A preliminary DFT analysis of phenolic acids in connection with their phytotoxic activity.

Juan S. Gómez-Jeria* and Héctor R. Bravo*

Department of Chemistry, Faculty of Sciences, University of Chile, Las Palmeras 3425, Santiago 7800003, Chile
_____

Abstract

We employed a formal QSAR method to find the main interactions regulating the variation of the ability to inhibit the germination of Lactuca sativa seeds by a small group of phenolic acids. The same techique was employed to obtain structure-retention factor relationships. All molecular geometries were fully optimized at the B3LYP/6-31G(d,p) level of theory. From the corrected Mulliken Population Analysis results the numerical values for all electronic local atomic reactivity indices (LARIS) were calculated. Statistically significant results were obtained for both properties. The variation of the numerical values of both properties seems to be associated with interactions with electron rich centers.

Keywords: *Lactuca sativa*, phenolic acids, QSAR, allelochemicals, phytotoxicity, retention factor.
_____

INTRODUCTION

This is the second part of our research concerning the electronic factors regulating the inhibition of the germination of *Lactuca Sativa* seeds by small molecules. In this paper we present the results corresponding to the inhibitory capacity of a small group of phenolic acids. The references regarding this field of research can be found in our previous work [1]. Also, and for the first time, we use in this paper a new tool that may in some cases improve the analysis of the resulting QSAR equations.

MATERIALS AND METHODS

The results presented here are obtained from what is now a routinary procedure [2]. For this reason, we employ here a general model for the paper's structure [2]. This model contains *standard* phrases for the presentation of the methods, calculations and results because these sections do not need to be rewritten repetitively. The method has been fully presented in earlier publications [3-8]. Therefore, we shall discuss only the results obtained in this study. The results of the successful application of the KPG method can be found elsewhere ([9-16] and references therein [17]).

Selection of molecules and biological activities

The selected molecules are a group of phenolic acids taken from a previous study [18]. Their general formula and biological activity are displayed, respectively, in Fig. 1 and Table 1. The biological activities analyzed here are the percentage of germination inhibition (I) and the capacity factor (or retention factor, k') of these compounds. The capacity factor is a means of determining the retention of an analyte on the chromatographic column. A high k value shows that the sample is highly retained and has spent a substantial amount of time interacting with the stationary

_____

phase. There is a linear relationship between log(k′) and the octanol/water partition coefficient (log(P)), a measure of the liphophilicity of chemicals.
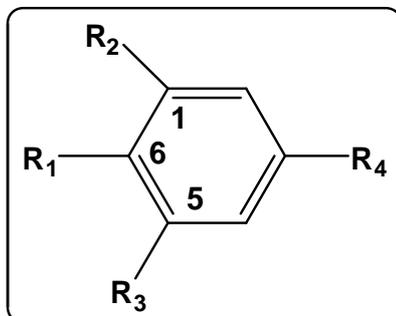
Figure 1. General formula of phenolic acids.

Table 1. Phenolic acids, percentage of germination inhibition of lettuce seeds and retention factors

Mol.	R_1	R_2	R_3	R_4	log(I) (250 µgmL^{-1})	log(k')
1	H	H	H	COOH	1.99	0.30
2	OH	H	H	COOH	1.94	0.23
3	Br	H	H	COOH	0.74	-0.41
4	CN	H	H	COOH	0.63	-0.37
5	Cl	H	H	COOH	--	-0.51
6	OH	OMe	H	COOH	--	-1.15
7	OH	OH	OH	COOH	1.92	0.98
8	H	H	H	CH=CHCOOH	1.73	-0.1
9	OH	OH	H	CH=CHCOOH	2.00	0.83

Calculations

The electronic structure of all molecules in their neutral form was calculated within the Density Functional Theory (DFT) at the B3LYP/6-311g(d,p) level with full geometry optimization. The Gaussian suite of programs was used [19]. All the information needed to calculate numerical values for the local atomic reactivity indices was obtained from the Gaussian results with the D-Cent-QSAR software [20]. All the electron populations smaller than or equal to 0.01 e were considered as zero [8]. Negative electron populations coming from Mulliken Population Analysis were corrected as usual [21]. Since the resolution of the system of linear equations is not possible because we have not enough molecules, we made use of Linear Multiple Regression Analysis (LMRA) techniques to find the best solution. For each case, a matrix containing the dependent variable (the biological activity of each case) and the local atomic reactivity indices of all atoms of the common skeleton as independent variables was built. The Statistica software was used for LMRA [22]. We worked with the *common skeleton hypothesis* stating that there is a definite collection of atoms, common to all molecules analyzed, that accounts for nearly all the biological activity. The action of the substituents consists in modifying the electronic structure of the common skeleton and influencing the right orientation of the drug. It is hypothesized that different parts or this common skeleton accounts for almost all the interactions leading to the expression of a given biological activity. The common skeleton is shown in Fig. 2. X7, X8 and X9 design, respectively, the atoms directly bonded to C1, C6 and C5 (Fig. 1).
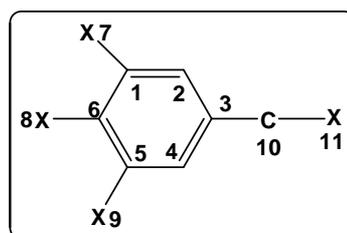
Figure 2. Common skeleton of phenolic acids

_____

RESULTS

Results for the percentage of germination inhibition of lettuce seeds.

The best equation obtained was:

$$\log(I) = -1.13 - 0.58 S_9^N(LUMO)* + 1.43 F_8(LUMO+2)* \tag{1}$$

with n=7, R=0.99, R^2=0.98, adj-R^2=0.98, F(2,4)=135.92 (*p*<0.0001) and SD=0.09. No outliers were detected and no residuals fall outside the $\pm 2\sigma$ limits. Here, $S_9^N(LUMO)*$ is the nucleophilic superdelocalizability of the lowest vacant MO localized on atom 9 and $F_8(LUMO+2)*$ is the Fukui index (the electron population) of the third lowesr vacant MO localized on atom 8. Tables 2 and 3 show the beta coefficients, the results of the t-test for significance of coefficients and the matrix of squared correlation coefficients for the variables of Eq. 1. There are no significant internal correlations between independent variables (Table 3). Figure 3 displays the plot of observed *vs.* calculated log(I).

Table 2. Beta coefficients and t-test for significance of coefficients in Eq. 1

Reactivity index	Beta	t(4)	p-level
$S_9^N(LUMO)*$	-1.24	-16.08	<0.00009
$F_8(LUMO+2)*$	0.55	7.19	<0.002

Table 3. Matrix of squared correlation coefficients for the variables in Eq. 1

	$S_9^N(LUMO)*$	$F_8(LUMO+2)*$
$S_9^N(LUMO)*$	1.00	
$F_8(LUMO+2)*$	0.38	1.00
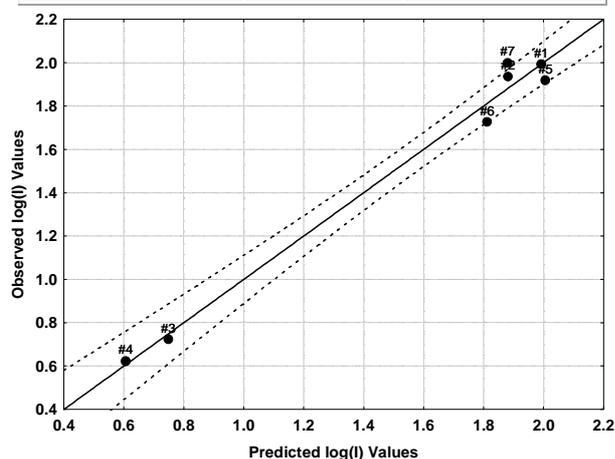
Figure 3. Plot of predicted *vs*. observed log(I) values (Eq. 1). Dashed lines denote the 95% confidence interval

The associated statistical parameters of Eq. 1 indicate that this equation is statistically significant and that the *simultaneous* variation of the numerical values of a group of two local atomic reactivity indices of atoms of the common skeleton explains about 98% of the variation of log(I) in this group of phenolic acids. Figure 3, spanning about 1.4 orders of magnitude, shows that there is a good correlation of observed *versus* calculated values and that almost all points are inside the 95% confidence interval. This can be considered as an indirect evidence that the common skeleton hypothesis works relatively well for this set of molecules. A very important point to stress is the following. When a local atomic reactivity index of a inner occupied MO (i.e., HOMO-1 and/or HOMO-2) or of a

_____

higher vacant MO (LUMO+1 and/or LUMO+2) appears in any equation, this means that the remaining of the upper occupied MOs (for example, if HOMO-2 appears, upper means HOMO-1 and HOMO) or the remaining of the empty MOs (for example, if LUMO+1 appears, lower means the LUMO) contribute to the interaction. Their absence in the equation only means that the variation of their numerical values does not account for the variation of the numerical value of the biological property.

Results for the retention factor.
The best equation obtained was:

$$\log(k') = 0.65 - 3.12 F_9(LUMO+2)* + 0.04 S_8^N(LUMO+2)* - 0.31 S_1^E(HOMO)* \qquad (2)$$

with n=9, R=0.99, R^2=0.98, adj-R^2=0.97, F(3,5)=92.06 (p<0.00009) and SD=0.11. No outliers were detected and no residuals fall outside the ±2σ limits. Here, $S_8^N(LUMO+2)*$ is the nucleophilic superdelocalizability of the third lowest vacant MO localized on atom 8, $S_1^E(HOMO)*$ is the electrophilic superdelocalizability of the highest occupied MO localized on atom 1 and $F_9(LUMO+2)*$ is the Fukui index of the third lowest vacant MO localized on atom 9. Tables 5 and 6 show the beta coefficients, the results of the t-test for significance of coefficients and the matrix of squared correlation coefficients for the variables of Eq. 2. There is only one significant internal correlation between independent variables (Table 6). Figure 4 displays the plot of observed *vs.* calculated log(k').

Table 4. Beta coefficients and t-test for significance of coefficients in Eq. 2

Reactivity index	Beta	t(5)	p-level
$F_9(LUMO+2)*$	-1.11	-15.74	<0.00002
$S_8^N(LUMO+2)*$	0.45	6.38	<0.001
$S_1^E(HOMO)*$	-0.27	-4.53	<0.006

Table 5. Matrix of squared correlation coefficients for the variables in Eq. 2

	$S_1^E(HOMO)*$	$S_8^N(LUMO+2)*$	$F_9(LUMO+2)*$
$S_1^E(HOMO)*$	1.00		
$S_8^N(LUMO+2)*$	0.14	1.00	
$F_9(LUMO+2)*$	0.07	0.54	1.00
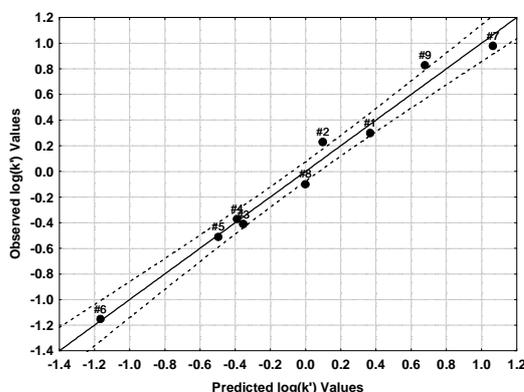
Figure 4. Plot of predicted *vs*. observed log(k') values (Eq. 2). Dashed lines denote the 95% confidence interval

_____

The associated statistical parameters of Eq. 2 indicate that this equation is statistically significant and that the variation of the numerical values of a group of three local atomic reactivity indices of atoms of the common skeleton explains about 97% of the variation of log(k') in this group of phenolic acids. Figure 4, spanning about 2.4 orders of magnitude, shows that there is a good correlation of observed *versus* calculated values and that almost all points are inside the 95% confidence interval. This can be considered as an indirect evidence that the common skeleton hypothesis works relatively well for this set of molecules.

Local Molecular Orbitals of phenolic acids.
Table 6 shows the local MO structure of atoms 1, 8 and 9 (see Fig. 2). Nomenclature: Molecule (HOMO) / (HOMO-2)* (HOMO-1)* (HOMO)* - (LUMO)* (LUMO+1)* (LUMO+2)*. Lp means lone pair.

Table 6. Local Molecular Orbitals of atoms 1, 8 and 9

	Atom 1	Atom 8	Atom 9
1 (32)	30σ31π32π-33π34π36π	23α25σ27σ-35σ37σ39σ	25α27σ28σ-35σ37σ38σ
2 (36)	34σ35π36π-37π38π41π	32π33π36π-37π39σ41π	27σ30σ31σ-39σ40σ42σ
3 (49)	47σ48π49π-50π51π52σ	46lp47lp49π-50π52σ53σ	40σ42σ43σ-52σ55σ56σ
4 (38)	36σ37π38π-39π40π41π	34π35π38π-39π41π42σ	30σ31σ34σ-42σ43σ44σ
5 (40)	38σ39π40π-41π42π45π	36π37lp40π-41π43σ44σ	31σ33σ34σ-43σ46σ47σ
6 (44)	42σ43π44π-45π46π50π	41π43π44π-45π47σ50π	34σ36σ37σ-47σ48σ51σ
7 (44)	42σ43π44π-45π46π47π	41π43π44π-45π46σ50π	41π43π44π-46σ47π50π
8 (39)	36π38π39π-40π41π42π	28σ31σ33σ-43σ44σ46σ	31σ33σ34σ-43σ44σ45σ
9 (47)	44π46π47π-48π49π51π	44π46π47π-48π51π52σ	36σ40σ42σ-50σ52σ53σ

DISCUSSION

Inhibition of germination.

Table 2 shows that the importance of variables in Eq. 1 is $S_9^N(LUMO)^* > F_8(LUMO+2)^*$.

A variable-by-variable analysis (this is an approximate approach) shows that a high germination inhibitory capacity is associated with small (positive) values for $F_8(LUMO+2)^*$ and with high values for $S_9^N(LUMO)^*$ when this reactivity index is positive. To obtain high positive values for $S_9^N(LUMO)^*$ we must shift downwards the corresponding eigenvalue and making this MO more reactive. Therefore atom 9 interacts with an electron rich center. Table 6 shows that all but one LUMOs have a σ nature (R_3 = H in eight molecules and OH in one, Table 1). To explain the appearance of this reactivity index in Eq. 1 it is necessary to invoke a σ-σ C-H..C interaction.

Low values for the electron population of $(LUMO+2)_8^*$ are required for high inhibition of germination. Table 6 shows that this MO has σ or π nature. To explain this finding we shall employ Hoffmann's classification of MO interactions according the the total number of involved electrons: two electron interactions (occupied-vacant MOs, attractive), four electron interactions (occupied-occupied MOs, repulsive) and zero electron interactions (vacant-vacant MOs, repulsive) [23-25]. Note that the last interaction has not direct energetic consequences. Interactions between vacant MOs have been used have been used to explain some experimental results [26-29]. Within the abovementioned scheme we suggest that $(LUMO+2)_8^*$ MO is engaged in a repulsive interaction with vacant MOs of the partner. A question that puzzled us is how to make an educated guess about the role of the MOs that do not appear in the final equations. In this particular case, what are the specific interactions of $(LUMO+1)_8^*$ and $(LUMO)_8^*$ with the partner? Within the variable-by-variable approach, we reasoned as follows. Figure 5 shows a graph of $F_8(LUMO+2)^*$ vs log(I).

_____
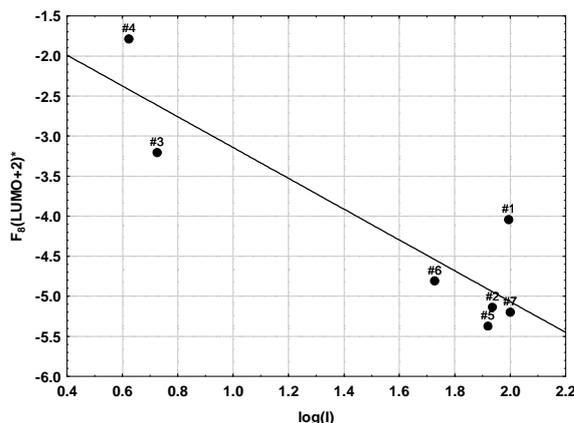
Figure 5. Plot of $F_8(LUMO+2)*$ *vs* **log(I)**

As expected from Eq. 1, the inhibitory capacity diminishes when the value of $F_8(LUMO+2)*$ increases. Figure 6 shows the relationship between $F_8(LUMO+1)*$ and log(I).
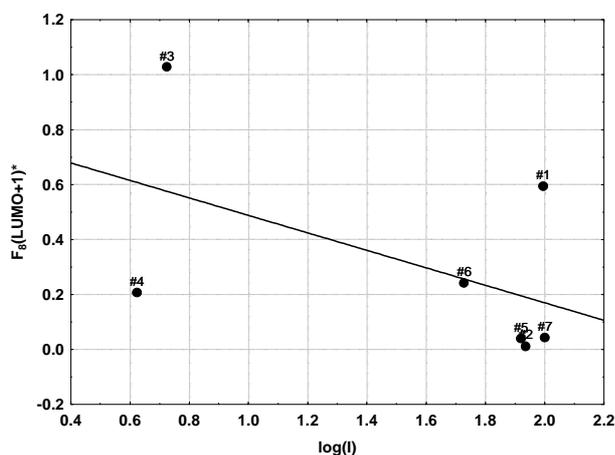
Figure 6. Plot of $F_8(LUMO+1)*$ *vs* **log(I)**

In general we can see that log(I) diminishes when the value of $F_8(LUMO+1)*$ increases. This plot and the ones shown below show that our analysis is approximate: it is the simultaneous variation of the values of all reactivity indices appearing in the equations that provides an account of the variation of the biological activity. In this case we may think that $(LUMO+1)_8^*$, having σ or π natures (Table 6), is also engaged in a repulsive interaction with vacant MOs of the partner. Figure 7 shows the relationship between $F_8(LUMO)*$ and log(I).

_____
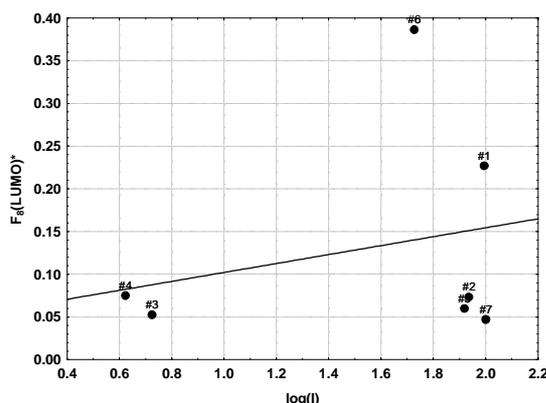
Figure 7. Plot of $F_8(LUMO)*$ *vs* **log(I)**

We can see that in this case the inhibitory activity remains almost constant when the value of $F_8(LUMO)*$ increases. The exception are molecules 1 and 8 (Table 1). In almost all molecules $(LUMO)_8^*$ has a π nature. Therefore, we suggest that, in most molecules, $(LUMO)_8^*$ interacts with an electron-rich center. Considering that molecules having a σ $(LUMO)_8^*$ also interact (molecules 1 and 8, $R_1=H$, Table 1), we are not in position for the moment to be more specific about the possible nature of the occupied MOs of the partner.

All the suggestions are displayed in the partial 2D pharmacophore of Fig. 8.
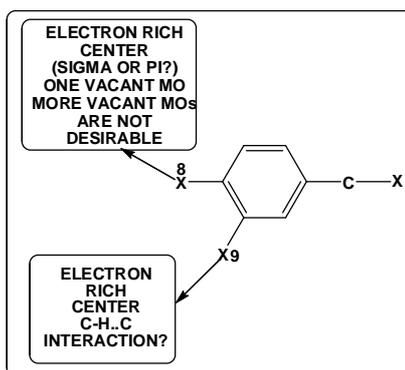
Figure 8. Partial 2D pharmacophore for log(I)

Retention factor, k'.

Table 4 shows that the importance of variables in Eq. 2 is $F_9(LUMO+2)* \gg S_8^N(LUMO+2)* > S_1^E(HOMO)*$. A small retention factor is associated with high (positive) values for $F_9(LUMO+2)*$, small (negative) values for $S_1^E(HOMO)*$ and, if positive, small values for $S_8^N(LUMO+2)*$. Small (negative) values for $S_1^E(HOMO)*$ are obtained by shifting downwards the associated eigenvalue, making this MO less reactive. It is possible then that this MO is engaged in a repulsive interaction with occupied MOs of the partner. Considering that the local frontier MOs of all molecules coincide with the molecules' frontier orbitals, we suggest that atom 1 is interacting with an electron rich center through its LUMO*. If positive, $S_8^N(LUMO+2)*$ needs to have a small numerical value. This value is obtained by shifting upwards the corresponding eigenvalue making the MO less reactive. Figure 9 shows the plot of $S_8^N(LUMO+2)*$ *vs* log(k').

_____
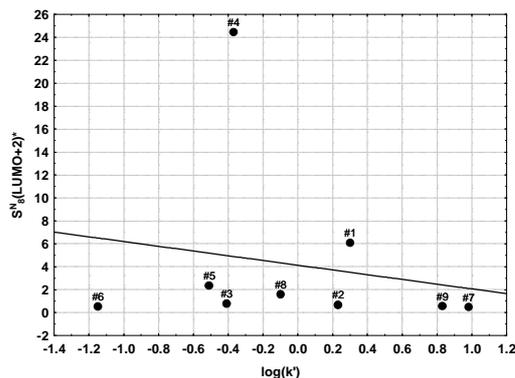
Figure 9. Plot of $S_8^N(LUMO+2)*$ *vs* **log(k')**

We can see that, with the exception of molecule 4, no noticeable changes of log(I) with changes in the values of $S_8^N(LUMO+2)*$ are observed. Therefore in this case the plot does not provide any valuable information. From the equation itself we may suggest that $(LUMO+2)_8^*$ seems to be engaged in a repulsive interaction with vacant MOs of the partner. Regarding the role of $(LUMO+1)_8^*$ and $(LUMO)_8^*$, figures 10 and 11 show, respectively, the plot of $S_8^N(LUMO+1)*$ *vs* log(k') and the plot of $S_8^N(LUMO)*$ *vs* log(k').
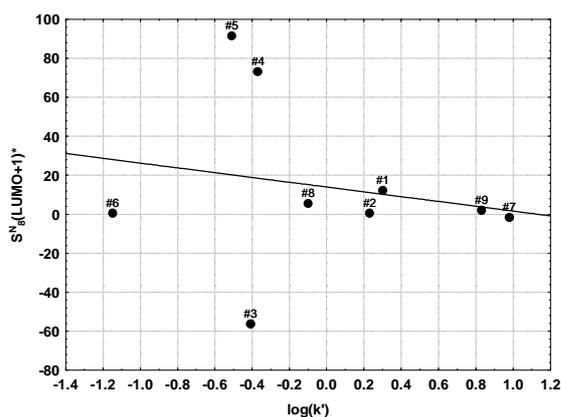
Figure 10. Plot of $S_8^N(LUMO+1)*$ *vs* **log(k')**
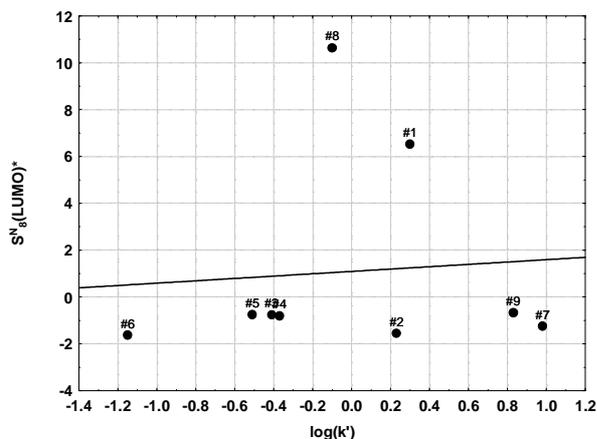
Figure 11. Plot of $S_8^N(LUMO)*$ *vs* **log(k')**

_____

We can see in Fig. 10 that, in general, log(k') increases when the value of $S_8^N (LUMO + 1)*$ diminishes. This can be an indication that $(LUMO+1)_8^*$ is also engaged in a repulsive interaction with vacant MOs of the partner. Figure 11 shows that, if we do not consider molecules 1 and 8, log(k') and $S_8^N (LUMO)*$ are independent. In this case it seems that $(LUMO+1)_8^*$ could be interacting with an electron rich center. Small retention factor is associated with high (positive) values for $F_9 (LUMO + 2)*$. Figure 12 shows the plot of $F_9 (LUMO + 2)*$ *vs* log(k').
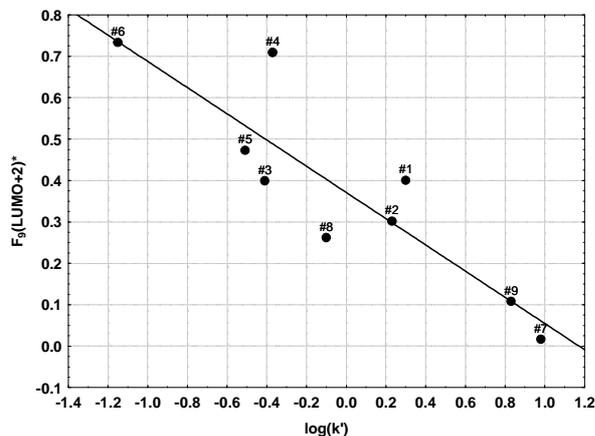
Figure 12. Plot of $F_9 (LUMO + 2)*$ *vs* **log(k')**

This plot coincides with the requirements for $F_9 (LUMO + 2)*$. The plots of $F_9 (LUMO + 1)*$ *vs* log(k') and $F_9 (LUMO)*$ *vs* log(k') show the same trend. This suggests that atom 9 is interacting with an electron rich center through its first three lowest vacant MOs. All the suggestions are displayed in the partial 2D pharmacophore of Fig. 13.
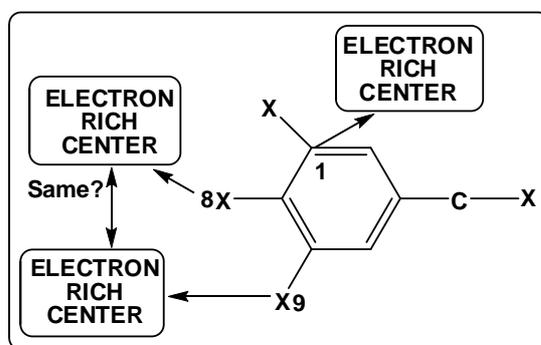
Figure 13. Partial 2D pharmacophore for the retention factor

In summary, we have employed a formal QSAR method to develop an equation to predict the phytotoxicity of phenolic acids and to find the main interactions regulating the variation of the ability to inhibit the germination of *Lactuca sativa* seeds by a small group of phenolic acids. The same techique was employed to obtain structure-retention factor relationships that predict the lipophylic properties of phenolic acids. The variation of the numerical values of both properties seems to be associated with the variation of some interactions with electron rich centers. Eq. 2 is then a suitable expression to predict the lipophilicity of phenolic acids. This last property is involved on the bioactivity of molecules.

REFERENCES

[1] HR Bravo, BE Weiss-López, J Valdebenito-Gamboa, JS Gómez-Jeria, *Res. J. Pharmac. Biol. Chem. Sci.*, **2016**,

_____

7, 792-798.

[2] Note. The results presented here are obtained from what is now a routinary procedure. For this reason, we built a general model for the paper's structure. This model contains *standard* phrases for the presentation of the methods, calculations and results because they do not need to be rewritten repeatedly.

[3] JS Gómez-Jeria, *Int. J. Quant. Chem.*, **1983**, 23, 1969-1972.

[4] JS Gómez-Jeria, "Modeling the Drug-Receptor Interaction in Quantum Pharmacology," in *Molecules in Physics, Chemistry, and Biology*, J. Maruani Ed., vol. 4, pp. 215-231, Springer Netherlands, **1989**.

[5] JS Gómez-Jeria, M Ojeda-Vergara, *J. Chil. Chem. Soc.*, **2003**, 48, 119-124.

[6] DA Alarcón, F Gatica-Díaz, JS Gómez-Jeria, *J. Chil. Chem. Soc.*, **2013**, 58, 1651-1659.

[7] JS Gómez-Jeria, *Elements of Molecular Electronic Pharmacology (in Spanish),* Ediciones Sokar, Santiago de Chile, **2013**.

[8] JS Gómez-Jeria, *Canad. Chem. Trans.*, **2013**, 1, 25-55.

[9] A Robles-Navarro, JS Gómez-Jeria, *Der Pharma Chem.*, **2016**, 8, 417-440.

[10] JS Gómez-Jeria, Í Orellana, *Der Pharma Chem.*, **2016**, 8, 476-487.

[11] JS Gómez-Jeria, C Moreno-Rojas, *Der Pharma Chem.*, **2016**, 8, 475-482.

[12] JS Gómez-Jeria, R Cornejo-Martínez, *Der Pharma Chem.*, **2016**, 8, 329-337.

[13] JS Gómez-Jeria, S Abarca-Martínez, *Der Pharma Chem.*, **2016**, 8, 507-526.

[14] J Valdebenito-Gamboa, JS Gómez-Jeria, *Der Pharma Chem.*, **2015**, 7, 543-555.

[15] MS Leal, A Robles-Navarro, JS Gómez-Jeria, *Der Pharm. Lett.*, **2015**, 7, 54-66.

[16] JS Gómez-Jeria, J Valdebenito-Gamboa, *Der Pharma Chem.*, **2015**, 7, 103-121.

[17] Note, *All papers of Gómez-Jeria can be found in https://www.researchgate.net/profile/Juan-Sebastian_Gomez-Jeria or http://repositorio.uchile.cl/discover filtertype=author&filter_relational_operator=equals&filter=G%C3%B3mez+Jeria%2C+Juan+Sebasti%C3%A1n*,

[18] HR Bravo, SV Copaja, M Lamborot, "Phytotoxicity of Phenolic Acids From Cereals," in *Herbicides - Advances in Research*, A. J. Andrew J. Price, J. A. Kelton Eds., pp. 37-49, http://www.intechopen.com/books/herbicides-advances-in-research/phytotoxicity-of-phenolic-acids-from-cereals#exportas, **2013**.

[19] MJ Frisch, GW Trucks, HB Schlegel, GE Scuseria, MA Robb, et al., "G03 Rev. E.01," Gaussian, Pittsburgh, PA, USA, **2007**.

[20] JS Gómez-Jeria, "D-Cent-QSAR: A program to generate Local Atomic Reactivity Indices from Gaussian 03 log files. v. 1.0," Santiago, Chile, **2014**.

[21] JS Gómez-Jeria, *J. Chil. Chem. Soc.*, **2009**, 54, 482-485.

[22]Statsoft, "Statistica v. 8.0," 2300 East 14 th St. Tulsa, OK 74104, USA, 1984-2007.

[23]R Hoffmann, *Ang. Chem. Int. Ed.*, **1987**, 26, 846-878.

[24]R Hoffmann, *Solids and surfaces: a chemist's view of bonding in extended structures,* VCH Verlagsgesellschaft, Weinheim, **1988**.

[25]R Hoffmann, *Rev. Mod. Phys.*, **1988**, 60, 601-628.

[26]RW Balk, G Boxhoorn, TL Snoeck, GC Schoemaker, DJ Stufkens, et al., *J. Chem. Soc. Dalton Trans.*, **1981** 1524-1530.

[27]RW Balk, T Snoeck, DJ Stufkens, A Oskam, *Inorg. Chem.*, **1980**, 19, 3015-3021.

[28]RW Balk, DJ Stufkens, A Oskam, *J. Chem. Soc. Dalton Trans.*, **1981** 1124-1133.

[29]E Joselevich, *Ang. Chem. Int. Ed.*, **2004**, 43, 2992-2994.