Science of the Total Environment 408 (2009) 277–285

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

QSAR analysis on *Spodoptera litura* antifeedant activities for flavone derivatives

Pablo R. Duchowicz [a,*], Mohammad Goodarzi [a], Marco A. Ocsachoque [b], Gustavo P. Romanelli [b,c], Erlinda del V. Ortiz [d], Juan. C. Autino [c], Daniel O. Bennardi [c], Diego M. Ruiz [c], Eduardo A. Castro [a]

[a] *Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA (UNLP, CCT La Plata-CONICET), Diag. 113 y 64, Sucursal 4, C.C. 16, 1900 La Plata, Argentina*
[b] *Centro de Investigación y Desarrollo en Ciencias Aplicadas "Dr. J. J. Ronco" (CINDECA), Departamento de Química, Facultad de Ciencias Exactas, UNLP-CONICET. Calle 47 N° 257, B1900AJK La Plata, Argentina*
[c] *Cátedra de Química Orgánica, Facultad de Ciencias Agrarias y Forestales, UNLP. Calles 60 y 119, B1904AAN La Plata, Argentina*
[d] *Facultad de Tecnología y Ciencias Aplicadas, Universidad Nacional de Catamarca, Av. Maximio Victoria 55, (4700), Catamarca, Argentina*

ARTICLE INFO

Article history:
Received 7 July 2009
Received in revised form 4 September 2009
Accepted 24 September 2009
Available online 20 October 2009

Keywords:
QSAR Theory
Flavone derivatives
Replacement Method
pED_{50}
Antifeedant activity

ABSTRACT

We establish useful models that relate experimentally measured biological activities of compounds to their molecular structure. The pED_{50} feeding inhibition on *Spodoptera litura* species exhibited by aurones, chromones, 3-coumarones and flavones is analyzed in this work through the hypothesis encompassed in the Quantitative Structure–Activity Relationships (QSAR) Theory. This constitutes a first necessary computationally based step during the design of more bio-friendly repellents that could lead to insights for improving the insecticidal activities of the investigated compounds.

After optimizing the molecular structure of each furane and pyrane benzoderivative with the semiempirical molecular orbitals method PM3, more than a thousand of constitutional, topological, geometrical and electronic descriptors are calculated and multiparametric linear regression models are established on the antifeedant potencies. The feature selection method employed in this study is the Replacement Method, which has proven to be successful in previous analyzes. We establish the QSAR both for the complete molecular set of compounds and also for each chemical class, so that acceptably describing the variation of the inhibitory activities from the knowledge of their structure and thus achieving useful predictive results. The main interest of developing trustful QSAR models is that these enable the prediction of compounds having no experimentally measured activities for any reason. Therefore, the structure–activity relationships are further employed for investigating the antifeedant activity on previously synthesized 2-,7-substituted benzopyranes, which do not pose any measured values on the biological expression. One of them, 2-(α-naphtyl)-4H-1-benzopyran-4-one, results in a promising structure to be experimentally analyzed as it has predicted $pED_{50} = 1.162$.

1. Introduction

The background of this study consists on developing useful models that relate experimental insect antifeedant activities of compounds to their molecular structure. The environmental contamination caused by an extensive use of chemical insecticides is a well-known problem, leading to the need of replacing these agents by insecticides of natural origin, posing reduced or no-harm effects to the environment (Picman et al., 1995). It is known that repellents are more bio-friendly than exterminating agents. Insect antifeedants act as repellents and often have only weak insecticidal activity (Morimoto et al., 1999). Several classes of organic compounds have been studied as antifeedants (Kokubun et al., 2003; Ley, 2005; Morimoto et al., 2007; Stevenson et al., 2003).

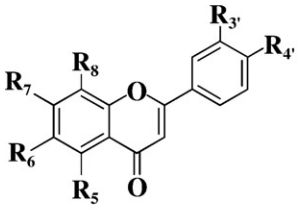
Flavones and chromones are important heterocyclic compounds belonging to the flavonoid family that occur naturally in plants. There exist more than 4000 chemically unique flavonoids that have become very popular during past years due to their health promoting effects, as they exhibit many biological activities such as anti-bacterial, anti-fungal, anti-oxidant and anti-cancer (Martens and Mithöfer, 2005). Such kind of compounds are important constituents of the human diet, being derived largely from fruits, vegetables, nuts, seeds, stems and flowers, and thus constitute one of the most important classes of metabolites.

Flavones and chromones have a significant impact on various aspects of plant biology. They are capable of absorbing harmful UV–B radiation, thus they can act as UV filters (Harborne and Williams, 2000). Moreover, they are involved in various interactions with other organisms, microbes as well as insects or other plants. They act as co-pigments in flowers providing colors to attract pollinators and fruit and seed dispersors (Martens and Mithöfer, 2005).

* Corresponding author. Tel.: +54 221 425 7430, +54 221 425 7291; fax: +54 221 425 4642.
E-mail addresses: pabloducho@gmail.com, prduchowicz@yahoo.com.ar (P.R. Duchowicz).

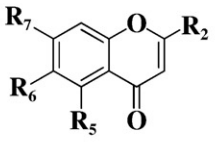
0048-9697/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.scitotenv.2009.09.041

Flavone derivatives: 1-17



N°	R_5	R_6	R_7	R_8	$R_{3'}$	$R_{4'}$
1	-OH	-H	-OH	-H	-H	-H
2	-OH	-H	-OH	-H	-H	-OH
3	-OH	-H	-OH	-H	-OH	-OH
4	-OH	-H	-OH	$-OCH_3$	-H	-H
5	-OH	-H	$-OCH_3$	-OH	-H	-H
6	-OH	-H	-OH	-OH	-H	-H
7	-OH	-H	$-OCH_3$	$-OCH_3$	-H	-H
8	-OH	-OH	-OH	-H	-H	-H
9	-OH	$-OCH_3$	-OH	-H	-H	-H
10	-OH	$-OCH_3$	$-OCH_3$	-H	-H	-H
11	$-OCH_3$	$-OCH_3$	$-OCH_3$	$-OCH_3$	$-OCH_3$	$-OCH_3$
12	-H	-H	-H	-H	-H	-H
13	-H	$-CH_3$	-H	-H	-H	-H
14	-H	-OH	-H	-H	-H	-H
15	-H	$-OCH_3$	-H	-H	-H	-H
16	-H	-H	-OH	-H	-H	-H
17	-H	-H	$-OCH_3$	-H	-H	-H

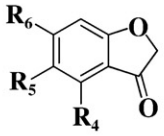
Chromone derivatives: 18-23



N°	R_2	R_5	R_6	R_7
18	-H	-H	-H	-H
19	$-CH_3$	-H	-H	$-CH_3$
20	$-CH_3$	$-OCH_3$	-H	$-OCH_3$
21	$-CH_3$	$-OCH_3$	$-OCH_3$	$-OCH_3$
22	$-CH_3$	-H	-H	$-OCH_3$
23	-H	-H	-H	$-OCH_3$

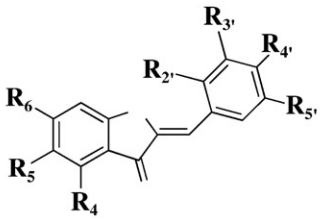
Fig. 1. Molecular structures of flavones, chromones, 3-coumarones, and aurones analyzed with linear QSAR models.

3-Coumarone derivatives: 24-27



N°	R_4	R_5	R_6
24	-H	-H	-H
25	$-OCH_3$	-H	$-OCH_3$
26	$-CH_3$	-H	$-CH_3$
27	$-OCH_3$	$-OCH_3$	$-OCH_3$

Aurone derivatives: 28-46



N°	R_4	R_5	R_6	$R_{2'}$	$R_{3'}$	$R_{4'}$	$R_{5'}$
28	-H	-H	-H	-H	-H	-H	-H
29	-H	-H	-H	$-OCH_3$	-H	-H	-H
30	-H	-H	-H	-H	$-OCH_3$	-H	-H
31	-H	-H	-H	-H	-Cl	-H	-H
32	-H	-H	-H	-H	-Br	-H	-H
33	-H	-H	-H	-H	-H	$-OCH_3$	-H
34	-H	-H	-H	-H	$-OCH_3$	$-OCH_3$	-H
35	-H	-H	-H	-H	$-OCH_2$-	-O-	-H
36	-H	-H	-H	-H	-OH	-OH	-H
37	-H	-H	-H	-H	$-OCH_3$	-OH	-H
38	-H	-H	-H	-H	$-OCH_2CH_3$	-OH	-H
39	-H	-H	-H	-H	$-OCH_3$	-H	$-OCH_3$
40	-H	-H	-H	-H	$-OCH_3$	$-OCH_3$	$-OCH_3$
41	$-OCH_3$	-H	-H	$-OCH_3$	-H	-H	-H
42	$-OCH_3$	-H	$-OCH_3$	-H	$-OCH_3$	$-OCH_3$	-H
43	$-OCH_3$	$-OCH_3$	$-OCH_3$	-H	$-OCH_3$	$-OCH_3$	-H
44	$-OCH_3$	-H	$-OCH_3$	-H	$-CH_3$	$-CH_3$	-H
45	$-CH_3$	-H	$-CH_3$	-H	$-CH_3$	$-CH_3$	-H
46	$-CH_3$	-H	$-CH_3$	-H	$-OCH_3$	$-OCH_3$	-H

Fig. 1 (*continued*).

Flavone related compounds are able to affect insects in various ways, for instance, they inhibit larvae feeding or act as feeding deterrent. Ohmura et al. (2000) studied antifeedant activity of some flavones and their related compounds against the subterranean termite *Coptotermes* sp. Morimoto et al. (2003) evaluated antifeedant property of various flavones (among them, flavone, 6-methylflavone, 6-methoxyflavone and 7-methoxyflavone) against the common cutworm *Spodoptera litura*. The relationship between compounds that affect the GABA(A) receptor and insect antifeedants has already been reported (Eichenseer and Mullin, 1997). The insect taste sensitivity is dominated by GABA(A) receptors and active compounds against one of the receptors act as the insect antifeedant.

Among the different modern methods available in the literature for predicting properties of substances based on their molecular structure is the Quantitative Structure–Activity Relationships (QSAR) Theory, whose pioneer works were developed by Hansch in 1964 (Hansch, 1990; Hansch and Leo, 1995). The main hypothesis involved in any QSAR is the assumption that the variation of the behaviour of chemical compounds, as expressed by any experimentally measured biological or physicochemical property on such compounds, can be correlated with numerical entities related to some aspect of the chemical structure termed molecular descriptors (Katritzky et al., 1995; Todeschini and Consonni, 2000; Trinajstic, 1992). Descriptors are generally used to describe different characteristics/attributes of certain structure in order to yield information about the activity/property being studied. QSAR techniques are usually based on statistically determined linear or non-linear models that relate the chemical behaviour of compounds with their descriptors. The main interest of developing predictive QSAR models is that these enable the prediction of compounds having no experimentally measured activity for many different reasons, either because they are unstable, toxic, or simply because their measurement requires too much time or is expensive.

This work collects the available experimental information from the literature and establishes useful QSARs on the *S. litura* antifeedant activities of aurones, chromones, 3-coumarones and flavones through linear regression models. Even though some linear regressions were established in previous analyzes of the inhibitory activity (Morimoto et al., 2007; Morimoto et al., 2003), a present existing gap is that these studies were rather qualitative and that none of them considered a greater number of compounds and more sophisticated modeling techniques as employed here. Furthermore, these studies considered single regressions of the antifeedant activity of active compounds and their melting points, hydrogen bonding parameters (R_f), and lipophilic parameters ($\log_{10}k$), without considering and a great number of molecular descriptors involving more elaborated definitions of the structure.

We consider that the linear methodology is the best statistical technique for analyzing present data set, as few experimental data are available and thus it is necessary to employ the lowest number of optimized parameters during the model development. Among the most recent advances in this field is the Spectral-Structure Activity Relationship (S-SAR) method (Lacrama et al., 2007; Putz and Lacrama, 2007; Putz et al., 2009), which enables to replace the Multivariable Linear Regression analysis by purely algebraic models with some conceptual and computational advantages, having both environmental and biological applications. In this work, we resort to the Replacement Method (RM) as variable subset selection approach, as this technique has been successful for selecting relevant structural descriptors (Duchowicz et al., 2006; Duchowicz et al., 2005; Duchowicz et al., 2008a, b; Goodarzi et al., 2009). Finally, another main interest of present research is to apply the so derived QSAR models for estimating the antifeedant activity of some new 2-,7-substituted benzopyranes (Bennardi et al., 2008), for which there still are no experimental activities. Few attempts were carried out in past years to synthesize flavonoids with substitutions of such types.

2. Methods

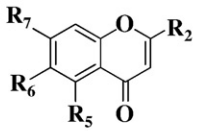
2.1. Experimental data

The experimental ED_{50} antifeedant activities of aurones, chromones, and flavones expand the range 0.035–5.6 [$\mu mol\,cm^{-2}$] and are collected from recent publications (Morimoto et al., 2003, 2007). The observations are converted into minus logarithm scale ($pED_{50} = -\log_{10}ED_{50}$) for modeling purposes. Fig. 1 displays the molecular structures of the compounds which are employed for establishing the various QSARs. The estimation set composed of substituted flavones and chromones with unknown experimental activities were prepared following a procedure described elsewhere (Bennardi et al., 2008), and are included in Fig. 2.

2.2. Calculation of molecular descriptors

The initial conformations of the compounds are drawn by means of the "model build" modulus available in HyperChem 6.03 (Hyperchem, 2009). Each molecular structure is firstly preoptimized with the Molecular Mechanics Force Field (MM+) procedure and the resulting geometry is further refined by means of the semiempirical method PM3 (Parametric Method-3). We choose a gradient norm limit of 0.01 kcal $Å^{-1}$.

The numerical descriptors for each compound are calculated with Dragon (2009) and include several variable types characterizing the 1D, 2D, and 3D aspects of structure: constitutional, topological, geometrical, charge, GETAWAY (GEometry, Topology and Atoms-Weighted AssemblY), WHIM (Weighted Holistic Invariant Molecular descriptors), 3D-MoRSE (3D-Molecular Representation of Structure based on Electron diffraction), molecular walk counts, BCUT



N°	R_2	R_5	R_6	R_7
47	$-C_6H_5$	-H	-H	$-OCH_3$
48	$-C_6H_5$	-H	-H	-Cl
49	$-C_6H_5$	-H	-H	-Br
50	α-napthyl	-H	-H	-H
51	β-napthyl	-H	-H	-H
52	α-napthyl	-H	-H	-Br
53	β-napthyl	-H	-H	-Br
54	α-napthyl	-H	-H	-Cl
55	β-napthyl	-H	-H	-Cl
56	α-napthyl	-H	-H	$-CH_3$
57	β-napthyl	-H	-H	$-CH_3$
58	α-napthyl	-H	-H	$-OCH_3$
59	β-napthyl	-H	-H	$-OCH_3$
60	α-napthyl	-H	-H	-F

Fig. 2. Estimation set of flavone and chromone derivatives without pED_{50} experimental activities.

descriptors, 2D-Autocorrelations, aromaticity indices, Randic molecular profiles, radial distribution functions, functional groups and atom-centred fragments. We also add quantum-chemical descriptors to the pool such as HOMO and LUMO energies, and HOMO–LUMO gap ($\Delta_{HOMO-LUMO}$). The total number of calculated descriptors resulted in 1500 variables.

2.3. Modeling strategy

2.3.1. Variable subset selection method

In recent years researchers have focused an increasing attention on finding the most efficient tool for variable selection in QSAR/QSPR studies. There are a lot of feature selection methods to find the best structural descriptors from a pool of variables and the Replacement Method (RM) (Duchowicz et al., 2005, 2006), employed here, was successfully used elsewhere (Duchowicz et al., 2008a, b; Goodarzi et al., 2009). In brief, the RM is an efficient optimization tool which generates multi-parametric linear regression QSPR models by searching the set **D** of D descriptors for an optimal subset **d** of $d \ll D$ ones with minimum model's standard deviation S. The quality of the results achieved with this technique is quite close to that obtained by performing an exact (combinatorial) full search (FS) of molecular descriptors, although, of course, requires much less computational work. We used the computer Matlab 5.0 system for all our calculations (Matlab, 2004).

2.3.2. Internal and external validation of QSAR models

In this sort of theoretical studies it results of crucial importance to check the predictive capability of the model through validation. In this way, one verifies that the linear relationships established behave not only correlative but would also function similarly well for the prediction of new data not contemplated during the training stage of the model. The consistency and reliability of a method can be explored using the Leave-One-Out Cross Validation procedure (*loo*) (Hawkins et al., 2003), which we employ here. Another sort of internal validation we use is the Y-randomization technique (Wold and Eriksson, 1995) consisting of scrambling the observed pED_{50} in such a way that they do not correspond to the respective molecules. After analyzing 500,000 cases of Y-randomization for each developed QSAR, the smallest S_{rand} achieved is compared to the one found when considering the true calibration (S). Therefore, in case $S_{rand} > S$, it is expected that the QSAR is not fortuitous and does not result from happenstance, and results in real structure–activity relationship.

In our study, the models are then further subjected to external validation by using a test set of fresh structures that are not contemplated during the model development. These structures are selected by hand in such a way that they share similar structural characteristics to the training compounds.

3. Results and discussion

Previous structure–activity studies have suggested that the insect antifeedant activity exhibited by the flavone and chromone derivatives stronghly depends both on the 2-position substituent and the substituted pattern on the A-ring of the benzopyranone (Morimoto et al., 2003). The antifeedant activity decreases due to the 2-position bulky substituents, while the introduction of a substituent to the 6- or 7-position tend to increase the activity for various compounds. For the the case of coumaranone and aurone derivatives, the introduction of alkoxy and alkyl groups to the A and B-rings seems to increase the antifeedant activity (Morimoto et al., 2007).

We establish a structure–activity relationship on all the active aurone and flavone compounds. On the other hand and as we already mention in the Introduction section, it is also our intention to apply the derived QSAR for estimating the antifeedant activity on new 2-,7-substituted benzopyranes. For this purpose, we need to have a model that is successful for describing inactive compounds. Therefore, it would also be necessary to individually model each class of compounds for acceptably describing the variation of the inhibitory activities and thus predicting the structures of the estimation set.

Table 1
Notation for molecular descriptors involved in QSAR models of insect antifeedant activity.

Type	Dim	Molecular descriptor	Description
3D-MoRSE	3D	*Mor24u*	3D-MoRSE-signal 24/unweighted
		Mor28u	3D-MoRSE-signal 28/unweighted
		Mor10u	3D-MoRSE-signal 10/unweighted
GETAWAY	3D	$R4u^+$	*R* maximal autocorrelation of lag 4/unweighted
		$R5e^+$	*R* maximal autocorrelation of lag 5/weighted by atomic Sanderson electronegativities
Topological	2D	*IC1*	Information content index (neighborhood symmetry of 1-order)
Atom-centred fragments	1D	*H-046*	H attached to C0(sp^3) no X attached to next C
Galvez topological charge indices	2D	*GGI7*	Topological charge index of order 7
Radial distribution function	3D	*RDF040u*	Radial distribution function 4.0/unweighted

3.1. QSAR for the combined set of active aurone and flavone analogues

This set includes the active compounds of the four classes: flavones, chromones, 3-coumarones, and aurones. We begin this analysis with the application of the RM variable subset selection technique on the complete set of active aurones, chromones and flavones (27 compounds). In this way, we expect to find the optimal linear regression model that minimizes its standard deviation (S) and

Table 2
Experimental and QSAR predicted pED_{50} inhibitory activities for active aurone and flavone analogues.

Number	Chemical name	Exp.	Eq. (1)
1	Chrysin	−0.398	0.041
4	Wogonin	−0.301	0.217
6	Norwogonin	−0.182	−0.068
7	Moslosooflavone	−0.114	−0.147
8	Baicalein	0.018	−0.125
11	Nobiletin	−0.748	−0.300
12	Flavone	0.959	0.581
13	6-Methylflavone	1.456	1.622
15	6-Methoxyflavone	0.824	1.029
17	7-Methoxyflavone	1.229	0.984
18	Chromone	0.775	0.801
19	2,7-Dimethylchromone	0.991	0.699
21	2-Methyl-5,6,7-trimethoxychromone	0.387	0.186
23	7-Methoxychromone	1.018	0.929
25	4,6-Dimethoxycoumaranone	−0.588	−0.271
26	4,6-Dimethylcoumaranone	0.168	0.261
27	4,5,6-Trimethoxycoumaranone	0.387	−0.493
30	3′-Methoxyaurone	−0.004	0.122
32	3′-Bromoaurone	−0.624	−0.545
33	4′-Methoxyaurone	−0.308	−0.134
34	3′,4′-Dimethoxyaurone	0.066	0.115
35	3′,4′-Methylendioxyaurone	−0.207	−0.144
38	3′-Ethoxy-4′-hydroxyaurone	−0.292	−0.353
39	3′,5′-Dimethoxyaurone	−0.538	−0.682
41	4,6-Dimethoxyaurone	0.071	0.003
42	3′,4,4′,6-Tetramethoxyaurone	0.921	0.671
46	3′,4′-Dimethoxy-4,6-dimethylaurone	−0.009	−0.043

includes the best "representative" molecular descriptors, extracted from the pool containing $D = 1500$ variables. The following three-descriptors linear QSAR is achieved which follows the semiempirical "Rule of Thumb" (Hansch, 1990), stating that at least five or six data points should be present for each fitting parameter:

$$pED_{50} = -2.416(\pm 0.3) + 2.002(\pm 0.4)\cdot Mor24u - 1.430(\pm 0.3)\cdot Mor28u + 38.152(\pm 5)\cdot R4u^{+} \quad (1)$$

$$N = 27,\ R = 0.890,\ S = 0.305,\ F = 28.993,\ p < 10^{-4},\ R_{\max} = 0.563$$
$$R_{loo} = 0.820,\ S_{loo} = 0.384,\ S_{\text{Rand}} = 0.480$$

In this equation, N is the number of compounds used, R is the correlation coefficient, S stands for the model's standard deviation of calibration, p is the significance of the model, $R_{\max}$ is the maximum intercorrelation coefficient between descriptors participating in Eq. (1), and subindex *loo* stands for the Leave-One-Out Cross Validation technique (Hawkins et al., 2003). Parameters of *loo* measure the internal validation of the developed QSAR upon inclusion/exclusion of compounds. The S_{Rand} parameter represents the standard deviation according to the Y-Randomization technique (Wold and Eriksson, 1995) (500,000 cases). A brief description for each molecular descriptor appearing in Eq. (1) and in the next QSAR equations are supplied by Table 1, which appear defined with degree of details in the specialized literature (Todeschini and Consonni, 2000).

Table 2 provides the predicted pED_{50} for all the active compounds according to Eq. (1). Fig. 3A includes a graphical representation of the predictions as function of the experimental values and Fig. 3B plots the residuals as function of the predictions.

3.2. QSAR for flavone derivatives

Present set includes only the benzopyranes of the flavone and chromone classes, as both of them share the same structural

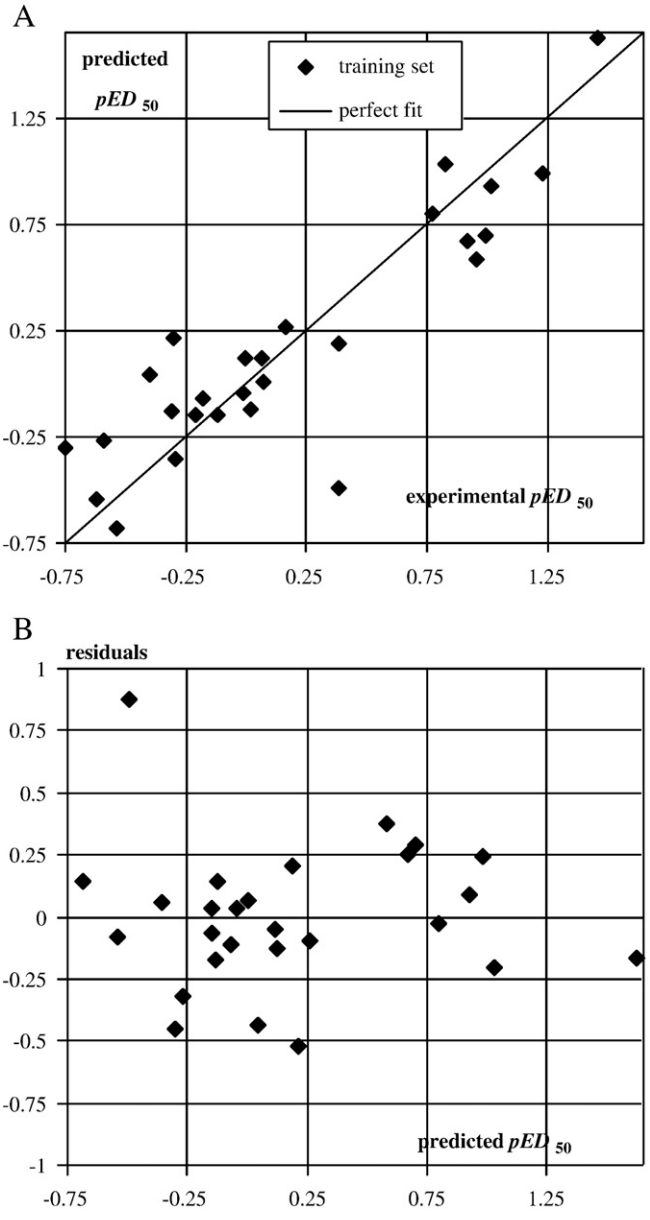


Fig. 3. A. Predicted insect antifeedant activity as function of experimental values for all the active compounds ($N = 27$). B. Residuals versus predicted pED_{50} activities.

Table 3
Experimental and QSAR predicted pED_{50} for flavone and chromone derivatives.

Number	Chemical name	Exp.[a]	Eq. (2)
Training set			
1	Chrysin	−0.398	−0.056
3	Luteolin	Inactive ($-\log_{10}10$)	−1.323
4	Wogonin	−0.301	−0.453
5	Isowogonin	Inactive ($-\log_{10}200$)	−0.996
6	Norwogonin	−0.182	−0.839
8	Baicalein	0.018	−0.731
10	Mosloflavone	Inactive ($-\log_{10}200$)	−0.982
11	Nobiletin	−0.748	−0.634
12	Flavone	0.959	1.087
13	6-Methylflavone	1.456	1.215
14	6-Hydroxyflavone	Inactive ($-\log_{10}15$)	−1.196
16	7-Hydroxyflavone	Inactive ($-\log_{10}10$)	−1.304
17	7-Methoxyflavone	1.229	1.382
18	Chromone	0.775	0.091
19	2,7-Dimethylchromone	0.991	1.232
20	2-Methyl-5,7-dimethoxychromone	Inactive ($-\log_{10}200$)	−1.075
21	2-Methyl-5,6,7-trimethoxychromone	0.387	−0.510
23	7-Methoxychromone	1.018	1.217
Test set			
2	Apigenin	Inactive	−0.840
7	Moslosooflavone	−0.114	−0.222
9	Oroxylin A	Inactive	−0.453
15	6-Methoxyflavone	0.824	2.034
Estimation set			
47	7-Methoxyflavone	–	0.405
48	7-Chloroflavone	–	−1.120
49	7-Bromoflavone	–	−1.120
50	2-(α-Naphtyl)-4H-1-benzopyran-4-one	–	1.162
51	2-(β-Naphtyl)-4H-1-benzopyran-4-one	–	−0.522
52	7-Bromo-2-(α-naphtyl)-4H-1-benzopyran-4-one	–	−0.705
53	7-Bromo-2-(β-naphtyl)-4H-1-benzopyran-4-one	–	−2.443
54	7-Chloro-2-(α-naphtyl)-4H-1-benzopyran-4-one	–	−0.651
55	7-Chloro-2-(β-naphtyl)-4H-1-benzopyran-4-one	–	−2.388
56	7-Methyl-2-(α-naphtyl)-4H-1-benzopyran-4-one	–	0.762
57	7-Methyl-2-(β-naphtyl)-4H-1-benzopyran-4-one	–	0.708
58	7-Methoxy-2-(α-naphtyl)-4H-1-benzopyran-4-one	–	−0.668
59	7-Methoxy-2-(β-naphtyl)-4H-1-benzopyran-4-one	–	−0.180
60	7-Fluoro-2-(α-naphtyl)-4H-1-benzopyran-4-one	–	−0.271

[a] Inactive compounds have $pED_{50} < -1$. Modeled value is indicated in parentheses.

backbone. As can be observed from Table 3 many of the molecules are inactive, being experimentally reported with a censored value exceeding certain threshold, as $pED_{50} < -1$. This would make it impossible to design quantitative models due to the unavailability of the numerical value. However, it is to be noted that a censored value represents very important information that a model has to capture during its calibration, in order to accurately represent the chemical universe of compounds.

We take into account inactive compounds by assigning to them a given experience-based number, in accordance with the threshold $pED_{50} < -1$, with the aim of guiding the model's performance on these missing data. The numbers assigned to the 6 inactive compounds appearing in the training set are indicated in parentheses in Table 3, and these values are manually selected in such a way that reasonable predictions are achieved for the activity of the training compounds. According to this, the best relationship found involves the next three theoretical descriptors:

$$pED_{50} = 11.483(\pm 2) - 6.202(\pm 1)\cdot IC1 + 54.296(\pm 12)\cdot R5e^{+} + 0.357(\pm 0.2)\cdot H\text{–}046 \quad (2)$$

$$N = 18,\ R = 0.856,\ S = 0.663,\ F = 12.789,\ p < 10^{-4},\ R_{\max} = 0.748$$
$$R_{loo} = 0.801,\ S_{loo} = 0.774,\ S_{\text{Rand}} = 0.770$$

The graphical representation of Eq. (2) is given in Fig. 4A and B. The test set for validating this QSAR involves 2 active and 2 inactive flavones and from Table 3 it is seen that these compounds are acceptably predicted. By inspection of the predictions of this table one may analyze the behaviour of this model, leading to the conclusion that compounds predicted to have low negative pED_{50} activities may or not be inactive, but those posing positive values are active. Therefore, this QSAR is able to guide the synthesis of relevant structures with insect antifeedant activity.

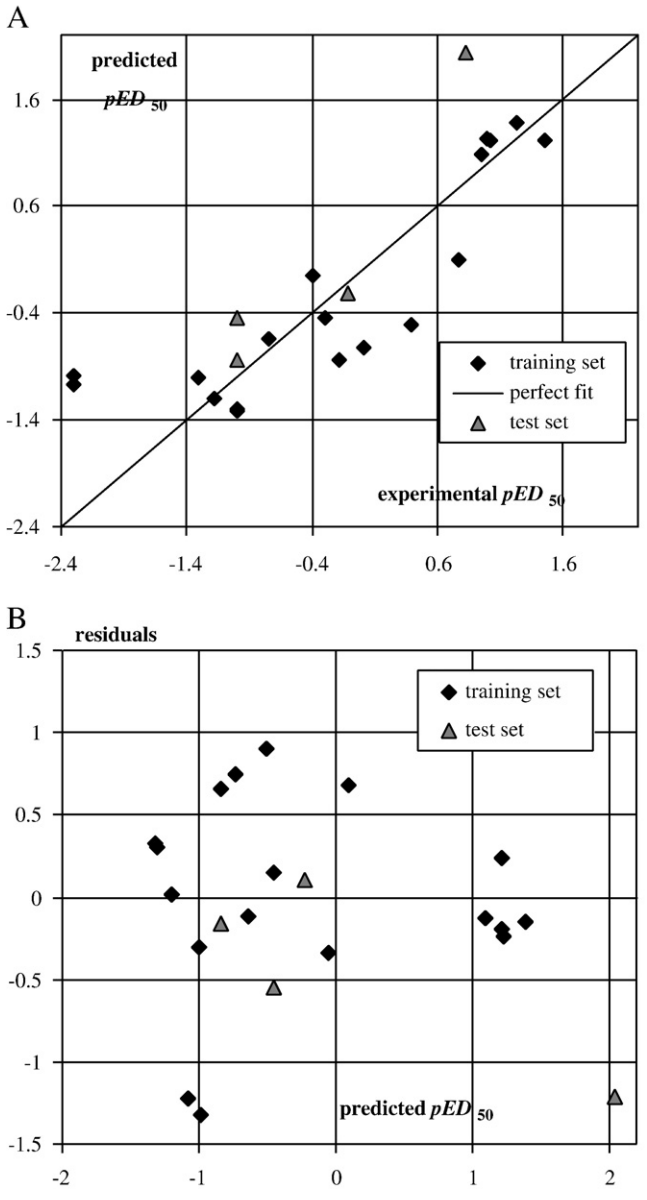


Fig. 4. A. Predicted insect antifeedant activity as function of experimental values for active and inactive flavones and chromones. B. Residuals versus predicted pED_{50} activities.

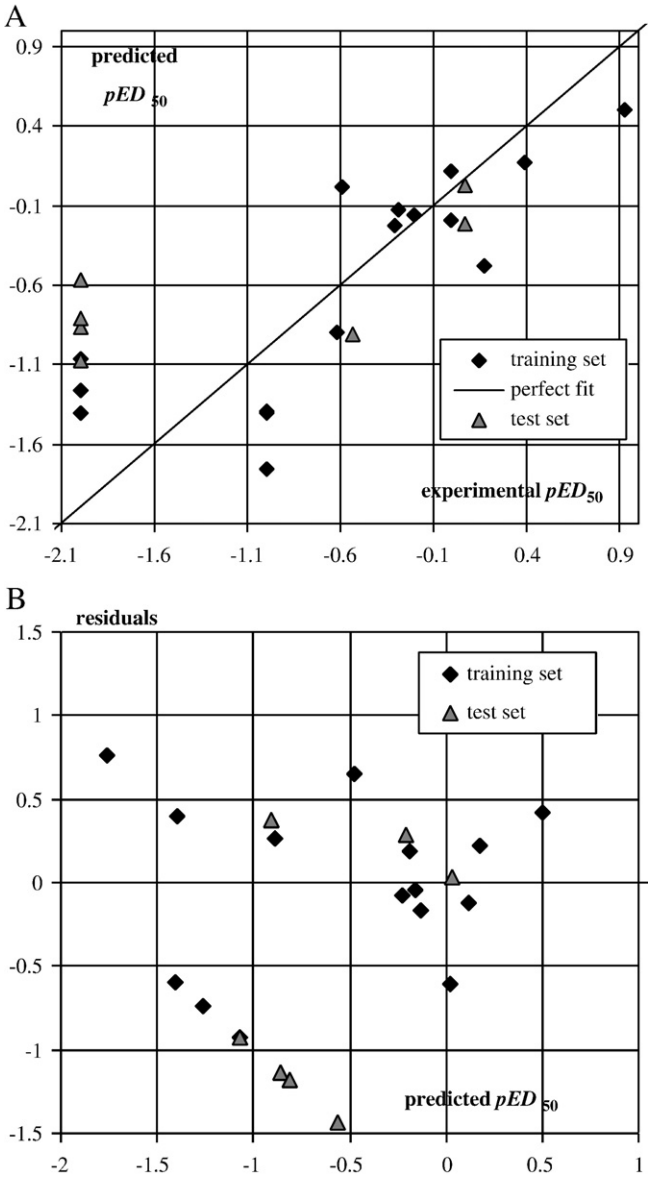


Fig. 5. A. Predicted insect antifeedant activity as function of experimental values for active and inactive aurones and 3-coumarones. B. Residuals versus predicted pED_{50} activities.

Table 4
Experimental and QSAR predicted pED_{50} for aurone and 3-coumarone derivatives.

Number	Chemical name	Exp.[a]	Eq. (3)
Training set			
24	Coumaranone	Inactive ($-\log_{10}100$)	−1.404
25	4,6-Dimethoxycoumaranone	−0.588	0.021
26	4,6-Dimethylcoumaranone	0.168	−0.479
27	4,5,6-Trimethoxycoumaranone	0.387	0.171
30	3′-Methoxyaurone	−0.004	−0.189
32	3′-Bromoaurone	−0.624	−0.893
33	4′-Methoxyaurone	−0.308	−0.230
35	3′,4′-Methylendioxyaurone	−0.207	−0.158
36	3′,4′-Dihydroxyaurone	Inactive ($-\log_{10}10$)	−1.758
38	3′-Ethoxy-4′-hydroxyaurone	−0.292	−0.127
40	3′,4′,5′-Trimethoxyaurone	Inactive ($-\log_{10}100$)	−1.265
42	3′,4,4′,6-Tetramethoxyaurone	0.921	0.503
43	3′,4,4′,5,6-Pentamethoxyaurone	Inactive ($-\log_{10}10$)	−1.396
44	4,6-Dimethoxy-3′,4′-dimethylaurone	Inactive ($-\log_{10}100$)	−1.068
45	4,6,3′,4′-Tetramethylaurone	Inactive ($-\log_{10}10$)	−1.400
46	3′,4′-Dimethoxy-4,6-dimethylaurone	−0.009	0.115
Test set			
28	Aurone	Inactive	−0.565
29	2′-Methoxyaurone	Inactive	−0.863
31	3′-Chloroaurone	Inactive	−0.814
34	3′,4′-Dimethoxyaurone	0.066	0.027
37	4′-Hydroxy-3′-methoxyaurone	Inactive	−1.075
39	3′,5′-Dimethoxyaurone	−0.538	−0.911
41	4,6-Dimethoxyaurone	0.071	−0.211

[a] Inactive compounds have $pED_{50} < -1$. Modeled value is indicated in parentheses.

We apply Eq. (2) for predicting the estimation set of structures provided by Fig. 2 (see Table 3), and find that one of them, 2-(α-naphtyl)-4H-1-benzopyran-4-one, results a promising structure to be experimentally analyzed as it has predicted $pED_{50} = 1.162$. In addition, it can be appreciated that among these structures the α-naphthyl derivatives tend to be more actives than the β-naphthyl ones.

3.3. QSAR for aurone derivatives

In this set appear the benzofuranes of the aurone and 3-coumarone classes. By considering inactive compounds as in the previous section, the following QSAR is obtained over 16 structures:

$$pED_{50} = -2.459(\pm 0.5) - 7.111(\pm 2)\cdot GGI7 + 0.305(\pm 0.07)\cdot RDF040u + 1.265(\pm 0.4)\cdot Mor10u \quad (3)$$

$$N = 16,\ R = 0.813,\ S = 0.564,\ F = 13.200,\ p < 10^{-4},\ R_{\max} = 0.840$$
$$R_{loo} = 0.640,\ S_{loo} = 0.773,\ S_{\mathrm{Rand}} = 0.590$$

Fig. 5A and B plots the predictions and residuals of Eq. (3). In this case, the test set for validating the QSAR involves 3 active and 4 inactive aurones whose predicted activities are included in Table 4, demonstrating that the model has predictive capability.

It is noted that in all the established models, there is no serious intercorrelation between the participating descriptors. In both structure–activity relationships expressed by Eqs. (2) and (3), the inclusion of inactive compounds greatly affect the statistical quality of the derived models. Therefore, it would also be possible to obtain improved models that consider only the active compounds on each chemical class although, however, this would not allow us to apply them for any kind of investigated structure (active or inactive), as it is done in present work.

4. Conclusions

One of the major problems commonly found in computational modeling is due to incomplete experimental information on a modeled system, which would prevent to reasonably validate the mathematical models depicting the variation of the observations. Present work surmounts this problem by including inactive compounds during the model design and allows achieving a better description of the chemical universe, thus having special value when the model is to be used for predicting molecules with unknown experimental activities. The best linear QSAR regression equations established in this study are able to acceptably predict the insect antifeedant activities of compounds and provide insights into active and inactive structures. The application of the designed QSAR on newly synthesized 2-,7-substituted benzopyranes leads to a favourable predicted inhibitory activity for 2-(α-naphtyl)-4H-1-benzopyran-4-one, and now it remains to be experimentally investigated in greater extent.

Acknowledgments

M.G., G.P.R., M.A.O., E.A.C. and P.R.D. would like to thank to the National Research Council of Argentina [Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina] and to Universidad Nacional de La Plata for financial support. M.G., G.P.R., E.A.C. and P.R.D. are Research Members of CONICET.

References

Bennardi DO, Romanelli GP, Jíos JL, Autino JC, Baronetti GT, Thomas HJ. Synthesis of substituted flavones and chromones using a Wells–Dawson heteropolyacid as catalyst. Arkivoc 2008;11:123–30.
Dragon. Milano Chemometrics and QSAR Research Group; 2009. http://michem.disat.unimib.it/chm.
Duchowicz PR, Castro EA, Fernández FM, González MP. A new search algorithm of QSPR/QSAR theories: normal boiling points of some organic molecules. Chem Phys Lett 2005;412:376–80.
Duchowicz PR, Castro EA, Fernández FM. Alternative algorithm for the search of an optimal set of descriptors in QSAR–QSPR studies. MATCH Commun Math Comput Chem 2006;55:179–92.
Duchowicz PR, Garro JCM, Castro EA. QSPR study of the Henry's law constant for hydrocarbons. Chemom. Intell. Lab. Syst. 2008a;91:133–40.
Duchowicz PR, Talevi A, Bruno-Blanch LE, Castro EA. New QSPR study for the prediction of aqueous solubility of drug-like compounds. Bioorg Med Chem 2008b;16:7944–55.
Eichenseer H, Mullin CA. Antifeedant comparisons of GABA/glycinergic antagonists for Diabroticite leaf beetle (Coloptera: Chrysomelidae). J Chem Ecol 1997;23:71–82.
Goodarzi M, Duchowicz PR, Wu CH, Fernández FM, Castro EA. New hybrid genetic based support vector regression as QSAR approach for analyzing flavonoids-GABA(A) complexes. J Chem Inf Model 2009;49:1475–85.
Hansch C. Comprehensive drug design. New York: Pergamon Press; 1990.
Hansch C, Leo A. Exploring QSAR. Fundamentals and applications in chemistry and biology. Washington, DC: American Chemical Society; 1995.
Harborne JB, Williams CA. Advances in flavonoid research since 1992. Phytochemistry 2000;55:481–504.
Hawkins DM, Basak SC, Mills D. Assessing model fit by cross validation. J Chem Inf Model 2003;43:579–86.
Hyperchem. 6.03 (Hypercube); 2009. http://www.hyper.com.
Katritzky AR, Lobanov VS, Karelson M. QSPR: the correlation and quantitative prediction of chemical and physical properties from structure. Chem Soc Rev 1995;24:279–87.
Kokubun TV, Veitch NC, Bridge PD, Simmonds MSJ. Dihydroisocoumarins and a tetralone from *Cytospora eucalypticola*. Phytochemistry 2003;62:779–82.
Lacrama A-M, Putz MV, Ostafe V. A Spectral-SAR model for the anionic–cationic interaction in ionic liquids: application to *Vibrio fischeri* ecotoxicity. Int J Mol Sci 2007;8:842–63.
Ley S. Development of methods suitable for natural product synthesis: the azadirachtin story. Pure Appl Chem 2005;77:1115–30.
Martens S, Mithöfer A. Flavones and flavone synthases. Phytochemistry 2005;66:2399–407.
Matlab. The MathWorks Inc; 2004. http://www.mathworks.com (accessed Apr 16, 2009).
Morimoto M, Urakawa M, Fujitaka T, Komai K. Structure–activity relationship for the insect antifeedant activity of benzofuran derivatives. Biosci Biotechnol Biochem 1999;63:840–6.
Morimoto M, Tanimoto K, Nakano S, Ozaki T, Nakano A, Komai K. Insect antifeedant activity of flavones and chromones against *Spodoptera litura*. J Agric Food Chem 2003;51:389–93.
Morimoto M, Fukumoto H, Nozoe T, Hagiwara A, Komai K. Synthesis and insect antifeedant activity of aurones against *Spodoptera litura* larvae. J Agric Food Chem 2007;55:700–5.
Ohmura W, Doi S, Aoyama M, Ohara S. Antifeedant activity of flavonoids and related compounds against the subterranean termite *Coptotermes formosaunus* Shiraki. J Wood Sci 2000;46:149–53.

Picman AK, Schneider EF, Picman J. Effect of flavonoids on mycelial growth of *Verticillium albo-atrum*. Biochem Syst Ecol 1995;23:683–93.
Putz MV, Lacrama A-M. Introducing spectral structure activity relationship (S-SAR) analysis. Application to ecotoxicology. Int J Mol Sci 2007;8:363–91.
Putz MV, Putz A-M, Lazea M, Ienciu L, Chiriac A. Quantum-SAR extension of the Spectral-SAR algorithm. Application to polyphenolic anticancer bioactivity. Int J Mol Sci 2009;10:1193–214.
Stevenson PC, Simmonds MSJ, Yule MA, Veitch NC, Kite GC, Irwin D, Legg M. Insect antifeedant furanocoumarins from *Tetradium daniellii*. Phytochemistry 2003;63:41–6.
Todeschini R, Consonni V. Handbook of molecular descriptors. Weinheim (Germany: Wiley VCH; 2000.
Trinajstic N. Chemical graph theory. Boca Raton (FL): CRC Press; 1992.
Wold S, Eriksson L. Statistical validation of QSAR results. In: van de Waterbeemd H, editor. Chemometrics methods in molecular design. Weinheim: VCH; 1995. p. 309–18.