

QSAR model based in the TOPS-MODE approach used to predict chromosomal aberrations in bioactive phenolic compounds

Estela Guardado Yordi ^{1,2*}, Maria João Matos ², Roxana Castro Pupo ¹, Lourdes Santana ²,
Eugenio Uriarte ² and Enrique Molina Pérez ^{3,4}

¹Universidad de Camagüey “Ignacio Agramonte Loynaz”, Facultad de Química, Departamento de Ciencia y Tecnología de los alimentos, Camagüey – Cuba, * estela.guardado@reduc.edu.cu

²Universidad de Santiago de Compostela, Facultad de Farmacia, Departamento de Química Orgánica, Santiago de Compostela – España

³Universidad de Camagüey “Ignacio Agramonte Loynaz”, Facultad de Química, Departamento de Química, Camagüey – Cuba

⁴Universidade do Estado do Amazonas, Centro de Estudos Superiores de Parintins, Parintins, Amazonas – Brasil

Abstract: The *in silico* characterization of bioactive substances which are constituents of functional foods or nutraceuticals is the methodology described in this communication. The aim of this work is to show the potential of the TOPS-MODE approach as chemical-informatics method to study the structure/clastogenic activity (chromosomal aberrations) and identify structural alerts related to genotoxicity. The results of QSAR studies were analyzed for several classes of phenolic compounds (flavonoids, phenolic acids and coumarins) for which was required the use of software STATISTIC and MODESLAB and a mathematical model encoding topological information of a substructural level. It was observed that the criteria for maximum clastogenicity are the methoxy and hydroxyl polisubstitutions (methoxy > hydroxyl) and the polarity of the substituents. This was observed for all the analyzed subclasses. The confirmation of these results is based on the percentage of good classification for the used external databases, the recognition of physical chemical descriptors of polarity (μ_1^{Pol} y μ_2^{Pol}) and in the calculation of the fragments contribution. It can be conclude that QSAR methods, and in particular the used topographical approach, may constitute a predictive tool for the design and evaluation of bioactive components of functional foods or nutraceuticals.

Keywords: phenolic acids; QSAR study; clastogenic activity; functional foods; nutraceuticals.

Introduction

The increasing advance of industrialization and the changing lifestyles have produced changes in the work environments and dietary habits. It is estimated that worldwide there are more than 70,000 chemicals and about 1,000 new substances annually. Exposure to these compounds, either in its development or production can sometimes have adverse effects on the health of workers. Also, their subsequent use as additives, new-type foods, nutraceuticals or drugs can potentially affect the consumers. These effects cannot be related to immediate and apparent injuries, but can take years to be manifested. Some of these substances may be genetically active, presenting the capacity of interacting with genetic material, causing DNA damage (1). This damage is commonly measured by the breakage of single or double chains and chromosomal aberrations (2,3). Chromosomal aberrations are easily observable structural changes in the metaphase of the cell cycle (clastogenic processes), which are caused by breakage of the DNA strands unrepaired or improperly repaired (4). This process is considered, together with the mutations, endpoints of oxidative damage to DNA (1).

There is a relationship between exposure to genotoxic substances (either occupational, accidental or due lifestyle) and the increased cancer risk. The exposure to potential genotoxic agents, both physical and chemical, can produce, depending on the type of damage induced on DNA, chromosomal abnormalities such as chromosomal aberrations, making these agents “clastogens” (4). Genotoxicity that can be caused by some bioactive compounds, such as phenolic compounds derived from plants and foods, is an area of current interest. Some phenolic compounds present in vegetables consumed by us, which have been studied for their antioxidant activity, also showed *in vitro* pro-oxidant and clastogenic activity (5).

The bioactive compounds are mainly used in the development and characterization of a “healthy food”. They are been giving rise to special interest after the development of new concepts such as functional foods, due to their claim to provide, through regular consumption, potential preventive effects as health protective or promoters. (6). Functional foods could be naturally occurred, but also can be modified by adding bioactive compounds, which must have been previously tested in different experimental systems. Due to the high structural variability that may be considered

for phenolic compounds in nature, structure-activity relationship studies based on chemoinformatic methods (QSAR/Quantitative Structure-Activity Relationships) can provide a low cost alternative for predicting possible chromosomal aberrations in response to exposure to such agents. Therefore, the aim of this work is to show the potential of the TOPS-MODE approach as chemoinformatic method to study the structure/clastogenic activity (chromosomal aberrations) and identify structural alerts related to genotoxicity. The objectives are: i) analyse the theoretical fundament of the QSAR methodologies, in particular the TOPS-MODE approach, ii) show evidences of virtual screening using this approach and a validated clastogenic model, which encodes sub-structural topological information from molecular descriptors based on Graph Theory.

Methods

The current study was focused on natural phenolic compounds present in foods. They formed a series for external prediction. The screening was carried out following a sequence shown in Figure 1.

Preparation of databases (DB) of chemical compounds presenting reported pro-oxidant activity. The search for compounds with pro-oxidant activity was performed on the basis of reports in electronic sources of scientific information, using the keywords: pro-oxidant (+ food, + chronic diseases, + oxidative damage), structure-activity relationship, functional food, healthy foods, antioxidant, oxidative stress and QSAR study. The selection of the compounds were carried out taking into account the following criteria: a) Inclusion: the presence of scientific reports, regardless of the method of determination of the activity and experimentation system; a chemical entity that was recognized by a CAS b) Exclusion: chemical entities as radicals, ions, salts, macromolecules and/or isomers.

Development of QSAR study using the TOPS-MODE approach. All the chemical compounds selected in the different DB (flavonoids, cinnamic acids and coumarins), were drawn in ChemDraw software version 10.0 (7), allowing the obtainment of the *Smiles* for each compound. Molecular descriptors were calculated (spectral moments) using software MODESLAB for Windows available on <http://www.modeslab.com>

The structural alerts were identified from the generation of new compounds congeners of each of the studied families, representing pro-oxidant activity structures capable of damaging the DNA. Calculations were made of the contributions of the fragments to a sample of compounds belonging to the identified alerts following the calculation procedure implemented in the TOPS-MODE "fragment contributions" implemented in MODESLAB.

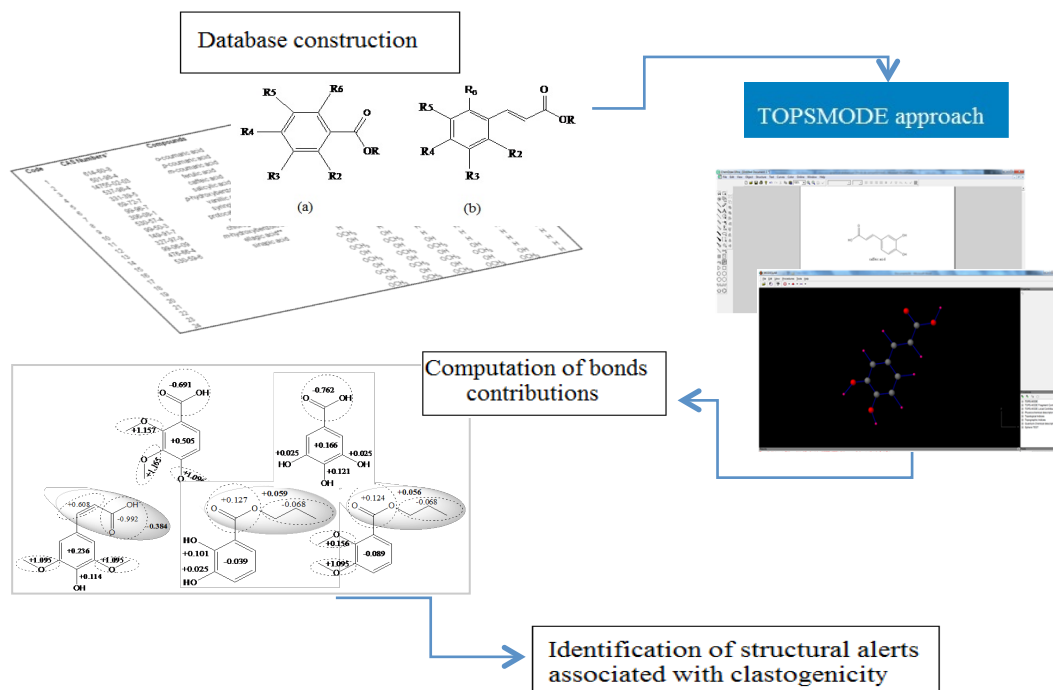


Figure 1. Methodology used for the QSAR studies.

Results and Discussion

QSAR studies and TOPS-MODE approach. QSAR methods are based on the physical, physicochemical, chemical and biological properties of the organic compounds, depending ultimately from their molecular structure. Its own philosophy makes possible to predict activities of new bioactive candidates, reducing the cost of experimental techniques (7). These methods can be applied to different sciences, among them food science, whenever it is important to reach one of the two main objectives of these studies: a) provide a way to estimate the studied activity/property of new compounds with a acceptable accuracy and b) obtain a structural interpretation in terms of the studied activity.

Generally, there are three basic components, which characterize these studies: i) the descriptors of molecular structure; ii) the property or biological activity and iii) the statistical technique used to establish the relationship.

TOPS-MODE is a molecular design approach with a graph-theoretical base. As cutting phenomenological needed to reach quantitative relationships using sample data that are processed statistically. This makes a general method, which requires no knowledge of the mechanisms involved in a process to describe it. This approach operates on the adjacency matrix between links in the molecule, excluding the hydrogen atoms. One of the topological descriptors used are so-called spectral moments (μ_k) of the matrix which were proposed by Estrada et al. (9). Some ideas in spectral moments have been generalized and extended to biomolecules by Estrada. The theoretical background about the spectral moments of bond adjacency matrix, has been described in many papers (9).

The spectral moment of order k is defined as the trace of the k -th power of the matrix E and its symbol is μ_k . Also, the trace is defined as the sum of the values of the main diagonal. One advantage is that μ_k can be expressed as linear combination of the number of occurrences certain fragments (sub-graphs). MODESLAB facilitates the calculation of global spectral moments selecting any weight (weight) for different bonds in the molecule, while local spectral moments for single bonds or a defined fragment in the molecule can be estimated in the same way after select the link/excerpt. The results are in file format, ready to be manipulated by statistical packages such as STATISTICA (10).

In the current work, it was selected the model of structure-activity clastogenic relationship proposed by Estrada et al. (11), which has been internally and externally validated (11-13) (Equation 1).

$$\begin{aligned}
 AC = & 0.0091[\Omega(\mu_1^{PS})] - 1.5520 \times 10^{-4}[\Omega(\mu_5^{vdW})] + 0.148[\Omega(\mu_4^{Ch})] - 0.0021[\Omega(\mu_2^{PS})] + \\
 & + 2.6261 \times 10^{-4}[\Omega(\mu_3^{PS})] - 3.8422 \times 10^{-5}[\Omega(\mu_4^{PS})] + 1.1520 \times 10^{-4}[\Omega(\mu_4^{MR})] + \\
 & + 1.2011 \times 10^{-6}[\Omega(\mu_5^{PS})] - 9.8202 \times 10^{-5}[\Omega(\mu_5^{MR})] - 3.8263 \times 10^{-5}[\Omega(\mu_8^H)] - \\
 & - 0.0626[\Omega(\mu_2^{Pol})] + 1.6689[\Omega(\mu_1^{Pol})] - 0.0078[\Omega(\mu_5^{Ch})] + 0.1123[\Omega(\mu_3^{Ch})] - 0.6517
 \end{aligned} \tag{1}$$

The clastogenic activity (AC)-relationship model was conformed by 372 organic compounds including known carcinogens, drugs, food additives, agrochemicals, cosmetic materials, medicinal products, and household materials (11). Here the letter Ω is used to indicate that the corresponding variable in brackets was orthogonalized respect to the rest of the variables included in the model. The classification model obtained is given below together with the statistical parameters of the linear discriminant is the squared analysis, where λ is the Wilks' statistics, D^2 Mahalanobis distance and F is the Fisher ratio (Wilks' - $\lambda= 0.629$; $F(14.194)=8.148$; $D^2=2.353$; $p<0.0000$).

Structural alerts generation and bond contribution. It was possible to study the contribution of different fragments at a substructural level, allowing definition of the structural alerts for the flavonoids, phenolic acids and natural coumarins. They are compounds with low molecular weight, which share a chemical skeleton (C6-C3-C6), (C6-C1), (C6-C3) skeleton, respectively (Figure 2).

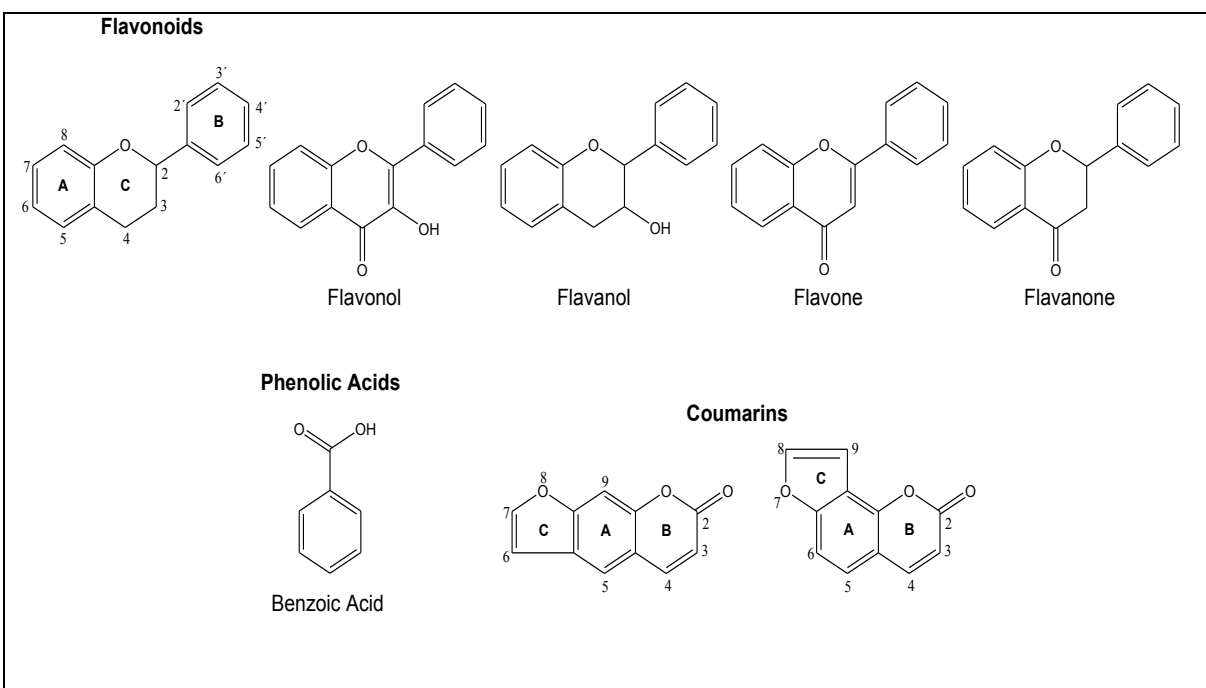


Figure 2. Numeration system of the flavonoids, phenolic acids and coumarins.

There were formed three external DB that allowed externally validate the model (Equation 1). These DB have been published (14,15) and a new report was also performed (Table 1).

The main foods and beverages rich in bioactive compounds of phenolic nature are: tea, wine, coffee, grape, apple, strawberry, artichoke, broccoli, carob and cocoa (16,17). They can be divided according to their basic structure in at least ten different classes. The most described are flavonoids (18), accounting that they represent two-thirds of the phenols from the diet (19). Flavonoids are formed by two phenyl rings (A and B) joined through a pyran ring (heterocyclic ring C) (Figure 2). We analyzed the functional group effect for flavonoid subclasses with reported pro-oxidant activity and compared them with 12 “none reported” flavonoids and 16 “designed compounds” (14).

Table 1. Classification Obtained from the Virtual Screening.

Phenolic compounds	Compounds	Classif. ^a	Prob. ^b	Phenolic compounds	Compounds	Classif. ^a	Prob. ^b
Flavonoids (Flavonol subclasses)	Datisctetin	G_2:1	56,1	Phenolic acids (Benzoic acid derivatives)	p-hydroxybenzoic acid	G_1:- 1	76,6
	Quercetin	G_2:1	67,6		Vanillic acid	G_2:1	59,2
	Myricetin	G_2:1	79,6		Syringic acid	G_2:1	91,6
	Morin	G_2:1	66,5		Protocatechuic acid	G_1:- 1	63,8
	Kaempferol	G_2:1	53,0		Gallic acid	G_2:1	54,7
	Fisetin	G_2:1	53,7		m-hydroxybenzoic acid	G_1:- 1	76,4
	Rutin	G_2:1	69,3		Ellagic acid	G_2:1	93,2
	Galangin	G_1:-1	57,5		Coumarins	Xanthotoxin	G_2:1
Flavonoids (Flavone subclasses)	Apigenin	G_1:-1	66,2	Bergapten		G_2:1	73,2
	Chrysin	G_1:-1	74,9	Xanthotoxol		G_1:- 1	50,8
	Luteolin	G_2:1	51,3	Angelicin		G_1:- 1	67,1
Flavonoids (Isoflavone subclasses)	Gesistein	G_1:-1	52,0	Oroselone		G_1:- 1	79,1
	Gesistin	G_2:1	51,2	Columbianetin		G_1:- 1	82,1
	Daidrein	G_1:-1	66,1				
	Daidrin	G_1:-1	63,2				
	Biochanin A	G_2:1	75,2				
Flavonoids	Naringenin	G_2:1	52,6				

^a classification generated by LDA (Lineal Discriminant Analysis), where G_2: 1 corresponds to the group of active compounds and G_1:-1 to the group of inactive compounds; ^b posterior probability.

(Flavanon subclasses)	Taxifolin	G_2:1	71,3
	Eriodictyol	G_2:1	65,5
	Hesperitin	G_2:1	85,9
	Baicalein	G_2:1	59,7
	Naringin	G_2:1	53,8
Flavonoids (Flavanol subclasses)	Catechin	G_2:1	60,5
	Salicylic acid	G_1:- 1	74,1

It was possible to analyze all the flavonoids' skeleton rings to generate the structural alert rules. This study included a hydroxyl group effect at C3, in rings A, B, and C; a methoxyl group effect; a C2:C3 phenyl group effect; a C2:C3 double-bond ring C effect; and a carbonyl group at ring C effect. This information was corroborated with the bond contributions. The criteria for maximal clastogenicity of pro-oxidant flavonoids were detected as: a) the presence of the 3-hydroxyl group in the C ring, b) the hydroxyl groups at the B ring, principally in *ortho* position and agglutinated, c) the presence of the methoxy and phenyl groups, d) the presence of the 4-cabonyl group at the C ring and e) no presence of 2,3-double bond in the C ring (14).

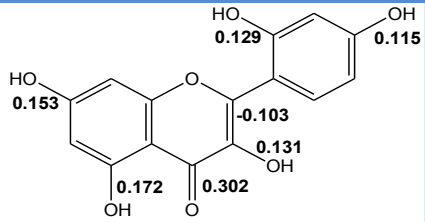
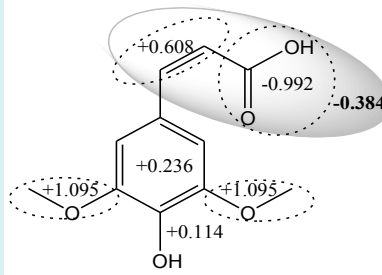
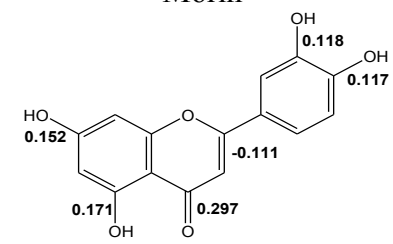
Benzoic acid derivatives are phenolic acids with pro-oxidant report (15). To identify the structural alerts, different aspects were analyzed: a) effect of the amount of hydroxyl groups on the benzene ring, b) effect of the amount of methoxy groups in the ring, c) effect of the position of the hydroxyl and methoxy groups in the ring (-*ortho*, -*meta* and -*para*), d) effect of hydroxyl and methoxy groups isolated or accumulated, e) effect of the replacement of hydroxyl by methoxy substituents on the benzene ring. It was identified that the main structural feature of the benzoic acid derivatives which are associated with a maximum clastogenicity is the amount of hydroxyl and methoxy groups. Methoxy groups will further increase clastogenic activity regardless of how they are near and far apart.

Coumarins are also found in fruits, green tea and other plants (20). The natural coumarins were the elements for the third prepared DB (21). As shown in Table 1, most of these compounds were predicted as inactive.

Methoxy and hydroxyl groups stimulate the molecule to be active ($\text{OCH}_3 > \text{OH}$). The presence of a methoxy group at positions 5 y 9 of the coumarin ring (compounds bergapten y xanthotoxin, respectively) is an important structural factor. The presence of alkyl chains inactivates the molecule, caused mainly by the presence of vinilic and methyl groups (compound oroselone y columbianetin, respectively).

The virtual screening performed with the three external DB (flavonoids, benzoic acid and coumarin derivatives) was validated through the percent of good classification (14,15). In addition, the analyzed alerts were corroborated from calculation of fragments. Table 2 shows an example of calculation of fragments. Methoxy and hydroxyl substituents contributed positively to the activity, aspect corroborated analyzing the fragment contribution. As said before, the contribution of methoxy groups is quantitatively higher than the hydroxyl ones.

Table 2. Representative Examples of the Bond Contributions Obtained from the TOPS-MODE Classification Model.

Class	Fragment contributions	Class	Fragment contributions
Flavonoid (Flavonol)	 <p>Morin^a</p>	Phenolic acid	 <p>Sinapic acid^b</p>
Flavonoid (Flavone)	 <p>Luteolin^a</p>		

^a Reference (14); ^b Reference (15)

Conclusions

Regarding possible clastogenic activity of different compounds described as pro-oxidants, it can be summarized that the main structural features, which are associated with the clastogenicity effect, are the amount of hydroxyl and methoxy groups present in the structure. These structural modifications represent an indicator of the toxicity and also a good strategy for the design of new derivatives, which do not exhibit this activity. This study represents an interesting tool to better understand the properties of natural substances in food. Therefore, it is also helpful in the development of functional foods or nutraceuticals and drug design.

Acknowledgements

The authors thank the Spanish Ministry (PS09/00501), Xunta da Galicia (PGIDIT09CSA030203PR, PGIDT08MMA011200PR and IN845B-2010/089) for the partial financial support. MJM thanks to Fundação para a Ciência e Tecnologia for a PhD grant (SFRH/BD/61262/2009).

References

1. Siân BA, DG. L. European Research on the Functional Effects of Dietary Antioxidants. *Molecular Aspects of Medicine* [serial on the Internet]. 2002 Available from: <http://www.elsevier.com/inca/publications/>.
2. Aruoma O. Methodological considerations for characterizing potential antioxidant actions of bioactive components in plant foods. *Mutation Research*. 2003;523-524:9-20.
3. Breimer L. Molecular Mechanisms of Oxygen Radical Car-cinogenesis and Mutagenesis. The Role of DNA Base Damage. *Mol Carcinog*. 1990;3:188-97.
4. Natarajan A. Chromosome aberrations: past, present and future. *MutatRes*. 2002;504(3-16).
5. Yordi EG, Molina E, Matos M, Uriarte E. Antioxidant and pro-oxidant effects of polyphenolic compounds and structure-activity relationship evidence. In: Bouayed J, Bohn T, editors. *Nutrition, Well-Being and Health*. Croatia: InTech; 2012. p. 23-48.
6. Bello J. *Calidad de vida, Alimentos y Salud humana. Fundamentos científicos*. n.d. ed. Madrid: Editorial Díaz de Santos; 2005.
7. Cambrige. ChemDraw. 2005.
8. ModesLab. Modeslab. 2002.
9. Estrada E, Molina E. Novel local (fragment-based) topological molecular descriptors for QSPR/QSAR and molecular design. *J Mol Graphics and Model*. 2001;20 54-64.
10. StatSoft I. STATISTICA (data analysis software system). 6. ed2001.

11. Estrada E, Molina E. Automatic extraction of structural alerts for predicting chromosome aberrations of organic compounds. *J Mol Graphics and Model.* 2006;25:275-88.
12. Yordi EG. Sistema de cribado computacional para la identificación de compuestos prooxidantes, presentes en alimentos, que dañan al ADN: flavonoides: Universidad de La Habana, Cuba; 2010.
13. Yordi EG, Molina EP, Matos MJ, Uriarte EV. Structural Alerts for Predicting Clastogenic Activity of Pro-oxidant Flavonoid Compounds: Quantitative Structure—Activity Relationship Study *J Biomol Screen.* [online first]. 2012 September 21.
14. Yordi EG, Molina E, Matos M, Uriarte E. Structural alerts for predicting clastogenic activity of pro-oxidant flavonoid compounds: quantitative structure-activity relationship study. *J Biomol Screen.* 2012;17 (2):216-24.
15. Yordi EG, Matos MJ, Pupo RC, Santana L, Uriarte E, Molina E. QSAR study of the potential clastogenic activity of phenolic acids in: 16th International Electronic Conference on Synthetic Organic Chemistry. Santiago de Compostela 2012 [cited 2012 noviembre]; Available from: <http://www.sciforum.net/conf/ecsoc-16/sections/computational/>.
16. Shahidi F, Nacsk M. Food Phenolics: Sources, Chemistry, Effects, and Application. n.d., editor. Lancaster: Technomic Publishing Company; 1995.
17. United States Department of Agriculture [USDA]. USDA Database for the Flavonoid Content of Selected Foods, Release 2.1 Agriculture Research Service; 2007 [cited 2011 April]; Available from: <http://www.ars.usda.gov/SP2UserFiles/Place/12354500/Data/Flav/flav02-1.mdb>.
18. Bors W, Heller W, Michel C, Stettmaier K. Flavanoids and Polyphenols: Chemistry and Biology. In: Packer L, Cadenas, E, editor. *Handbook of Antioxidants.* New York: Marcel Dekker; 1996. p. 409-66.
19. Scalbert A, Williamson G. Dietary intake and bioavailability of polyphenols. *J Nutr.* 2000;130:2073S-85S.
20. Mirunalini S, Krishnaveni M. Coumarin: A Plant derived Polyphenol with wide Biomedical Applications. *International Journal of PharmTech Research.* 2011;3(3):1693-6.
21. Ng T, Liu F, Wang Z. Antioxidative activity of natural products from plants. *Life Sciences.* 2000;66(8):709-23.