 2:

$$\textbf{Activity } (\mathbf{I_{LPO}}) = 0.561(\pm 0.059)\mathbf{I_{OH}} - 0.036(\pm 0.016)\mu - 0.239(\pm 0.072)\mathbf{I_{Glc}} + 0.273(\pm 0.081), (5)$$

N = 27, R = 0.933, Q^2 = 0.821, s = 0.146, F = 51.42, SPRESS = 0.171

Eqn. (5) provides a good correlation for all sets of considered compounds, taking into account the effects of the prooxidant compounds. It is obviously that the presence of the O-glucose and OCH_3 groups contributes to the prooxidant properties of the flavonoids.

As can be seen in Figure 3, the most active compounds have the dipole moments of 5 D and less. It is known that at the beginning of the interaction between two molecules the electrostatic field has the largest effect. So, we can assume that during flavonoids action there is a combination of several processes, where in addition to the phenolic groups (radical formation) also the characteristics that include the defined value of dipole moment are required. The value of dipole moment defines orientation of flavonoid and interaction rate and thereby play important role in driving of interaction. The data analysis has shown that with an increase in the dipole moment, the inhibition constant is decreased. It indicates that a good inhibitor of lipid peroxidation should have a small dipole moment (no more than 5 – 6 D).

These predictions contain a small error value which represents the prediction ability of the basic QSAR equations, Eqn. (1) and Eqn. (2).

On the basis of these results it appears that the most effective inhibitors of lipid peroxidation are flavonoids with the 3′,4′-substitution pattern on the B-ring (hydroxyl and/

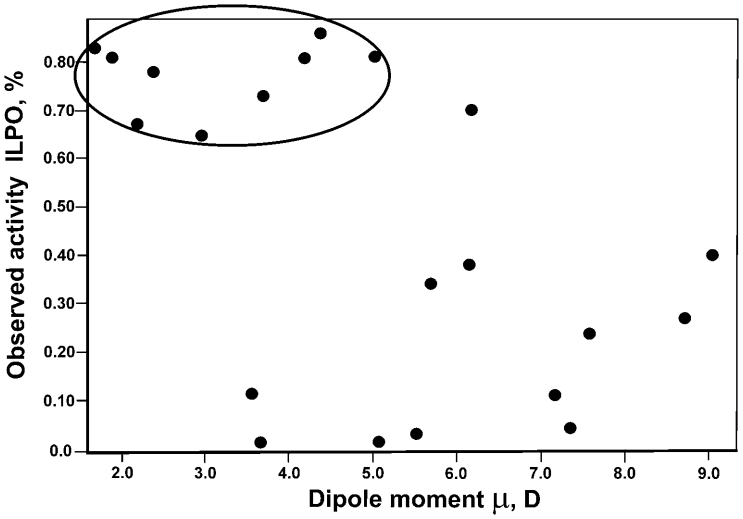


Figure 3. A plot of the dipole moments values and inhibition of lipids peroxidation $\mathbf{I_{LPO}}$ for 21 compounds.

Table 4. Statistical parameters (R, R^2 and Q^2) for all obtained models and results of scrambling tests for Models (1) and (5).

Model	Number of compounds	Training set			Scrambling test	
		R	R^2	Q^2	R^2 (min-max)	Q^2(for R^2max)
Model 1	27	0.935	0.874	0.808	0.025 – 0.443	0.191
Model 2	21	0.938	0.880	0.846		
Model 3	21	0.776	0.602	0.530		
Model 4	27	0.813	0.661	0.612		
Model 5	27	0.933	0.871	0.821	0.013 – 0.335	0.066

or prenyl group) and/or the hydroxyl group at the C-3 position. Apparently, the presence of an o-dihydroxy structure on the B-ring confers a higher degree of stability on the flavonoid phenoxyl radicals by participating in electron delocalization and is an important feature for the antioxidant activity. The high antioxidant activities of the Glabrol and Lehmanine flavanones are probably related to the favorable 3′,4′-substitution pattern (hydroxyl and/or prenyl) on the B-ring. The lowest inhibition of lipids peroxidation of flavonoids lack both the C-3 hydroxyl group and the 3′,4′-dihydroxy (or prenyl) occupied B-ring.

One can conclude that the Eqn. (5) is useful for the description not only antioxidant, but also prooxidant compounds (i.e., compounds with "bulky" glucose-like groups and OCH_3 groups). In another case, when there are flavonoids with only hydroxyl groups, the Eqn. (2) can be used.

The models obtained in our study show not only statistical significance, but also good predictive ability. The estimated predictive ability ("leave-one-out" Q^2 cross-validation coefficient) of the models has been within 0.530 – 0.846. Because of relatively small set of compounds we can not apply internal "leave-many-out" technique for estimation the predictive ability of models, therefore we used the "leave-one-out" technique with cross-correlation coefficient Q^2. The models with Q^2 less than 0.8 are often considered not particularly good. A low Q^2 value can be due to two reasons: (1) variability (uncertainty) in the dependent variable measures of the training set and/or (2) a limited and/or poor selection of the independent variables for constructing the model. In the case of our training set, and inhibition of lipid peroxidation in general, there is a considerable variability in the inhibition of lipid peroxidation measures that are the dependent variables. Thus, the low Q^2 values from the inhibition of lipid peroxidation arise predominantly from "noise" in the inhibition of lipid peroxidation measures, and not necessarily from a poor selection of descriptors.

For the avoiding of overfitting results we carried out scrambling (randomization) tests. The experimental activity of compounds was randomly mixed between all of them. The models were built with mixed activities and was validated using R^2 and Q^2 (leave-one-out) technique. As it known, after mixing of activities data, the model should lose its predictive ability because the relationship between the structure and mutagenicity is broken. The test showed breakdown of predictive power of the model. The test was realized for two basic models – Model (1) and Model (5). Scrambling tests results for two models and correlation and validity results for all models are presented in Table 4.

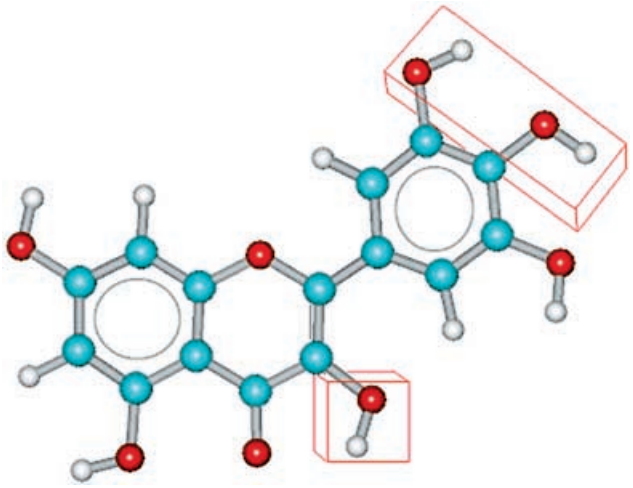


Figure 4. Structural features of flavonoids with a high antioxidant activity ($\mathbf{I_{LPO}}$).

Based on the described above results the significant molecular descriptors related to the compounds with antioxidant activity are: molecular shape index descriptor (Petijean shape indices), dipole moment, indicator variable of the presence of hydroxyl groups (OH) and the presence of O-Glc groups in relevant positions. These variables lead to a molecular level explanation of the geometrical and electronic molecular property contributions to the potency of lipid peroxidation inhibition.

The proposed models can be used to estimate the antioxidant activities of new substituted phenolic compounds or flavonoid derivatives. In addition, a good distinctive feature of all models derived is their ability to be easily explained and reasonably predict the antioxidant activity using only calculated parameters.

5 Conclusions

We have shown in this work that a set of specific parameters define the antioxidant activity of flavonoids. One of the main parameters is the indicator variable which embraces the OH groups of the flavonoid core. Other parameters are quantum-chemical descriptor and topological descriptor that makes it possible to build reliable QSARs. The developed structure-antioxidant activity relationships (given by Eqn.2 and Eqn.5) indicate that highly active flavonoids possess a OH group at C-3 atom and a 3′,4′-dihydroxy fragment at the B-ring in addition to the specific topological shape of the molecule. Thus these flavonoids should also possess a low dipole moment and a lack of O-glucose groups. Among the electronic molecular properties, the dipole moment is the most important for the interaction between a drug and the target molecule. The value of dipole moment defines orientation of flavonoid and interaction rate that are important in driving of interaction. Data analysis indicates that a good inhibitor of lipid peroxidation should have a small dipole moment (no more then 5 – 6 D).

Therefore the specific topological shape of the molecule and the availability of the aforementioned conditions lead to high activity of the considered class of compounds. The present analysis of flavonoid QSARs offers insight into their possible mechanisms of action and could be used in the design of new flavonoid-based drugs for the treatment of free radical-mediated disease conditions.

The significance of the obtained models has been confirmed by validation and scrambling/randomization tests.

Acknowledgements

This work was supported in part by National Science Foundation through RISE grant No. HRD-0401730, National Institute of Health SCORE grant No.3S06 GM008047-31S1 and by grant No. 1804/02 (Foundation of Support for Fundamental Research, Academy of Science, Uzbekistan).

References

[1] O. P. Agarval, *Agents and Actions* **1982**, *12*, 298.
[2] M. K. Church, *Drugs of Today* **1978**, *14*, 281.
[3] M. M. Iwy, *Experintia* **1985**, *25*, 3 – 4, 321.
[4] T. Ozawa, *Physiol. Rev.* **1997**, *77*, 425.
[5] A. L. Miller, *Alt. Med. Rev.* **1996**, *1*(2), 103.
[6] N. C. Cook, S. Samman, *J. Nutr. Biochem*. **1996**, *7*, 2.
[7] C. Hansch , T. Fujita, *J. Am. Chem. Soc*. **1964**, *86*, 1616.
[8] T. Fujita, *QSAR and Drug Design: New Developments and Applications*, Elsevier: Amsterdam **1995.**
[9] D. L. Lee, P. A. Kollman, F. J. Marsh, M. E. Wolff, *J. Med. Chem*. **1977**, *20*, 1139.
[10] H. Kubinyi, *3D QSAR in Drug Design: Theory, Methods and Applications*; ESCOM: Leiden **1993.**
[11] Y. X. Zhou, L. Xu, Y. P. Wu, B. L. Liu, *J. Chem. Intell. Lab. Sys.* **1999**, *45*, 95.
[12] T. J. Hou, J. M. Wang, N. Liao, X. J. Xu, J. Chem. Inf. Comput Sci. **1999**, *39*, 775.
[13] R. Leardi, R. Boggia, M. Terrile, *J. Chemomertrics* **1992**, *6*, 267.
[14] R. David, A. J. Hofinger, *J. Chem. Inf. Comput. Sci*. **1994**, *34*, 854.
[15] M. Randic, *J. Am. Chem. Soc*. **1975**, *97*, 6609.
[16] B. W. Clare, *J. Med. Chem*. **1998**, *41*, 3845.
[17] K. Baumann, *J. Anal. Chem*. **1999**, *18*, 36.
[18] R. M. Sweet, D. Eisenberg, *J. Mol. Biol*. **1983**, *171*, 479.
[19] M. Kitaoka, H. Kadokawa, M. Sugano, K. Ichikawa, M. Taki, S. Takaishi, Y. Iijima, S. Tsutsumi, M. Boriboon, T. Akiyama, *Plant Medica* **1998**, *64*, 511 – 515.
[20] Gaussian 98, Revision A.11.1, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Ada-

mo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, P. Salvador, J. J. Dannenberg, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, Gaussian, Inc., Pittsburgh PA **2001.**

[21] J. S. Wright, E. R. Johnson, G. A. DiLabio, *J. Am. Chem. Soc.* **2001**, *123*, 1173.

[22] M. J. Huang, J. Leszczynski, *J. Comp. – Aided Mol. Design* **2001**, *15*, 323.

[23] D. Venkateshwarlu, J. Leszczynski, *J. Comp. – Aided Mol. Design* **1998**, *12*, 373.

[24] Z. A. Khushbaktova, S. M. Yusupova, M. V. Zamaraeva, V. N. Syrov et al., *Chem. Nat. Comp.* **1996**, *32*, 338 – 343.

[25] HyperChem 6.01, HyperCube Inc. **2000.**

[26] M. J. S. Dewar, E. F. Zoebisch, E. F. Healy, J. J. P. Stewart, *J. Am. Chem. Soc.* **1985**, *107*, 3902.

[27] R. Todeschini, V. Consonni, DRAGON software for the Calculation of Molecular Descriptors, web version 3.0 for Windows **2003.**

[28] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors, Wiley-VCH, Weinheim and New York **2000.**

[29] L. Davis, Handbook of Genetic Algorithms, Van Nostrand Reinhold, N. Y. (USA) **1991.**

[30] J. Devillers, Genetic Algorithms in Molecular Modeling, Academic Press, Ltd., London **1996.**

[31] M. Petitjean, *J. Chem. Inf. and Comp. Sciences* **1992**, *32*(4), 331 – 337.

[32] D. B. de Oliveira, A. C. Gaudio, *Quant. Struct.-Act. Relat.* **2000**, *19*(6), 599 – 601.

[33] E. Sergediene, K. Jonsson, H. Szymusiak et al., *FEBS Letters* **1999**, *462*, 392.

[34] D. Amic, D. D. Amic, D. Beslo, N. Trinajstic, *Croatica Chemica Acta* **2003**, *76* (1), 55.

[35] B. Yang, A. Kotani, K. Arai, F. Kusu, *Analytical Sciences* **2001**, *17*, 599.

