Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 25 (2007) 680–690
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Quantitative structure–activity relationship studies for the prediction of antifungal activity of *N*-arylbenzenesulfonamides against *Botrytis cinerea*

Liane Saíz-Urra [a], Maykel Pérez González [a,b,*], Isidro G. Collado [c], Rosario Hernández-Galán [c]

[a] *Chemical Bioactive Center, Central University of Las Villas, Santa Clara, Villa Clara, C.P. 54830, Cuba*
[b] *Service Unit, Experimental Sugar Cane Station "Villa Clara-Cienfuegos", Ranchuelo, Villa Clara, C.P. 53100, Cuba*
[c] *Department of Organic Chemistry, Faculty of Sciences, Cadiz University, Puerto Real, 11510 Cadiz, Spain*

Received 8 February 2006; received in revised form 8 May 2006; accepted 8 May 2006
Available online 12 May 2006

Abstract

The *Botrytis cinerea* is one of the most interesting fungal pathogens. It can infect almost every plant and plant part and cause early latent infections which damage the fruit before ripening. The QSAR is an alternative method for the research of new and better fungicides against *B. cinerea*. This paper describes the results of applying a topological sub-structural molecular design (TOPS-MODE) approach for predicting the antifungal activity of 28 *N*-arylbenzenesulfonamides. The model described 86.1% of the experimental variance, with a standard deviation of 0.223. Leave-one-out and leave-group-out cross validation was carried out with the aim of evaluating the predictive power of the model. The values of their respective squared correlations coefficients were 0.754 and 0.741. The TOPS-MODE approach was compared with three other predictive models, but none of these could explain more than 72.8% of the variance with the same number of variables.

In addition, this approach enabled the assessment of the contribution of different bonds to antifungal activity, thereby making the relationships between structure and biological activity more transparent. It was found that the fungicidal activity of the chemicals analyzed was increased by the presence of a sulfonamide group bonded to two aromatics rings, making this group the most important of the molecule. The majority of the substituents present in the aromatic rings have an electron withdrawing effect and they contribute to a smaller degree than the sulfonamide group to the property under study. The aromatic moiety plays an important role in this activity; its contribution changes with different substituents. Generally, the nitro group has a positive and great contribution to the biological property but when this group is involved in some compounds in *ortho* effect with the SO_2 moiety of the sulfonamide group a lower value of contribution is observed for both groups.

Keywords: Antifungal activity; *Botrytis cinerea*; *N*-arylbenzenesulfonamides; Spectral moments; 3D descriptors; Constitutional descriptors; QSAR

1. Introduction

The *Botrytis cinerea* is certainly one of most interesting fungal pathogens because of its very unique characteristics: it can live pathogenically but also saprophytically, it can be devastating to some crops but also beneficial under certain conditions. It is found throughout the world and can infect almost every plant and plant part. Additionally, it causes early latent infections which damage fruit before ripening [1].

Since the mid-1990s, new compounds with excellent activity against *B. cinerea* have been commercialized. However, strategies to control the fungus with classic fungicides may produce side-effects, notably environmental contamination and the development of multi-resistant fungal strains. Some of these fungicides, such as dicarboximide procymidone, are persistent enough to be detected in vegetables [2], soil [3] and even after vinification [4]. For all these reasons, the development of alternative fungicides has been attracting considerable attention in recent times.

Quantitative structure–activity relationships (QSAR) have been broadly used for some years mainly in medical research [5–9]. This methodology makes use of the molecular

* Corresponding author.
E-mail address: mpgonzalez76@yahoo.es (M.P. González).

1093-3263/$ – see front matter
doi:10.1016/j.jmgm.2006.05.006

descriptors offering valuable and simple information about the structure of the molecules which is used later in the elaboration of the predictive models. Employing it methodology allows cost savings by reducing the laboratory resources needed, and the time required to create and investigate new drugs with certain desired biological activity.

There have been limited studies aimed at exploring the utilization of the QSAR techniques in the rational search for new fungicides against *B. cinerea*. For example, Akagi et al. [10] investigated the antifungal activity of fluazim and the related fungicidal *N*-phenylpyridinamines against this fungus. Fluazinam was selected after lead optimization of the *N*-phenylpyridinamine skeleton which was obtained by lead development from acaricidal compound. The preventive activity against *B. cinerea* of *N*-phenylpyridinamines was analyzed by the technique of QSAR (ALS method) to elucidate the role of substituents on both the pyridine and benzene rings, and also to obtain insight into the mode of action of this series of compounds. Structure–activity relationships on the substituents of the pyridine ring were explained by a combination of electric, steric and hydrophobic parameters, while those on the substituents of the benzene ring were more complicated. Also this research group [11] investigated the structure–activity relationships of pyridylcarbamates active against both benzimidazole-sensitive and -resistant isolates of *B. cinerea*. QSAR analyses and molecular modeling studies were carried out to investigate the structural requirements for highly active compounds and the structural feature of the binding site of each strain. Significantly different QSAR equations were obtained only for substituents at the 6-position of the pyridine ring.

Humeres et al. [12] measured the octanol–water partition coefficients of a series of alkyldithiocarbamates and ethyldithiocarbamates heavy metal salts (Zn, Cu, Fe). The hydrophobicity constant of the dithiocarbamic fragment >NC(S)SH was calculated to be −1.80, following the assumption of additivity of log *P*. The authors postulated that assuming that the fungitoxic activity of dithiocarbamates is due to a neutral free acid species. The partition coefficients of a series of alkyldithiocarbamic acids were calculated by the fragmental method and quantitative measurements of fungitoxicity were made on *Phytium aphanidermatum*, *B. cinerea* and *Rhizoctonia solani*. The diameter of the colonies decreased exponentially with increasing concentrations of the congeners, but for a shorter range of concentrations straight lines were obtained. Iterative optimization and regression analysis showed that the QSAR of fungicidal activity, measured as pC50, depends only on log *P* and that substituent electronic and steric parameters are not statistically significant. The QSAR describes as much as 91% of the variation of the biological response. Another interesting report is the review about a QSAR of 1-*N*-substituted azoles active against *B. cinerea* by Kataoka [13].

On the other hand, the 2D and 3D molecular descriptors have been broadly used in much QSAR research. These descriptors have not been used before in the development of models to predict the fungicidal activity against *B. cinerea*.

In addition, the sulfonamides are an important group of drugs and have been characterized as bactericidal due to their ability to inhibit the synthesis of dihydrofolic acid from *p*-aminobenzoic acid, pteridine, and glutamic acid by means of dihydropteroate synthase, and their competitive antagonism against *p*-aminobenzoic acid [14–16]. This interesting group also plays a role as a fungicide against *Pythium ultimum*, *Phytophthora capsici*, *R. solani* and *B. cinerea* [17–19].

For these reasons, our aim is to develop a predictive model of the antifungal activity for a set of *N*-arylbenzenesulfonamides with various substituents at the arylamine and benzenesulfonyl rings and to make a comparison with other methodologies based on the quality of their statistical parameters of the regression as well as predictive capability of the models generated. In addition, we aim to establish the contribution of several fragments to the properties of these compounds with the intention of later synthesizing new and better fungicidal compounds belonging to the family.

2. Materials and methods

2.1. The spectral moments

The spectral moments are based on the calculation of the bond matrix, whose theoretical basis has been widely described in previous reports [20,21]. Essentially, the bond matrix is the square and symmetric matrix whose entries are ones or zeros according to whether the corresponding bonds are adjacent or not. The order of this matrix (m) is the number of bonds in the molecular graph, two bonds being adjacent if they are incident to a common atom. The spectral moments of the edge adjacency matrix are defined as the traces, i.e., the sum of the main diagonal of the different powers of such a matrix.

In order to apply the above approach to the development of a model for predicting the antifungal activity, the following steps were followed:

(a) an adequate training set of chemicals was selected;
(b) the molecular graphs for each of the training set were drawn;
(c) the molecular bonds with appropriate weights were differentiated;
(d) the spectral moments of the bond matrix for each molecule in the data set was computed;
(e) a quantitative structure–activity relationship (QSAR) was arrived at by using multiple regression analysis (MRA);
(f) the predictive performance of the model was assessed employing the leave one out (LOO) and leave group out (LGO) cross-validation methodologies;
(g) the contribution of the different fragments was ascertained in order to determine their quantitative contribution to the antifungal activity of the molecules studied.

For the development of MRA we used the following expression:

$$P = a_0\mu_0 + a_1\mu_1 + a_2\mu_2 + \cdots + a_k\mu_k + b \quad (1)$$

where P is the activity under study. It is the negative logarithm of the *concentration required to cause 50% inhibition of fungal* (ED_{50}) represented in the following way $-\log(ED_{50})$, μ_k is the kth spectral moment, and the a'_ks are the coefficients obtained by the regression equation.

2.2. Selection of bond weights and calculation of molecular descriptors

The bond distance (Std) and the standard bond dipole moments (Dip) were used as bond weights. On the other hand, Gasteiger–Marsilli charge was used as bond weight too. This charge is an atomic contribution and it was transformed into bond contributions, according to Eq. (2). This transformation has been previously described [22]:

$$w(i,j) = \frac{w_i}{\delta_i} + \frac{w_j}{\delta_j} \quad (2)$$

where, w_i and δ_i are the atomic weight and vertex degree of the atom i, respectively. The calculation of the spectral moment descriptors was carried out with Modeslab 1.0 software [23]. The input of the chemical structures into the Modeslab software was carried out using the familiar SMILES notation. We calculated the first 15 spectral moments (μ_1–μ_{15}) for each bond weight and the number of bonds in the molecules (μ_0). Due to the non-linearity of the biological process under study (fungicidal activity) the interactions between μ_0 and μ_1 and with all variables (descriptors) were evaluated. Spectral moments were calculated considering the molecules without hydrogen atoms.

2.3. Data set

We employed 28 *N*-arylbenzenesulfonamides with various substituents at the arylamine and benzenesulfonyl position to develop a predictive model taking into account the concentration required to cause 50% inhibition of fungal growth (ED_{50}) [24].

In Table 1 is shown the compounds structures with their biological activity.

Each compound was dissolved in methanol or dimethyl sulphoxide (DMSO) and was tested for its in vitro antifungal activity against the plant pathogenic fungi *B. cinerea*. Each serially diluted compound was incorporated in 1/5 strength potato dextrose agar (PDA) plates at 0–200 μg/ml, and a 5-mm agar of the *B. cinerea* was placed at the center of the plate. Three replicate plates of each concentration for the fungus was incubated at 25 °C. After incubating for 2–6 days, the radial growth was measured, and the ED_{50} value for the mycelial growth inhibition was calculated by a probit analysis [24].

2.4. Computational strategies

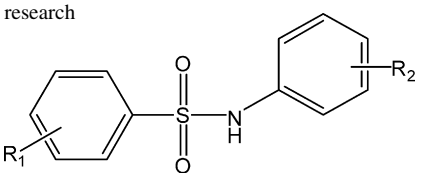
The mathematical model was obtained by means of the MRA as implemented in the STATISTICA software version 6.0 [25]. The genetic algorithm was used as the variable selection strategy for the variable included in the equation. The most significant parameters were identified from the data set using genetic algorithm (GA) analysis of the descriptors obtained by TOPS MODE computer software. The GA is a class of methods based on biological evolution rules [26–28]. The first step is to create a population of linear regression models. These regression models mate with each other, mutate, cross-over, reproduce, and then evolve through successive generations toward an optimal solution. The GA simulation conditions were 10,000 generations and 300 populations.

Analysis of residuals and deleted residuals from the regression equations was used to identify outliers. The statistical significance of the models was determined by examining the squared regression coefficient (R^2), the standard deviation (S), the number of variables, the Fisher ratio (F) and the Akaike information criterion (AIC).

2.5. Validation of the models

The models obtained were validated calculating the cross-validated squared regression coefficient (q^2) values. The q^2 values are calculated from "leave-one-out" (LOO) test and

Table 1
The compounds structures with their biological activity used in the current research



Compounds	R_1	R_2	$-\log(ED_{50})$
1	H	H	−2.379
2	H	*p*-NO_2	−1.818
3	H	*p*-CF_3	−1.170
4	H	*o*-Cl, *p*-CF_3	−1.447
5	H	*o*-CF_3, *p*-NO_2	−1.520
6	H	*m*-CF_3, *p*-F	−1.029
7	*p*-CH_3	H	−1.839
8	*p*-CH_3	*p*-NO_2	−1.550
9	*p*-CH_3	*p*-CF_3	−0.978
10	*p*-CH_3	*o*-Cl, *p*-CF_3	−1.083
11	*p*-CH_3	*o*-CF_3, *p*-NO_2	−0.996
12	*p*-CH_3	*m*-CF_3, *p*-F	−0.898
13	*o*-NO_2	H	−2.547
14	*o*-NO_2	*p*-NO_2	−1.856
15	*o*-NO_2	*p*-CF_3	−1.265
16	*o*-NO_2	*o*-Cl, *p*-CF_3	−1.540
17	*o*-NO_2	*o*-CF_3, *p*-NO_2	−2.612
18	*o*-NO_2	*m*-CF_3, *p*-F	−1.083
19	*m*-NO_2	H	−1.860
21	*m*-NO_2	*p*-CF_3	−1.134
22	*m*-NO_2	*o*-Cl, *p*-CF_3	−2.184
23	*m*-NO_2	*o*-CF_3, *p*-NO_2	−2.157
24	*m*-NO_2	*m*-CF_3, *p*-F	−0.987
25	*p*-NO_2	H	−1.833
26	*p*-NO_2	*p*-NO_2	−1.931
27	*p*-NO_2	*p*-CF_3	−0.792
28	*p*-NO_2	*o*-Cl, *p*-CF_3	−1.803
30	*p*-NO_2	*m*-CF_3, *p*-F	−0.544

from ''leave-group-out'' (LGO) test, also known as cross-validation.

For LOO a data point is removed from the set, and the regression recalculated; the predicted value for that point is then compared to its actual value. This is repeated until each datum has been omitted once; the sum of squares of these deletion residuals can then be used to calculate q^2, an equivalent statistic to R^2. In the LGO method, 25% of the data were eliminating every time in 100 different forms. In this way we guaranteed not to use the same group of compounds for every validation. The q^2 values can be considered a measure of the predictive power of a regression equation: whereas R^2 can always be increased artificially by adding more parameters (descriptors), q^2 decreases if a model is overparameterized [29], and is therefore a more meaningful summary statistic for QSAR models.

2.6. Orthogonalization of descriptors

The orthogonalization process of molecular descriptors was introduced by Randić several years ago as a way of improving the statistical interpretation of the model built by using interrelated indices [30–34]. The main philosophy of this approach is to avoid the exclusion of descriptors on the basis of their collinearity with other variables previously included in the model. The acceptable level of collinearity to avoid is a more subjective issue. In our view, the collinearity of the variables should be as low as possible because the interrelatedness among the different descriptors can result in a highly unstable regression coefficient, which makes it impossible to know the relative importance of an index and underestimates the utility of the regression coefficient model.

The Randić method of orthogonalization has been described in detail in several publications [30–34]. Thus, we will only give a general overview here. The first step in orthogonalizing the molecular descriptors in a model is to select the appropriate order of orthogonalization, which, in this case, is the order in which the variables were selected in the genetic algorithm search procedure of the linear regression analysis. The first variable $\mu_1\mu_9^{GM}$ is taken as the first orthogonal descriptor $\Omega^1\mu_1\mu_9^{GM}$, and the second one is orthogonalized with respect to it by taking the residual of its correlation with $\Omega^1\mu_1\mu_9^{GM}$. The process is repeated until all the variables are completely orthogonalized, and the orthogonal variables are then used to obtain the new model.

2.7. Comparison with other approaches

The use of spectral moments for the prediction of antifungal activity, explained in the previous section, was compared with other methodologies by using DRAGON [35] computer software, version 2.1. Before calculating the DRAGON descriptors; each molecule was optimized geometrically by using the quantum chemical semi-empirical method Austin Model 1 (AM1) included with MOPAC 6.0 software [36]. The DRAGON descriptors used were Constitutional descriptors, Randić molecular profile and geometrical descriptors. The development of these three models involved the use of the same data set that we reported in the previous section. That it is to say, the same data that we used to model the spectral moments. The comparison was based on the quality of the statistical parameters of the regression as well as the predictive capability of the models generated.

In this sense, additional criteria exist to compare the quality of different models. One of these criteria was formulated by Akaike in 1973 [37,38]. Akaike's information criterion (AIC) take into account the statistical goodness of fit and the number of parameters that have to be estimated to achieve that degree of fit. This criterion is calculated using the following equation:

$$\mathrm{AIC} = \mathrm{RSS}\frac{n + p'}{(n - p')^2} \tag{3}$$

where RSS is the sum of the squared differences between the observed (y) and estimated response (y'), n the number of compounds in the training set, and p' is the number of adjustable parameters in the model. When comparing models, the model that produces the minimum value of this statistical parameter should be considered potentially the most useful. We calculated the Akaike values for the four approaches and included these in the comparison.

2.8. Computation of bond contributions

The computation of the fragment contribution has been described in numerous recent reports [39,40]. Next, we will provide an overview of the methodology used. Bond contributions to antifungal activity are calculated on the basis of the local spectral moments, which are defined as the diagonal entries of the different powers of the weighted bond matrix

$$\mu_k^T(i) = b_{ii}(T)^k \tag{4}$$

where $m_k^T(i)$ is the kth local moment of the bond i and $b_{ii}(T)$ are the diagonal entries of the weighted bond matrix, and T is the type of the bond weight: Std, Dip and GM.

It is straightforward to realize that the total moments are the sum of the local moments. Consequently we can substitute Eq. (4) into the QSAR (Eq. (5)) in such a way that the total contribution of the different bonds in a specific molecule is obtained as follows:

$$P = b_0 + \sum_k a_k\mu_k^T \tag{5}$$

These contributions represent the additive features of the property modeled and they can be expressed as fragment contributions, summing the contributions of different bonds that are inside the substructure whose contribution is under determination.

3. Results and discussion

The preliminary model that we found taking into account the descriptors derived by the TOPS-MODE approach was

Table 2
The statistical parameters of the linear regression models obtained for the four kinds of descriptors involved in the comparison

Descriptors	Variables	R^2	S	F	p
Spectral moments	$\mu_1\mu_9^{GM}$, $\mu_1\mu_2^{Dip}$, $\mu_1\mu_{14}^{Std}$, $\mu_1\mu_{10}^{GM}$	0.861	0.223	35.71	$<10^{-5}$
Constitutional	MW, AMW, Mv, Ms	0.719	0.311	14.75	$<10^{-5}$
Randić molecular profile	DP04, SP09, SP10, SP11	0.696	0.331	13.13	$<10^{-5}$
Geometrical	W3D, G1, SPH, FDI	0.728	0.313	15.334	$<10^{-5}$

Table 3
The validation parameters of the linear regression models obtained for the four kinds of descriptors involved in the comparison

Descriptors	Variables	q^2_{CV-LOO}	S_{CV-LOO}	q^2_{CV-LGO}	S_{CV-LGO}	AIC
Spectral moments	$\mu_1\mu_9^{GM}$, $\mu_1\mu_2^{Dip}$, $\mu_1\mu_{14}^{Std}$, $\mu_1\mu_{10}^{GM}$	0.754	0.298	0.741	0.322	0.072
Constitutional	MW, AMW, Mv, Ms	0.569	0.394	0.467	0.483	0.145
Randić molecular profile	DP04, SP09, SP10, SP11	0.603	0.378	0.471	0.474	0.157
Geometrical	W3D, G1, SPH, FDI	0.630	0.365	0.582	0.432	0.141

the following:

$$-\log(ED_{50}) = -18.86\mu_1\mu_9^{GM} + 1.75\mu_1\mu_2^{Dip} - 1.38\mu_1\mu_{14}^{Std} + 17.90\mu_1\mu_{10}^{GM} - 1.53 \quad (6)$$

$$N = 28, \quad R^2 = 0.861, \quad S = 0.223, \quad F = 35.71, \quad p < 10^{-5}, \quad AIC = 0.072$$

where N is the number of compounds included in the model, R^2 the square of the correlation coefficient, S the standard deviation of the regression, F the Fisher ratio and AIC is the Akaike's information criterion.

For the validation we calculate q^2 and the standard deviation S_{cv} of the LOO and LGO procedures and arrive at the following values:

$$q^2_{CV-LOO} = 0.754, \quad S_{CV-LOO} = 0.298, \quad q^2_{CV-LGO} = 0.741, \quad S_{CV-LGO} = 0.322$$

In spite of achieving adequate values of statistical parameters, we thought that it was not enough to say that our model was appropriate. Therefore, we carried out a comparison with other methodologies to demonstrate the superiority of our model.

Table 4
Descriptors of the QSARs regression reported in this study

Symbol	Definition
MW	Molecular weight
AMW	Average molecular weight
Mv	Mean atomic van der Waals volume (scaled on carbon atom)
Ms	Mean electrotopological state
DP04	Molecular profile no. 04
SP09	Shape profile no. 09
SP10	Shape profile no. 10
SP11	Shape profile no. 11
W3D	3D-Wiener index
G1	Gravitational index G1
SPH	Spherosity
FDI	Folding degree index

The results obtained from this comparison are given in Tables 2 and 3. In this connection, the meaning of the variables derived using DRAGON is given in Table 4. The meanings of the variables calculated from the TOPS MODE are explained above.

As we can see in a general way, the model proposed using the spectral moments methodology, shows a better statistical significance. That is to say, the spectral moments explain more of the variance of the data than the other ones using the same number of descriptors in the equation (four variables). The value of the squared regression coefficient for the TOPS-MODE approach is 0.861 while the other methodologies have values never higher than 0.730.

The spectral moments also present the best validation parameters, regarding the better predictive power. The parameter q^2, equivalent statistically to R^2, shows its greater value in this methodology for the two types of cross validation LOO and LGO (0.754 and 0.741, respectively). The rest of the methodologies have lower values of q^2 than the spectral moments for both cross validation techniques, being the second better methodology the one that uses the geometric descriptors with values of q^2 equal to 0.63 and 0.582 for these cross validation techniques.

When comparing the values of the Akike's information criterion (AIC) for each of the models we demonstrate once again the superiority of the spectral moments methodology (lowest value (AIC = 0.072)).

Collinearity among variables was avoided by making an orthogonalization of molecular descriptors because interrelatedness among different descriptors can result in highly unstable models.

The QSAR model obtained with the spectral moments (Eq. (6)) after orthogonalization and standardization is given below, together with the statistical parameters of regression analysis. The chart relative to the relationship between the

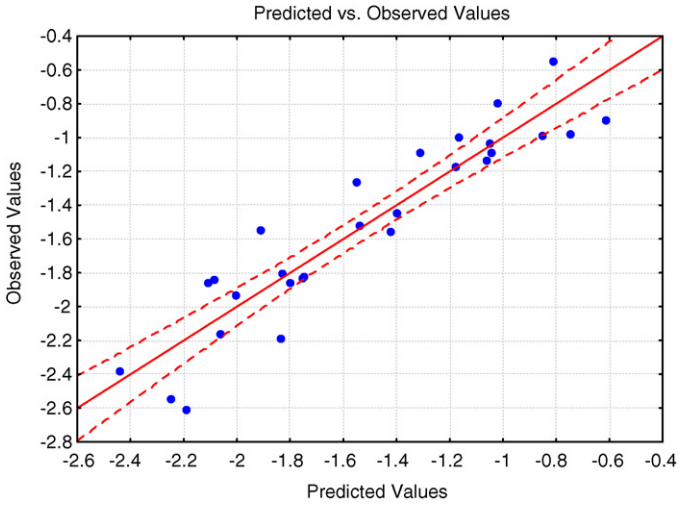


Chart 1. Relationship between the observed and predicted values according to Eq. (7).

observed and predicted values is provided too (Chart 1).

$$-\log(\mathrm{ED}_{50}) = -0.29\Omega^{1}\mu_1\mu_9^{\mathrm{GM}} + 0.325\Omega^{2}\mu_1\mu_2^{\mathrm{Dip}} - 1.11\Omega^{3}\mu_1\mu_{14}^{\mathrm{Std}} + 17.90\Omega^{4}\mu_1\mu_{10}^{\mathrm{GM}} - 1.53 \quad (7)$$

$$N = 28, \quad R^2 = 0.861, \quad S = 0.223, \quad F = 35.71,$$

$$p < 10^{-5}, \quad q^2_{\mathrm{CV-LOO}} = 0.754, \quad S_{\mathrm{CV-LOO}} = 0.298,$$

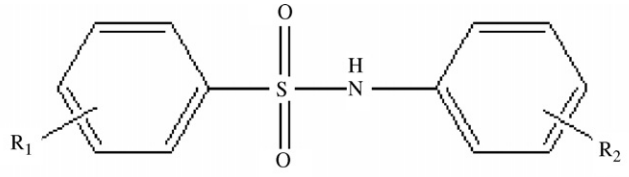
$$q^2_{\mathrm{CV-LGO}} = 0.741, \quad S_{\mathrm{CV-LGO}} = 0.322$$

The presence of outliers in QSAR models can to become in a serious problem due to that the model is unable to predict its "real" biological activity. In this connection, we looked for the presence of outliers in Eq. (7), where any outlier was found because of all the compounds included in the training set had a standard residual value lower than 2δ where δ is equivalent to the standard deviation and a relative small deleted residual, as have been see in Table 5.

Table 5
Predicted, residual and deleted residual according to Eq. (7)

Compounds	−log (ED_{50}) obs.	−log (ED_{50}) pred.	Residual	Deleted residual
1	−2.379	−2.442	0.063	0.097
2	−1.818	−1.750	−0.067	−0.077
3	−1.170	−1.181	0.010	0.012
4	−1.447	−1.402	−0.045	−0.052
5	−1.520	−1.538	0.018	0.020
6	−1.029	−1.054	0.025	0.032
7	−1.839	−2.088	0.249	0.313
8	−1.550	−1.422	−0.128	−0.151
9	−0.978	−0.750	−0.228	−0.271
10	−1.083	−1.047	−0.035	−0.045
11	−0.996	−1.167	0.172	0.188
12	−0.898	−0.614	−0.284	−0.372
13	−2.547	−2.250	−0.297	−0.353
14	−1.856	−2.109	0.253	0.290
15	−1.265	−1.552	0.287	0.318
16	−1.540	−1.912	0.371	0.420
17	−2.612	−2.191	−0.421	−0.782
18	−1.083	−1.316	0.233	0.262
19	−1.860	−1.801	−0.059	−0.066
20	−1.134	−1.066	−0.067	−0.074
21	−2.184	−1.835	−0.349	−0.505
22	−2.157	−2.062	−0.095	−0.131
23	−0.987	−0.855	−0.132	−0.152
24	−1.833	−1.757	−0.075	−0.086
25	−1.931	−2.008	0.077	0.093
26	−0.792	−1.020	0.227	0.251
27	−1.803	−1.833	0.029	0.045
28	−0.544	−0.811	0.267	0.312

The analysis of this equation reflects the influence of the variables in the fungal activity in question. In a general way we can say that the electronic factors (variables weighted with



Groups	R_1
I	H
II	*p*-CH_3
III	*o*-NO_2
IV	*m*-NO_2
V	*p*-NO_2

R_2
H
p-NO_2
p-CF_3
o-Cl. *p*-NO_2
o-CF_3. *p*-NO_2
m-CF_3. *p*-F

Fig. 1. Basic structure of the family of compounds analyzed.

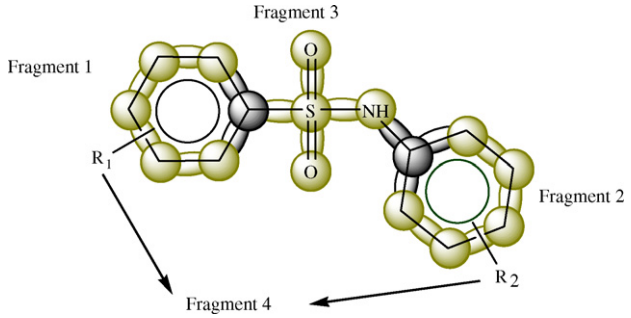


Fig. 2. The fragments employed in the analysis (shown in green).

bond dipole moment ($\Omega^2\mu_1\mu_2^{\mathrm{Dip}}$) and Gasteiger–Marsili charge ($\Omega^1\mu_1\mu_9^{\mathrm{GM}}$, $\Omega^4\mu_1\mu_{10}^{\mathrm{GM}}$)), have a positive influence and, when weighted with bond distance ($\Omega^3\mu_1\mu_{14}^{\mathrm{Std}}$), the opposite effect.

More specifically the Gasteiger–Marsili charge explains 39.8% of the variance, bond dipole moment 29.4% and the bond distance 16.9% for a total of 86.1%.

Among the data we can observe quite generally how the presence of substituents is a key function to focus on in the search for compounds with antifungal activity.

3.1. Analysis of fragment contributions

As a way to gain a better understanding of the influence of variables and the contribution of the different substituents, we carried out an analysis of molecules fragments involved in the development of the predictive model.

The basic structure employed in our analysis is shown in Fig. 1.

The compounds for every group have the same substituent R_1 and different substituents R_2 which have been shown previously in Fig. 1. The calculations of the fragments were focused towards the following principal fragments (see Fig. 2):

1. The fragment containing the aromatic ring linked to S.
2. The fragment related to the aromatic ring linked to N.
3. The sulfonamide group.
4. The different substituents of both aromatic rings.

In the analysis of the first group of compounds we found that the contribution of the aromatic ring linked to N it is not significant when the *para* position in this ring is substituted. The contribution is slightly lower when there is another substituent in the same ring.

For the other ring linked to S, the contribution takes a value less negative than the ring previously analyzed and when the ring is linked to N it has a second substituent and the contribution of the ring is modularly lower.

The substituents in the ring are electron withdrawing and contribute in a positive way. The NO_2 group, in particular, increases the contribution when there are other electron withdrawing substituents in the ring. Further, it is important to note that when the second substituent is more electron withdrawing, the contribution of NO_2 increases to an even greater degree (see Fig. 3).

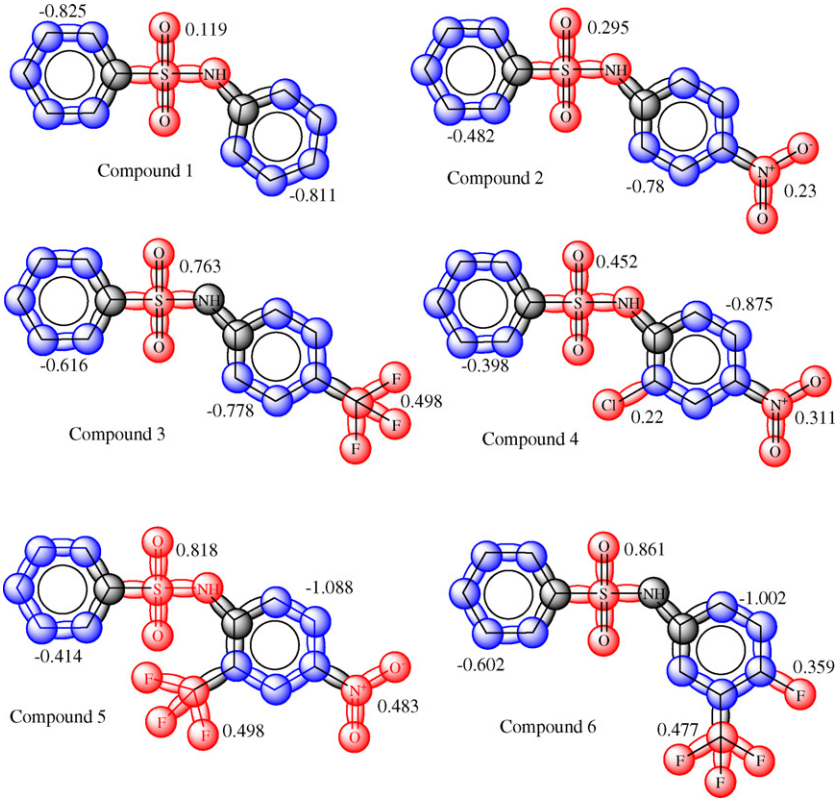


Fig. 3. The structures of group number I. The blue fragments have a negative contribution, those in red a positive one.

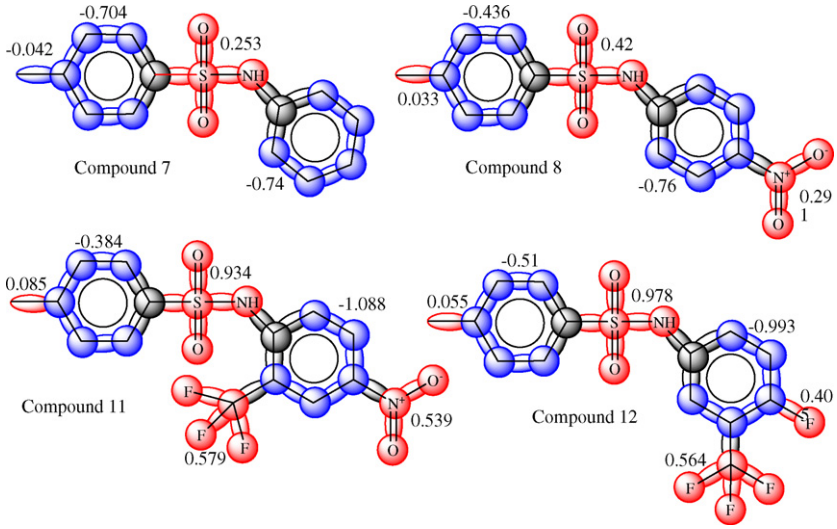


Fig. 4. Contribution of fragments of compounds belonging to the second group (the fragments in blue have a negative contribution and the fragments in red have a positive contribution).

These tendencies remain in the second group. The contribution is even less negative in the aromatic ring linked to S, when the electron donor substituent CH_3 is linked to this ring in the compounds in the second group, as compared to the compounds analyzed in the previous group (see Fig. 4).

If we compare compounds **2**, **19** and **24** that belong to the groups I, IV and V, respectively, we can see an interesting effect of the substituents in the aromatic rings.

Wherever NO_2 is present in either of the aromatic rings, the other one increase its contribution, that is to say, the contribution becomes less negative than if the substituent were not present (see Fig. 5).

Compounds **20** and **21** have a similar contribution. Worthy of note is the small difference in the contribution between these compounds in the aromatic rings and the presence of electron withdrawing substituents in both rings of these compounds. There is no significant difference between them (see Fig. 6).

We can make tentative conclusions about the contribution of the group NO_2:

It is possible that the NO_2 group undergoes reduction to oxime and later to amine and participates in this way in the process of inhibition of the growth fungal. This reduction might be easier in presence of the other electron withdrawing substituents (*o*-Cl and/or *o*-CF_3) in the same aromatic ring, because they leave lacking electrons to the ring and in this way the NO_2 will have fewer electrons at its disposal therefore this group will accept the electrons easier and to be reduced.

There are four interesting observations regarding the contribution of the aromatic rings to the biological property:

- *When the aromatics rings are unsubstituted* the contribution is similar in both rings, for example in the compound number **1**.
- *When only one aromatic ring has electron withdrawing substituents* the other ring makes a significantly less negative contribution. For example, the rest of the compounds of group number I and the compounds **13**, **19** and **24**.
- *When one aromatic ring has electron withdrawing substituents and the other has an electrodonor substituent such*

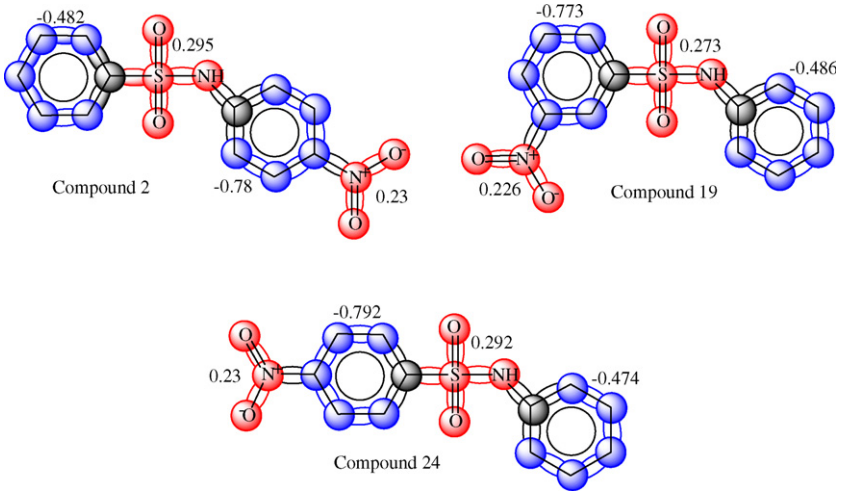


Fig. 5. Comparison between compounds **2**, **19** and **24** belonging to the groups I, IV and V, respectively.

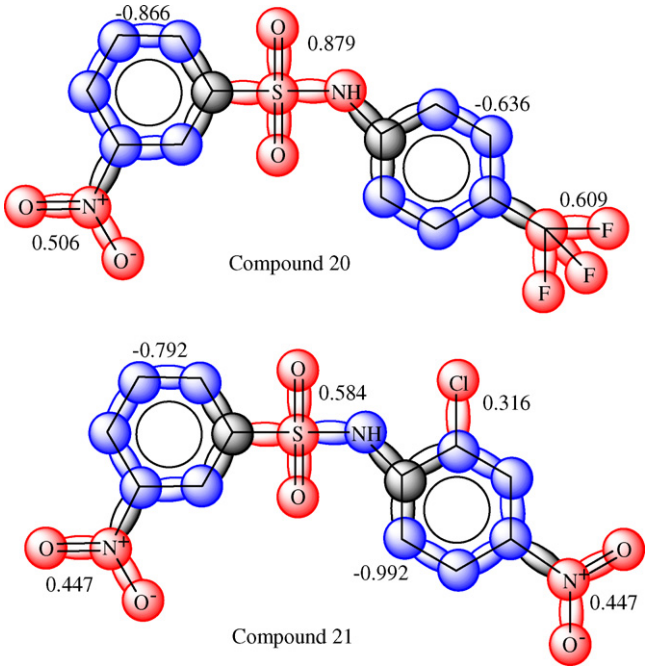


Fig. 6. Effect of the substituents in both aromatics rings.

as CH_3 the increase in the contribution of the ring with the electrodonor substituent is enhanced. For example, the compounds of the group number II.

- *When both rings contain electron withdrawing substituents* there is not significant difference between the aromatic rings regarding their contribution. For example, in the compounds of groups IV and V.

From these four observations we can derive an important conclusion about the contribution of the individual rings. As the deactivation increases in a ring, the contribution in the other one increases. This may be due to the fact that the second ring receives a higher share of the electronic density and this contributes in some way to an interaction with the receptor.

On the other hand, we found interesting facts in the comparison between compounds of the groups III, IV and V. As specific examples we take the compounds **13**, **19** and **24** because the tendencies explained above are also to be seen in them.

In these specific compounds the contribution of the aromatic ring linked to N is according with our previous explanation, less negative. In addiction, it is interesting the changes in the contribution of this ring when in the other aromatic ring the nitro group is in *meta* or *para* position respect to the SO_2 moiety. For these ones (**19** and **24**, see Fig. 7) the contribution of the aromatic ring linked to N atom is less negative than in the compound with nitro NO_2 in the *ortho* position (compound **13**, see Fig. 7). For the remainder of the compounds in the three groups (III, IV and V) this tendency remains.

However, in compounds in which the NO_2 is in *ortho* position with respect to the S atom (group III), we can say that the contribution of the ring linked to S is always the less negative. This is accentuated for the rest of compounds of the group III compared to other groups, when the ring linked to N has electron withdrawing substituents.

The contribution of especially the NO_2 group is similar among all compounds of the groups III, IV and V although we can see a slight decrease when this substituent is in an *ortho* position with respect to the S atom. All these observations are reflected in Fig. 7.

This behavior is seen for the *ortho* effect observed in the structures of the group III while in the compounds of the other groups involved in the comparison this effect is not present.

This effect might involve the NO_2 group in an interaction with the SO_2NH making its reduction more difficult, and resulting in it having the smallest contribution.

About the contribution of aromatic rings linked to S atom we observe that when the NO_2 is in the *ortho* position, it will be out of the plane of the ring. Therefore, this substituent cannot exert such a strong attraction on the electrons as when it is in *meta* and *para* position and the electronic density of this ring will be

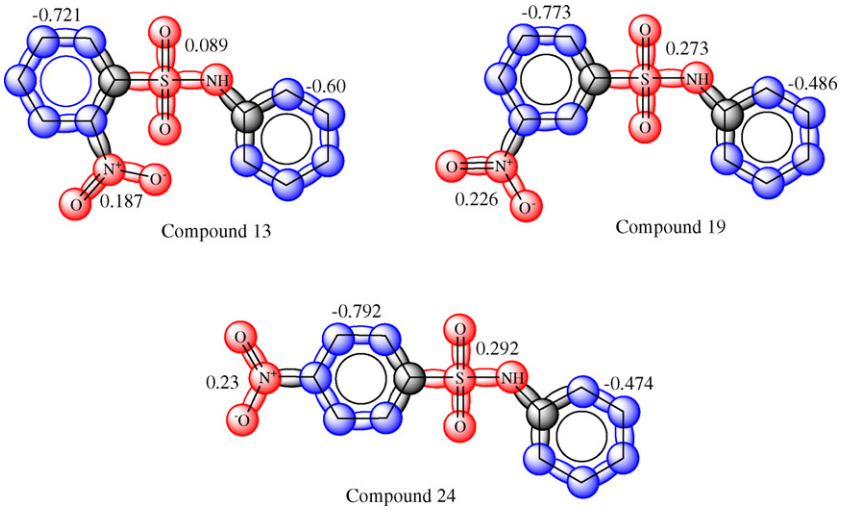


Fig. 7. Comparison among compounds **13**, **19** and **24** belonging to the groups III, IV and V, respectively.

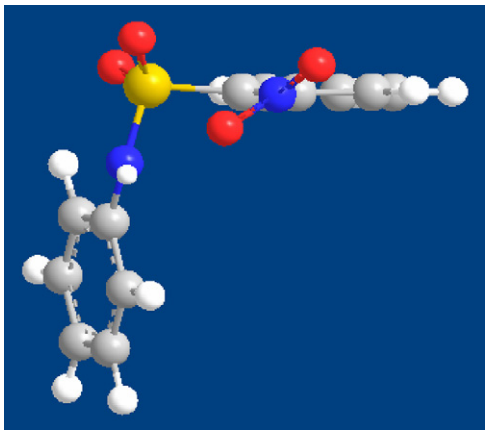


Fig. 8. Compound **13** (*o*-NO_2).

not be so strongly affected. As a result, the contribution of the ring linked to S atom, for the compounds with NO_2 in the *ortho* position will be the less negative one coinciding with the explanation given above. In the following figures we can observe the *ortho* effect in the compound **13**. In the other compounds, **19** and **24** the NO_2 group is co-planar with the aromatic ring.

Another very important question is the contribution of the SO_2NH group. The sulfonamides have a very well-known chemotherapeutic bactericidal activity. They compete with

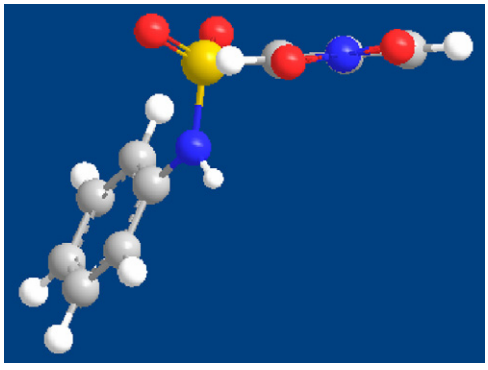


Fig. 9. Compound **19** (*m*-NO_2).

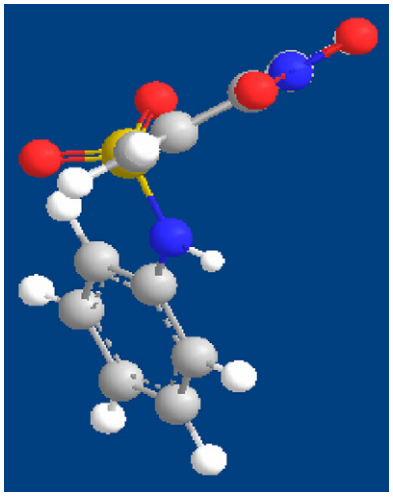


Fig. 10. Compound **24** (*p*-NO_2).

p-aminobenzoic acid to be linked to the enzyme dihydropteroate synthase, inhibiting the synthesis of folic acid and the synthesis of nitrogen bases [14–16]. However, the sulfonamides present antifungal activity, too [17–19].

The contribution of sulfonamides fragment is the most important in these molecules. It is more positive when there are electron withdrawing substituents in the aromatic rings. However in the case of compounds of group III, with NO_2 in the *ortho* position respect to the S atom we can observe how the *ortho* effect affects the contribution of the sulfonamide fragment taking its lowest value. In this sense it highlights how the sulfonamide group is out of the plane with respect to the aromatic ring when the *ortho* effect is present (see Fig. 8 and compare with Figs. 9 and 10).

4. Conclusions

In this work, we have achieved some conclusions enabling a better understanding of the interpretation of the predictive model developed. However, as general conclusions of the work we can say that a model of multiple regression analysis (MRA) was developed by using the TOPS-MODE approach. The model presented appropriate statistical quality as endorsed by their adjustment parameters and validation. The methodology proposed by us was superior to other descriptors such as constitutional descriptors, Randić molecular profile and geometrical descriptors taking into account the statistical parameters of the model and of the cross validation developed.

As for the analysis of the bond contributions, we conclude that the sulfonamide group makes the largest contribution to the antifungal effect. Most of the substituents present in the aromatic rings are electron withdrawing and they make a smaller contribution than the sulfonamide group to the property under study. In addition, the nitro group could be an active center because it may undergo reduction. Generally, the nitro group has a positive and great contribution to the inhibition of growth fungal. Nevertheless, when this group is in some compounds in *ortho* position regarding the SO_2 moiety of the sulfonamide group a lower value of contribution is observed for both groups. This can be given by the presence of the *ortho* effect.

The fragments comprising aromatic rings make negative contributions.

In spite of the presence of a distortion of planarity at the sulfonamide group, a change can be discerned in the contribution of one of these rings. This probably results from the presence of electron withdrawing substituents in the neighboring ring.

Acknowledgement

The authors acknowledge the owners of the Software MODESLAB for their kind donation of this tool.

References

[1] H.J. Rosslenbroich, D. Stuebler, *Botrytis cinerea*-history of chemical control and novel fungicides for its management, Crop Prot. 19 (2000) 557–561.

[2] P. Aplada-sarlis, K.S. Liapis, G.E. Miliadis, Study of procymidone and propargite residue levels resulting from application to greenhouse tomatoes, J. Agric. Food Chem. 42 (1994) 1575–1577.
[3] S. Paris-Palacios, S. Biagianti-Risbourg, G. Vernet, Hepatic disorders induced by a sublethal concentration of procymidone (fungicide dicarboximide) in Brachydanio rerio(Teleostei Cyprinidae): histo-cytological approach, Bull. Fr. Peche Piscic. 350 (1998) 547–562.
[4] C. Sala, F. Fort, O. Busto, F. Zamora, L. Arola, J. Guasch, Fate of some common pesticides during vinification process, J. Agric. Food Chem. 44 (1996) 3668–3671.
[5] M.P. González, L.C. Dias, A.M. Helguera, Y.M. Rodriguez, L.G. de Oliveira, L.T. Gomez, H.G. Diaz, TOPS-MODE based QSARs derived from heterogeneous series of compounds. Applications to the design of new anti-inflammatory compounds, Bioorg. Med. Chem. 12 (2004) 4467–4475.
[6] M.P. González, H. Gonzalez Diaz, R. Molina Ruiz, M.A. Cabrera, R. Ramos de Armas, TOPS-MODE based QSARs derived from heterogeneous series of compounds. Applications to the design of new herbicides, J. Chem. Inf. Comput. Sci. 43 (2003) 1192–1199.
[7] H. Gonzalez-Diaz, I. Bastida, N. Castañedo, O. Nasco, E. Olazabal, A. Morales, H.S. Serrano, R.R. de Armas, Simple stochastic fingerprints towards mathematical modelling in biology and medicine. 1. The treatment of coccidiosis, Bull. Math. Biol. 66 (2004) 1285–1311.
[8] H. Gonzalez-Diaz, O. Gia, E. Uriarte, I. Hernadez, R. Ramos, M. Chaviano, S. Seijo, J.A. Castillo, L. Morales, L. Santana, et al., Markovian chemicals "in silico" design (MARCH-INSIDE), a promising approach for computer-aided molecular design. I. Discovery of anticancer compounds, J. Mol. Model (Online) 9 (2003) 395–407.
[9] M.P. González, J. Caballero, A. Tundidor-Camba, A.M. Helguera, M. Fernandez, Modeling of farnesyltransferase inhibition by some thiol and non-thiol peptidomimetic inhibitors using genetic neural networks and RDF approaches, Bioorg. Med. Chem. 14 (2006) 200–213.
[10] T. Akagi, S. Mitani, T. Komyoji, K. Nagatani, Quantitative structure–activity relationships of fluazinam and related fungicidal *N*-phenylpyridinamines. Preventive activity against *Botrytis cinerea*, J. Pestic. Sci. 20 (1995) 279–290.
[11] T. Akagi, S. Mitani, K. Ito, I. Shigehara, T. Komyoji, N. Matsuo, Structure–activity relationships of pyridylcarbamates active against both benzimidazole-sensitive and -resistant isolates of *Botrytis cinerea*, Pestic. Sci. 44 (1995) 39–48.
[12] E. Humeres, G.A. Cantos, N.A. Debacher, R.J. Nunes, Hydrophobicity constant of the dithiocarbamic fragment and QSAR of dithiocarbamate fungicides, Org. React. 31 (1997) 45–49.
[13] T. Kataoka, QSAR of 1-*N*-substituted azoles active against *Botrytis cinerea*, in: W. Draber, T. Fujita (Eds.), In Rational Approaches to Structure, Activity, and Ecotoxicology of Agrochemicals, CRC, 1992, pp. 465–482.
[14] P.G. De Benedetti, in: E. Jucker (Ed.), Progress in Drug Research, Birkhäuser Verlag, Basel, 1991.
[15] F. Zani, P. Vicini, Antimicrobial activity of some 1,2-benzisothiazoles having a benzenesulfonamide moiety, Arch. Pharm. Pharm. Med. Chem. 331 (1998) 219–223.
[16] G.L. Mandell, W.A. Petri Jr., in: J.G. Hardman (Ed.), The Pharmacological Basis of Therapeutics, Mcgraw-Hill, New York, 1995.
[17] K.K. Kim, J.G. Kang, S.S. Moon, K.Y. Kang, Isolation and identification of antifungal *N*-butylbenzenesulphonamide produced by Pseudomonas sp. AB2, J. Antibiot. (Tokyo) 53 (2000) 131–136.
[18] C.R. Worthing, R.J. Hance, The Pesticide Manual, British Crop Protection Council, UK, 1991.
[19] H. Shimotori, H. Yanagida, Y. Enomoto, K. Igarashi, M. Yoshinari, M. Umemoto, Evaluation of benzenesulfonanilide derivates for the control of crucifers clubroot, J. Pestic. Sci. 21 (1996) 31–35.
[20] E. Estrada, Spectral moments of the edge-adjacency matrix of molecular graphs. 2. Molecules containing heteroatoms and QSAR applications, J. Chem. Inf. Comput. Sci. 37 (1997) 320–328.
[21] E. Estrada, Spectral moments of the edge adjacency matrix in molecular graphs. 1. Definition and applications to the prediction of physical properties of Alkanes†, J. Chem. Inf. Comput. Sci. 36 (1996) 844–849.
[22] E. Estrada, E. Uriarte, Y. Gutierrez, H. González, Quantitative structure-toxicity relationships using TOPS-MODE. 3. Structural factors influencing the permeability of commercial solvents through living human skin, SAR QSAR Environ. Res. 14 (2003) 145–163.
[23] Y. Gutierrez, E. Estrada, MODESLAB 1.0 (Molecular Descriptors Laboratory) for Windows, 2002.
[24] J.G. Kang, J.H. Hur, S.J. Choi, G.J. Choi, K.Y. Cho, L.N. Ten, K.H. Park, K.Y. Kang, Antifungal activities of *N*-arylbenzenesulfonamides against phytopathogens and control efficacy on wheat leaf rust and cabbage club root diseases, Biosci. Biotechnol. Biochem. 66 (2002) 2677–2682.
[25] Statsoft Inc., STATISTICA (data analysis software system), edn version 6.0, edited by; 2002.
[26] A. Vedani, M. Dobler, Multi-dimensional QSAR in drug research. Predicting binding affinities, toxicity and pharmacokinetic parameters, Prog. Drug Res. 55 (2000) 105–135.
[27] A. Tropsha, W. Zheng, Identification of the descriptor pharmacophores using variable selection QSAR: applications to database mining, Curr. Pharm. Des. 7 (2001) 599–612.
[28] K. Hasegawa, K. Funatsu, Partial least squares modeling and genetic algorithm optimization in quantitative structure–activity relationships, SAR QSAR Environ. Res. 11 (2000) 189–209.
[29] D.M. Hawkins, The problem of overfitting, J. Chem. Inf. Comput. Sci. 44 (2004) 1–12.
[30] D.J. Klein, M. Randić, D. Babić, B. Lučić, S. Nikolić, N. Trinajstić, Hierarchical orthogonalization of descriptors, Int. J. Quantum Chem. 63 (1991) 215–222.
[31] M. Randić, Correlation of enthalphy of octanes with orthogonal connectivity indices, J. Mol. Struct. (Teochem.) 233 (1991) 45–59.
[32] M. Randić, Orthogonal molecular descriptors, New J. Chem. 15 (1991) 517–525.
[33] M. Randić, Resolution of ambiguities in structure-property studies by use of orthogonal descriptors, J. Chem. Inf. Comput. Sci. 31 (1991) 311–320.
[34] B. Lučić, S. Nikolić, N. Trinajstić, D. Jurić, The structure-property models can be improved using the orthogonalized descriptors, J. Chem. Inf. Comput. Sci. 35 (1995) 532–538.
[35] R. Todeschini, V. Consonni, M. Pavan, Dragon Software, edn version 2.1, edited by; 2002.
[36] J. Frank, MOPAC, edn version 6.0, edited by: Seiler Research Laboratory, US Air Force Academy, Colorado Springs, CO, 1993.
[37] H. Akaike, Information theory and an extension of the maximum likelihood principle, in: B.N. Petrov, F. Csaki (Eds.), In 2nd International Symposium on Information Theory; Budapest, Akademiai Kiado, (1973), pp. 267–281.
[38] H. Akaike, A new look at the statistical model identification, IEEE Trans. Automat. Contr. AC-19 (1974) 713–716.
[39] E. Estrada, G. Patlewicz, Y. Gutierrez, From knowledge generation to knowledge archive. A general strategy using TOPS-MODE with DEREK to formulate new alerts for skin sensitisation, J. Chem. Inf. Comput. Sci. 44 (2004) 688–698.
[40] M.P. González, C. Teran, M. Teijeira, P. Besada, M.J. González-Moa, BCUT descriptors for predicting affinity toward A_3 adenosine receptors, Bioorg. Med. Chem. Lett. 15 (2005) 3491–3495.