Quantitative Structure–Antifungal Activity Relationships for cinnamate derivatives

Laura M. Saavedra [a], Diego Ruiz [b], Gustavo P. Romanelli [b,c], Pablo R. Duchowicz [a,*]

[a] Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA (UNLP, CCT La Plata-CONICET), Diagonal 113 y 64, Sucursal 4, C.C. 16, 1900 La Plata, Argentina
[b] Curso de Química Orgánica, Facultad de Ciencias Agrarias y Forestales, Universidad Nacional de La Plata (UNLP), 60 y 119, B1904AAN La Plata, Buenos Aires, Argentina
[c] Centro de Investigación y Desarrollo en Ciencias Aplicadas "Dr. J.J. Ronco" (CINDECA), Departamento de Química, Facultad de Ciencias Exactas, UNLP-CCT-CONICET, Calle 47 No. 257, B1900AJK La Plata, Argentina

ARTICLE INFO

ABSTRACT

Quantitative Structure–Activity Relationships (QSAR) are established with the aim of analyzing the fungicidal activities of a set of 27 active cinnamate derivatives. The exploration of more than a thousand of constitutional, topological, geometrical and electronic molecular descriptors, which are calculated with Dragon software, leads to predictions of the growth inhibition on *Pythium sp* and *Corticium rolfsii* fungi species, in close agreement to the experimental values extracted from the literature. A set containing 21 new structurally related cinnamate compounds is prepared. The developed QSAR models are applied to predict the unknown fungicidal activity of this set, showing that cinnamates like **38**, **28** and **42** are expected to be highly active for *Pythium sp*, while this is also predicted for **28** and **34** in *C. rolfsii*.

© 2015 Published by Elsevier Inc.

1. Introduction

The cinnamic acid esters or cinnamates are a family of aromatic esters with a structure as shown in Fig. 1.

There are several procedures that have been used to prepare cinnamates, via cinnamoyl chloride (Womack and McWhirter, 1955), using N,N'-dicyclohexylcarbodiimide (DCC) (Isaacs and Najem, 1988) as coupling agents to perform the direct reaction between cinnamic acid and a alcohol, or between a phosphorane and a benzaldehyde (Mali and Papalkar, 2003). Most recently, our research group has described two green procedure methods to prepare these compounds for the direct esterification of cinnamic acids with alcohol or phenol using heteropolyacids in bulk or silica-supported form (Palermo et al., 2012; Romanelli et al., 2010a; Ruiz et al., 2008).

Cinnamic acids and the corresponding cinnamates (α, β unsaturated esters) are natural compounds belonging to the family of the phenylpropanoids. The most important natural cinnamates are derived from the following cinnamic acids such as ferulic acid (4-hydroxy-3-methoxycinnamic acid), cumaric acid (p-hydroxycinnamic acid), caffeic acid (3,4-dihydroxycinnamic acid) and sinapic acid (4-hydroxy-3,5-dimethoxycinnamic). In general, they are bioactive compounds and exhibit varied and specific properties (Clifford, 2000).

Cinnamates are important organic compounds owing to their application in a wide range of industrial products such as graphics, lubricants, and plasticizers, particularly alkyl cinnamates are used in different areas, as perfume essence, soap and flavoring essence due to pronouncing fruit or flower aromas (Shu and Hongjun, 2013; Sinha et al., 2007). In fact, cinnamates with diverse structures posse a variety of biological properties such as: antioxidant, antimicrobial, anticancer, anti-inflammatory, and antithrombotic, among others (Ruiz, 2009).

Fig. 1 also shows the structure of three relevant cinnamates: the rosmarinic acid, a potential anxiolytic which acts as a GABA transaminase inhibitor, more specifically on 4-aminobutyrate transaminase (Awad et al., 2009); the imidoalkyl cinnamate, 2-(N-β-phthalimido) ethyl cinnamate, a potent inhibitor of the 17-β-hydroxysteroid dehydrogenase enzyme related to diseases such as breast cancer, Alzheimer disease and benign prostatic hyperplasia (Gobec et al., 2004; Kristan et al., 2006); and the chlorogenic acid, which has antihypertensive effects (Zhao et al., 2011) and has been reported to be a chemical sensitizer responsible for human respiratory allergy to certain types of plant materials (Freedman et al., 1964).

Several aryl cinnamates have been also used as intermediates for the synthesis of diverse heterocyclic compounds, such as flavanones (Moghaddam et al., 1999), chromones (Pinto et al., 1999),
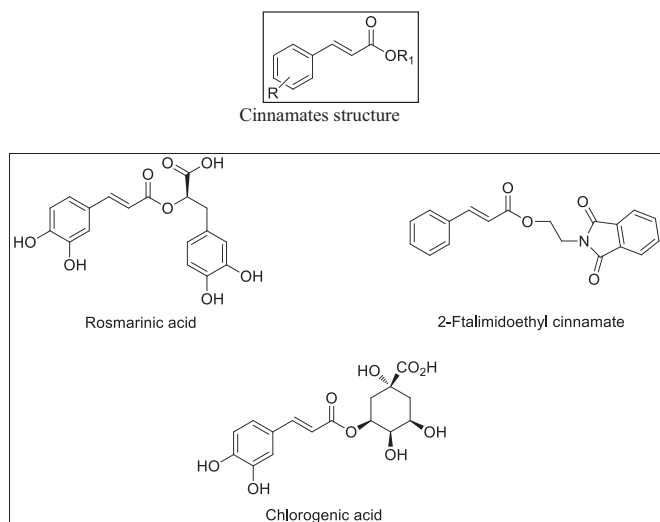
Fig. 1. Cinnamates structure and relevant biologically active cinnamates.

pyrazols (Li et al., 2005), dihydrocumarins (Aoki et al., 2005) and benzofuranones (Shankaran et al., 1985).

Cinnamates, whether natural or synthetic, have been reported with biological activity: repellence, antifeedant and insecticide activities and have attracted the attention of many researchers. Specifically, methyl cinnamate has insecticide effect against adults of *Sitophylus oryzae* (Coleoptera: Curculionidae) (Lee et al., 2008) and *Musca domestica* (Diptera: Muscidae) (Peterson et al., 2000); ethyl cinnamate has antifeedant effect against *Spodoptera littoralis* (Lepidoptera: Noctuidae) (Abdelgaleil et al., 2008) and *Hylobius abietis* (Coleoptera: Curculionidae) (Sunnerheim et al., 2007); and propyl cinnamate, shows insecticidal effect on *M. domestica* adults (Peterson et al., 2000). For this reason, the cinnamates represent a good alternative to replace the conventional insecticides and behave as potential green insecticides.

In this work, we study the antifungal activity of some cinnamate compounds having various substitution patterns along with some of their derived products against *Phytium sp.* and *Corticium rolfsii.* Pythium is a genus of parasitic oomycetes, while Corticium is a type of agaricomycetes. Several species are plant parasites, and they have formerly been classified as fungi. Pythium-induced root rot is a common crop disease and *C. rolfsii* is the causal agent of southern blight disease. When the organism kills newly emerged or emerging seedlings it is known as damping off, and is a common problem in fields and greenhouses (Jarvis, 1992).

The fundamental basis of the Quantitative Structure–Activity Relationships (QSAR) Theory (Hansch and Fujita, 1995; Hansch and Leo, 1995; Puzyn et al., 2010) is the hypothesis that the biological activity exhibited by a chemical compound is completely determined by its molecular structure. Within this framework, it is possible to get some insight into the complex underlying mechanism of action that leads to the final biological effect, by means of the predicted activities. There exist advanced QSAR models, which use other very known and popular statistical methods such as Linear Discriminant Analysis, and they can simultaneously predict multiple activities or toxicities of different classes of pesticides. Such models have a great performance from statistical and phenomenological points of view (Speck-Planche et al., 2012a, 2012b, 2011a, 2011b, 2011c). Here, multi-parametric linear regression models are established for modeling the antifungal activity, as this is considered one of the most popular statistical techniques, and explore more than a thousand theoretical descriptors (Duchowicz et al., 2009, 2007; Pasquale et al., 2012; Romanelli et al., 2010b).

We collect from the literature the 27 experimental antifungal activity of cinnamate derivatives (Tawata et al., 1996), expressed as the growth inhibition (*GI*, 100 ppm) of the *Pythium sp* and *C. rolfsii* fungi species (Table 1S), and establish useful QSAR models. The main objective of this research is to apply the developed structure–activity relationships to predict a set of 21 alkyl and aryl cinnamates derivatives, with unknown experimental fungicidal activities and recently synthesized in our laboratory.

2. Materials and methods

2.1. General procedures

All the reactions are monitored by thin layer chromatography (TLC) on precoated silica gel plates (254 mm). Flash column chromatography is performed with 230–400 mesh silica gel. All the yields are calculated from pure products. All the products are identified by comparison of physical data (mp, TLC, NMR) with those reported ones. Melting points of the compounds are determined in sealed capillary tubes and are uncorrected. ^{13}C NMR and ^{1}H NMR spectra are recorded at room temperature on Bruker AC-250 and Bruker Avance DPX-400 spectrometers using TMS as internal standard.

2.2. Catalysis preparation

The Preyssler salt, $K_{12.5}Na_{1.5}[NaP_5W_{30}O_{110}] \cdot 15H_2O$ (PWK), is prepared from $Na_2WO_4 \cdot 2H_2O$ according to a previously reported method (Creaser et al., 1993), and converted to the corresponding acid $H_{14}[NaP_5W_{30}O_{110}]$ (PW), by passing it through a Dowex-50Wx8 ion-exchange column.

2.3. General procedure for the synthesis of phenyl cinnamates

The esterification reaction is performed in a round-bottom flask, which is equipped with a condenser and immersed in an oil bath. A solution of cinnamic acid **1** (5 mmol) and phenol **2** (5 mmol) in toluene (20 mL) and the bulk HPA (5×10^{-2} mmol) are refluxed with stirring for the indicated time. In both cases, the catalyst is removed by filtration and washed twice with toluene (3 mL each). The organic solution is washed with cold 1 M NaOH (2×10 mL) and H_2O (2×10 mL) and then dried over anh. Na_2SO_4. Evaporation of the solvent under reduced pressure and silica flash column chromatography gives the pure cinnamate **3**.

2.4. Quantitative Structure–Activity Relationships

The initial conformations of the compounds are drawn by means of the "model build" modulus available in HyperChem 6.03 (HyperChem). Each molecular structure is first preoptimized with the molecular mechanics force field (MM+) procedure, and the resulting geometry is further refined by means of the Semi-empirical Method PM3 (Parametric Method-3). We choose a gradient norm limit of 0.01 kcal mol^{-1} Å^{-1}. The numerical descriptors for each compound are calculated with Dragon software (E-Dragon) and include several variable types characterizing the multidimensional aspects of the chemical 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 descriptors, 2D autocorrelations, aromaticity indices, Randic molecular profiles, radial distribution functions, functional groups, and atom-centered fragments. We also add quantum-chemical descriptors to the pool such as the highest

occupied (HOMO) and lowest unoccupied (LUMO) molecular orbital energies, HOMO–LUMO gap (ΔHOMO $-$ LUMO), total dipole moment, and the number of electrons. The total number of calculated descriptors is 1502.

We resort to the Replacement Method (RM) (Duchowicz et al., 2006) as molecular descriptor selection approach, an algorithm that has been proposed by our research group some years ago for generating multivariable linear regression QSAR models with minimized error, measured through the standard deviation (S) or the root mean square deviation ($RMSD$). 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, it requires much less computational work (Duchowicz et al., 2006). In addition, the RM provides models with statistical parameters better than those from the forward stepwise regression procedure and similar to those from the more elaborated genetic algorithms approach (Mercader et al., 2010, 2011).

The most efficient validation strategy for any QSAR model consists on using an external test set of molecules, never seen by the relationship during the calibration of its parameter values. For this purpose, it is used a number of $N_{train} = 23$ training set cinnamate derivatives and $N_{test} = 4$ test set cinnamate compounds (15% from 27 molecules). This partition is chosen for each bioactivity in such a way that both sets share similar qualitative structure–activity characteristics, as the training set should be representative of the molecular diversity of all the compounds under study and uniformly span over the whole activity range.

The Cross-Validation technique of Leave-One-Out (loo) is practiced (Golbraikh and Tropsha, 2002). The parameters R^2_{loo} and S_{loo} (squared correlation coefficient and standard deviation of Leave-One-Out) measure the stability of the model upon exclusion of molecules. According to the literature, R^2_{loo} should be greater than 0.5 for a validated model. The Y-Randomization procedure (Rücker et al., 2007) is also applied in order to verify that the model is robust. This technique consists on scrambling the experimental property values in such a way that they do not correspond to the respective compounds. After analyzing 200,000 cases of Y-Randomization, the standard deviation obtained (S^{Rand}) has to be a poorer value than that of the true calibration (S).

The applicability domain (AD) for the QSAR models is also explored, as not even a predictive model is expected to reliably predict the modeled activity for the whole universe of molecules. The AD is a theoretically defined area that depends on the descriptors and the experimental activity (Gramatica, 2007). Only the molecules falling within this AD are not considered model extrapolations. One possible way to characterize the AD is based on the leverage approach, which allows to verify whether a new compound can be considered as interpolated (with reduced uncertainty, reliable prediction) or extrapolated outside the domain (unreliable prediction). Each compound i has a calculated leverage value (h_i) and there exists a warning leverage value (h^*); Table 2S includes the definitions for h_i and h^*. When $h_i > h^*$ or close to h^* for a test set compound, then a warning should be given: it means that the prediction is the result of substantial extrapolation of the model and cannot be treated as reliable. When $h_i > h^*$ or close to h^* for a training set compound having a low residual value, then it means that such compound would reinforce the model and it is considered a 'good high leverage compound'.

In order to find out the relative importance of the j-th descriptor in the linear model, the regression coefficients are standardized (b^s_j, see Table 2S). The larger is the absolute value of b^s_j, the greater is the importance of such descriptor (Draper and Smith, 1981).

3. Results and discussion

3.1. Synthesis of new cinnamate derivatives

The set of 21 cinnamates derivatives is prepared (**28–48**) by the esterification reaction of cinnamic acid with the corresponding alcohols and phenols using toluene as solvent and PW as catalyst. The reactivity of different cinnamic acids and alcohol or phenols is tested under the same conditions (110 °C, cinnamic acid/alcohol or phenol ratio 1:1, catalyst 0.5 mmol%, and a reaction time of 7–18 h.). Results of the obtained yields are listed in Table 1. The results show that, in general, the reactions are clean and products are isolated by flash column chromatography in pure form without further purification (^{1}H and ^{13}C NMR).
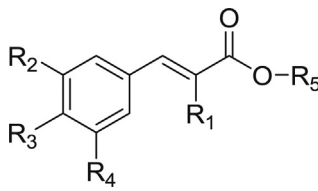
3.2. QSAR Study

The QSAR analysis is carried out by searching for various predictive linear relationships for the antifungal activities of the cinnamate structures. For modeling purposes, we first convert the experimental GI values into $\log_{10}(GI + 50)$. The optimal linear regression equations are established by means of the RM approach, which minimizes their standard deviation (S) and includes the best "representative" $d = 1 - 4$ molecular descriptors. The size of such linear regressions (d) has to obey the "rule of thumb", which states that at least six data points should be present per fitting parameter, in order to reduce overfitting problems ($N_{train}/d \geq 6$) (Tute, 1971).

3.2.1. QSAR for the growth inhibition of Pythium sp

Table 2 summarizes the main statistical results found for this fungi species, revealing that a three-descriptors model (QSAR-3) works better for the bioactivity, having a ratio $N_{train}/d = 7.7$ that accomplishes with the rule of thumb, while leading to an acceptable statistical quality for the training and test sets. Fig. 2A plots the predicted antifungal activities as function of the observed ones revealing that a straight line trend is established; Fig. 2B reveals that the residuals are randomly distributed and do not follow any kind of strange pattern which would indicate the presence of non-modeled factors.

It is appreciated from Fig. 2B that compound **22** has a residual value close to the 2S limit. The removal of such compound would allow to improve the calibration statistics for the QSAR-3 model (refer to Table 2) from $R_{train} = 0.84$, $S_{train} = 0.086$ to $R_{train} = 0.88$, $S_{train} = 0.078$. We believe that this cinnamate derivative involves a different structure–activity behavior when compared to the other training compounds. However, the prediction error of the improved model, measured through the test set and Leave-One-Out results, is not affected in a great extend ($RMSD_{test} = 0.076$, $S_{loo} = 0.10$) when compared to that of QSAR-3, and so we decide to include compound **22** into the analysis in order to have a more general model. The R^2_m powerful statistical index (Pratim Roy et al., 2009) is also acceptable for this model ($R^2_m > 0.5$). Table 3S includes the predicted $\log_{10}(GI + 50)$ values achieved with QSAR-3 for both the training and test sets of cinnamate analogs.

The applicability domain of QSAR-3 reveals that the four cinnamate compounds of the test set belong to the AD of the model ($h_i < h^* = 0.52$), and thus their predicted antifungal activities can be considered as reliable. In the training set, only compound **26** exceeds h^* and **24** is close to h^*, but these are good leverage compounds having low residuals. The leverage values are provided in Table 4S. Finally, the Y-Randomization technique demonstrates that the QSAR-3 model has $S_{train} < S^{rand} = 0.090$ and thus a valid structure–activity relationship is achieved. We also check that model accomplishes with the external validation criteria recommended in Ref. (Golbraikh and Tropsha, 2002) to assure

Table 1
Cinnamates synthesis.

ID	Time (h)	R_1	R_2	R_3	R_4	R_5	Yield (%)
28	4	H	H	H	H	Phenyl	88
29	4	H	H	H	H	4-Methylphenyl	90
30	4	H	H	H	H	2-Methylphenyl	85
31	4	H	H	H	H	4-Methoxyphenyl	85
32	4	H	H	H	H	2-Methoxyphenyl	83
33	4	H	H	H	H	4-Chloro	85
34	4	H	H	H	H	4-Bromo	82
35	4	H	H	H	H	2-Isopropyl-5-methyl	81
36	4	H	H	H	H	2-Biphenylyl	85
37	4	H	H	H	H	4-Biphenylyl	90
38	4	H	H	H	H	4-Propanoyl	83
39	4	H	H	H	H	2-Naphthyl	85
40	4	H	H	H	H	1-Naphthyl	83
41	4	H	H	CH_3	H	Phenyl	88
42	4	H	H	CH_3	H	4-Methylphenyl	90
43	4	H	H	CH_3	H	4-Methoxyphenyl	92
44	18	Phenyl	H	H	H	Phenyl	82
45	18	H	H	H	H	Phthalimidoethyl	80
46	18	H	H	CH_3	H	Phthalimidoethyl	80
47	18	H	Cl	H	H	Phthalimidoethyl	35
48	18	CH_3O	CH_3O	CH_3O	CH_3O	Phthalimidoethyl	70

Table 2
QSAR for the growth inhibition of *Pythium sp* by cinnamate derivatives (100 ppm).

Regression coefficient (standard error)		Statistical quality
QSAR-1		
Intercept	2.184 (0.09)	$N_{train}/d = 23$, $R^2_{train} = 0.27$, $S_{train} = 0.13$, $o \geq 2S = 0$[a],
RDF020m	−0.660 (0.2)	$R^2_{loo} = 0.13$, $S_{loo} = 0.15$, $R^2_m = 0.39$, $R^2_{test} = 0.91$, $S_{test} = 0.13$, $RMSD_{test} = 0.090$
QSAR-2		
Intercept	2.659 (0.2)	$N_{train}/d = 11.5$, $R^2_{train} = 0.50$, $S_{train} = 0.11$,
RDF040e	−0.0327 (0.008)	$R^2_{max} = 0.41$[b], $o \geq 2S = 0$,
$R6u^+$	−9.139 (2)	$R^2_{loo} = 0.31$, $S_{loo} = 0.13$, $R^2_m = 0.40$, $R^2_{test} = 0.96$, $S_{test} = 0.17$, $RMSD_{test} = 0.085$
QSAR-3		
Intercept	2.564 (0.09)	$N_{train}/d = 7.7$, $R^2_{train} = 0.71$, $S_{train} = 0.086$, $R^2_{max} = 0.33$, $o \geq 2S = 0$,
RDF020u	−0.141 (0.02)	$R^2_{loo} = 0.55$, $S_{loo} = 0.11$, $R^2_m = 0.68$,
RDF150m	0.238 (0.05)	$R^2_{test} = 0.82$, $S_{test} = -$, $RMSD_{test} = 0.072$
HATS2m	−1.719 (0.4)	
QSAR-4		
Intercept	1.049 (0.2)	$N_{train}/d = 5.7$, $R^2_{train} = 0.82$, $S_{train} = 0.070$,
DISPv	0.0534 (0.007)	$R^2_{max} = 0.71$, $o \geq 2S = 0$,
RDF065m	−0.106 (0.01)	$R^2_{loo} = 0.72$, $S_{loo} = 0.088$, $R^2_m = 0.60$,
Mor27v	−2.753 (0.3)	$R^2_{test} = 0.94$, $S_{test} = -$, $RMSD_{test} = 0.076$
E2e	1.473 (0.3)	

  [a] $o \geq 2S$ indicates the number of molecules with a predicted residual being equal or greater than three times the S_{train} value.
  [b] R^2_{max} represents the maximum squared correlation coefficient between two given descriptors of the model.

predictive capability, that is to say:

$$1 - R^2_0/R^2_{test} < 0.1 \text{ or } 1 - R'^2_0/R^2_{test} < 0.1$$

$$0.85 \leq k \leq 1.15 \text{ or } 0.85 \leq k` \leq 1.15$$

$$R^2_m > 0.5$$

The R^2_0, R'^2_0, k, $k`$ and R^2_m parameters are defined in Table 2S and their values in Table 5S.

Now, it is possible to examine the role of the molecular descriptors appearing in the developed QSAR models from Table 2; Table 6S includes a brief description for them. The QSAR-3 equation includes three conformation-dependent descriptors: two Radial Distribution Functions (RDF): *RDF*020u, RDF number 2.0/unweighted and *RDF*150m: RDF number 15.0/weighted by atomic masses; and a GETAWAY descriptor: *HATS*2m, leverage-weighted autocorrelation of lag 6/weighted by atomic masses. The correlation matrix for the model is provided in Table 7S, revealing the absence of very high correlations between descriptors pairs.

A Radial Distribution Function (Todeschini and Consonni, 2009) of an ensemble of atoms can be interpreted as the probability distribution of finding an atom in a spherical volume of certain radius, and incorporates atomic properties as weights, such as polarizabilities, volumes, masses or electronegativities, in order to differentiate the contributions of atoms to the activity under study. The GETAWAY (GEometry, Topology, and Atom-Weights AssemblY) type of descriptors (Todeschini and Consonni, 2009) have been designed with the main purpose of matching the 3D-molecular geometry. These numerical variables are derived from the elements h_{ij} of the Molecular Influence matrix (**H**), obtained through the values of atomic Cartesian coordinates. The diagonal elements of **H** (h_{ii}) are called leverages, and are considered to represent the influence of each atom on the whole shape of the molecule. For instance, the mantle atoms always have higher h_{ii} values than atoms near the molecule center, while each off-diagonal element h_{ij} represents the degree of accessibility of the *j*th atom to interactions with the *i*th one. Table 7S includes the numerical values for such molecular descriptors.
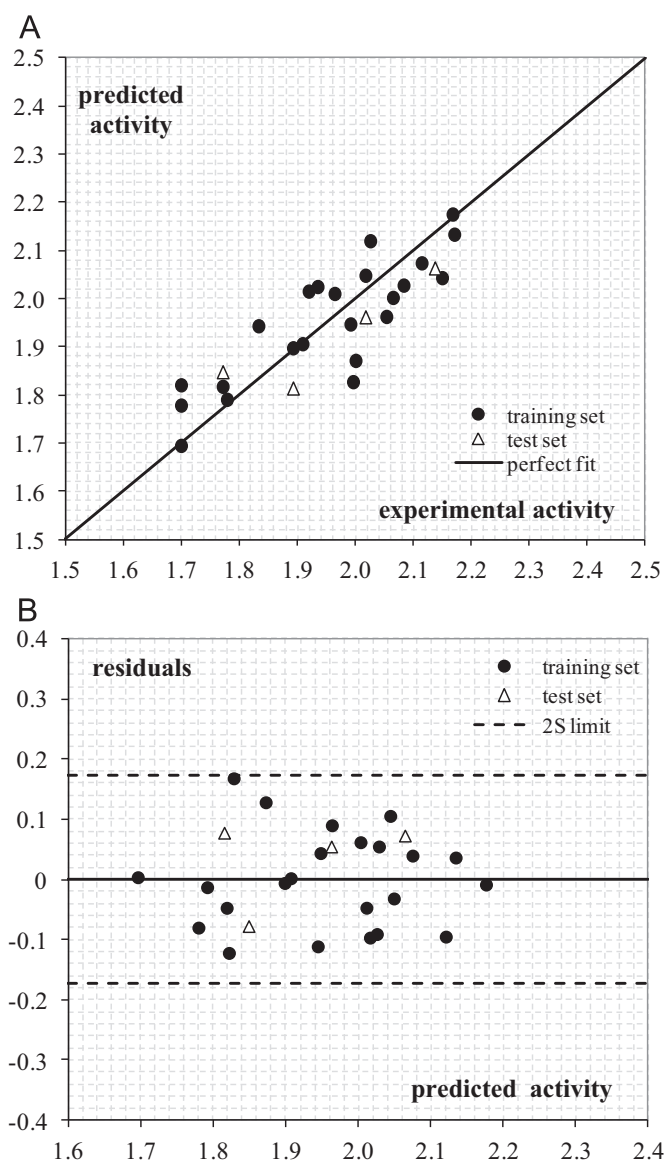
Fig. 2. (A) QSAR-3 predicted $\log_{10}(GI + 50)$ antifungal activities as function of experimental values for *Pythium sp.* (B) Residuals and predicted antifungal activities.

Table 3
QSAR for the growth inhibition of *Corticium rolfsii* by cinnamate derivatives (100 ppm).

Regression coefficient (standard error)		Statistical quality
QSAR-5		
Intercept	2.181 (0.07)	$N_{train}/d = 23$, $R^2_{train} = 0.32$, $S_{train} = 0.10$, $o \geq 2S = 1^{a}$,
PHI	−0.0425 (0.01)	$R^2_{loo} = 0.18$, $S_{loo} = 0.11$, $R^2_m = 0.81$,
		$R^2_{test} = 0.82$, $S_{test} = 0.078$,
		$RMSD_{test} = 0.055$
QSAR-6		
Intercept	2.247 (0.05)	$N_{train}/d = 11.5$, $R^2_{train} = 0.62$,
BELm8	−0.700 (0.1)	$S_{train} = 0.078$, $R^2_{max} = 0.80^{b}$, $o \geq 2S = 0$,
RDF070v	0.0794 (0.02)	$R^2_{loo} = 0.52$, $S_{loo} = 0.088$, $R^2_m = 0.036$,
		$R^2_{test} = 0.74$, $S_{test} = 0.19$, $RMSD_{test} = 0.095$
QSAR-7		
Intercept	2.700 (0.1)	$N_{train}/d = 7.7$, $R^2_{train} = 0.76$, $S_{train} = 0.064$,
BELv5	−0.709 (0.1)	$R^2_{max} = 0.77$, $o \geq 2S = 1$,
RDF080u	−0.0247 (0.006)	$R^2_{loo} = 0.60$, $S_{loo} = 0.092$, $R^2_m = 0.70$,
L3e	0.338 (0.06)	$R^2_{test} = 0.98$, $S_{test} = -$, $RMSD_{test} = 0.037$
QSAR-8		
Intercept	4.645 (0.5)	$N_{train}/d = 5.7$, $R^2_{train} = 0.88$, $S_{train} = 0.047$,
Mv	−2.585 (0.6)	$R^2_{max} = 0.76$, $o \geq 2S = 0$,
BELv5	−1.063 (0.1)	$R^2_{loo} = 0.80$, $S_{loo} = 0.062$, $R^2_m = 0.61$,
RDF080u	−0.0255 (0.004)	$R^2_{test} = 0.97$, $S_{test} = -$, $RMSD_{test} = 0.036$
L3u	0.432 (0.05)	

[a] $o \geq 2S$ indicates the number of molecules with a predicted residual being equal or greater than three times the S_{train} value.
[b] R^2_{max} represents the maximum squared correlation coefficient between two given descriptors of the model.

It is possible to perform a mathematical interpretation of these structure–activity findings. After standardization, the resulting order of importance of the descriptors is as follows:

$$RDF020u(1.1) > RDF150m(0.7) > HATS2m(0.6) \qquad (1)$$

where the standardized regression coefficients are shown in parentheses. The most important descriptor for the antifungal activity of this set of cinnamate derivatives is $RDF020u$, whose numerical values change most in accordance to the numerical variations of the experimental activity. However, the relative magnitudes of the standardized coefficients (1.1, 0.7 and 0.6) suggest that the numerical descriptors complement each other inside the linear equation, and that result of comparable relevance for predicting the studied activity.

According to Table 8S, the three descriptors take positive numerical values. In addition, from Table 2 it is noted that $RDF020u$ and $HATS2m$ have negative regression coefficients in the linear model, while $RDF150m$ has a positive coefficient. Therefore, we conclude that increased numerical values of $RDF150m$ or decreased values of $RDF020u$ or $HATS2m$ descriptors would lead to structures having a higher growth inhibitory activity on *Pythium*

sp. This result is in line with the predictions found for the most active cinnamate structures: **25**, **11**, **13**, **9** and **10**.

3.2.2. QSAR for the growth inhibition of Corticium rolfsii

Table 3 clearly demonstrates that three-descriptors lead to a better quality structure–activity relationship in QSAR-7, as it is the case for the previous fungi species, having a favorable $N_{train}/d = 7.7$ ratio. The graphical representation of the QSAR-7 model results appropriate in Fig. 3A and B. There is only one outlier cinnamate derivative (**12**) having a residual value in the interval $[2S, 2.5S]$; again, we also decide to keep this compound inside the model in order to have a generally applicable model. The condition given by the Y-Randomization technique is fulfilled, $S_{train} < S^{rand} = 0.070$, and also the external validation criteria in Table 5S. The R^2_m index is also acceptable for this model ($R^2_m > 0.5$). Inspecting the AD, Table 9S indicates that the four cinnamate compounds of the test set belong to the AD of the models ($h_i < h^* = 0.52$). In the training set, only compound **2** exceeds h^* but it is a good leverage compound.

The QSAR-7 involves a mixture of 2D and 3D molecular descriptors in order to explain the fungicidal activity: a BCUT: $BELv5$, lowest eigenvalue no. 5 of Burden matrix/weighted by atomic van der Waals volumes; a RDF descriptor: $RDF080u$, RDF number 8.0/unweighted; and a WHIM descriptor: $L3e$, third component size directional WHIM index/weighted by atomic Sanderson electronegativities. Such descriptors are not seriously intercorrelated in Table 7S.

BCUT descriptors are the eigenvalues of a modified connectivity matrix, the Burden matrix (**B**) (Todeschini and Consonni, 2009). The ordered sequence of the n smallest eigenvalues of **B** has been proposed as a molecular descriptor based on the assumption that the lowest eigenvalues contain contributions from all the atoms and thus reflect the molecular topology. The BCUT descriptors are
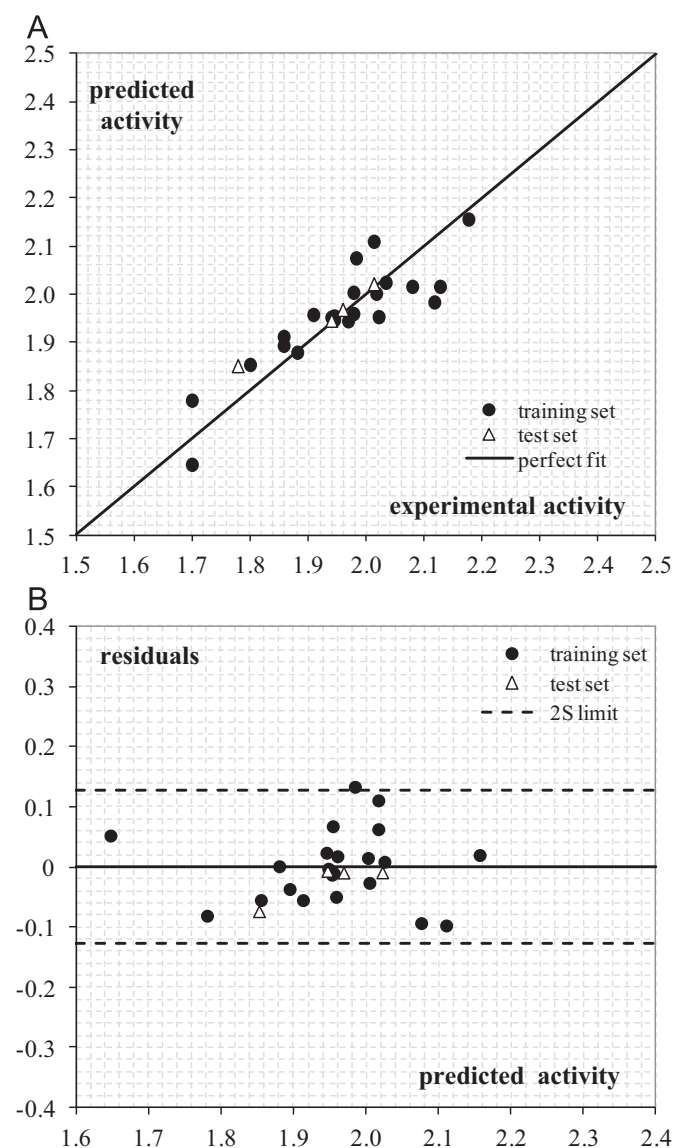
Fig. 3. (A) QSAR-7 predicted $\log_{10}(GI+50)$ antifungal activities as function of experimental values for *Corticium rolfsii*. (B) Residuals and predicted antifungal activities.

Table 4
QSAR predicted $\log_{10}(GI+50)$ antifungal activities for new synthesized cinnamate derivatives.

ID	*Pythium sp* QSAR-3	*Corticium rolfsii* QSAR-7
28	2.1147	1.9159
29	2.0758	1.9025
30	2.1023	1.8292
31	1.8877	1.8571
32	1.6432	1.7826
33	2.0768	1.9081
34	1.9477	1.9187
35	1.9502	1.8552
36	1.9352	1.5717
37	2.3610	1.7740
38	2.2090	1.7929
39	2.0929	1.7973
40	2.0566	1.6907
41	2.0782	1.8921
42	2.1135	1.7370
43	1.9361	1.6981
44	1.7490	1.6270
45	2.1782	1.9264
46	2.0941	1.8569
47	2.2495	1.9442
48	2.2494	1.5973

3.2.3. Antifungal activities prediction for new cinnamate compounds

As a final step of present QSAR analysis, we apply the established QSAR-3 and QSAR-7 models to predict the new synthesized 21 alkyl and aryl cinnamate derivatives (**28–48**), with unknown experimental fungicidal activities. These results are presented in Table 4. From Tables 4S and 9S, an inspection of the leverage values of some cinnamates inhibiting *Pythium sp*, such as **37**, **45–48**, and compounds inhibiting *C. rolfsii*, like **36**, **40**, **44** and **48**, reveals that such molecules have high leverages and are outside the AD, so their predicted activities should be considered as a substantial extrapolation of the model and cannot be treated as reliable. Compounds **45–48** have a molecular structure different to the ones investigated in the training set, due to the phthalimidoethyl functional group, so these results are consistent and one should not use QSAR-3 or QSAR-7 to predict **45–48**. Among the new compounds synthesized, some promising structures that deserve to be experimentally analyzed in forthcoming bioassays are compounds like **28**, **38** and **42** for the case of *Pythium sp*, and **28** and **34** for the *C. rolfsii* genus species.

4. Conclusions

The appropriate development of linear QSAR models on the available experimental fungicidal activities enables to predict these values for new cinnamate derivatives, that still do not have experimental data for the growth inhibition on *Pythium sp* and *C. rolfsii* fungi species. The exploration of more than a thousand of constitutional, topological, geometrical and electronic molecular descriptors leads to predictions of the growth inhibition in close agreement to the experimental values extracted from the literature.

The aforementioned method provides a clean, simple, and useful alternative for preparing substituted cinnamate derivatives; the use of silica-supported Preyssler heteropolyacid catalysts provides very good yields, also leading to an easy separation and recovery of the catalysts for further use. The catalytic activity, which is practically constant in consecutive reaction batches, and the high recovery of the catalysts allow for both low environmental impact and low cost. Other "green" advantages of the

an extension of the Burden eigenvalues and consider three classes of matrices, whose diagonal elements account for atomic charge related values, atomic polarizability related values and atomic H bond abilities. WHIM (Weighted Holistic Invariant Molecular Descriptors) descriptors are based on statistical indices calculated on the projections of atoms along principal axes (Todeschini and Consonni, 2009). The aim is to capture 3D-information regarding size, shape, symmetry and atom distributions with respect to invariant reference frames. In order to calculate them, a weighted covariance matrix is obtained from different weighting schemes for the atoms.

The order of descriptors importance is:

$$L3e(1.6) > BELv5(1.2) > RDF080u(1.0) \tag{2}$$

Again, the three descriptors take positive numerical values (Table 8S). Taking into account the signs of the regression coefficients, increased numerical values of $L3e$ or decreased values of $BELv5$ or $RDF080u$ descriptors would lead to structures having a higher growth inhibitory activity on *C. rolfsii*.

method are the low formation of wastes and the replacement of corrosive, soluble mineral acids.

Acknowledgments

Appendix A. Supplementary information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.ecoenv.2015.09.024.

References

Abdelgaleil, S.A.M., Abbassy, M.A., Belal, A.-S.H., Abdel Rasoul, M.A.A., 2008. Bioactivity of two major constituents isolated from the essential oil of *Artemisia judaica* L. Bioresour. Technol. 99, 5947–5950.

Aoki, S., Amamoto, C., Oyamada, J., Kitamura, T.A., 2005. A convenient synthesis of dihydrocoumarins from phenols and cinnamic acid derivatives. Tetrahedron 67, 9291–9297.

Awad, R., Muhammad, A., Durst, T., Trudeau, V., Arnason, J., 2009. Bioassay-guided fractionation of lemon balm (*Melissa officinalis* L.) using an in vitro measure of GABA transaminase activity. Phytother. Res. 23, 1075–1081.

Clifford, M.N., 2000. Chlorogenic acids and other cinnamates – nature, occurrence, dietary burden, absorption and metabolism. J. Sci. Food Agric. 80, 1033–1043.

Creaser, I., Heckel, M., Neitz, R., Pope, M., 1993. Rigid nonlabile polyoxometalate cryptates [ZP5W30O110](15–n)- that exhibit unprecedented selectivity for certain lanthanide and other multivalent cations. Inorg. Chem. 32, 1573–1578.

Draper, N.R., Smith, H., 1981. Applied Regression Analysis. John Wiley & Sons, New York.

Duchowicz, P.R., et al., 2009. QSAR analysis on *Spodoptera litura* antifeedant activities for flavone derivatives. Sci. Total Environ. 408, 277–285.

Duchowicz, P.R., Castro, E.A., Fernández, F.M., 2006. Alternative algorithm for the search of an optimal set of descriptors in QSAR–QSPR studies. MATCH – Commun. Math. Comput. Chem. 55, 179–192.

Duchowicz, P.R., Vitale, M.G., Castro, E.A., Fernández, M., Caballero, J., 2007. QSAR analysis for heterocyclic antifungals. Bioorg. Med. Chem. 15, 2680–2689.

E-Dragon. Milano Chemometrics and QSAR Research Group. VCCLAB, Virtual Computational Chemistry Laboratory. ⟨http://michem.disat.unimib.it/chm⟩.

Freedman, S.O., Shulman, R., Krupey, J., Sehon, A.H., 1964. Antigenic properties of chlorogenic acid. J. Allergy 35, 97–107.

Gobec, S., Sova, M., Kristan, K., Rižner, T.L., 2004. Cinnamic acid esters as potent inhibitors of fungal 17beta-hydroxysteroid dehydrogenase-a model enzyme of the short-chain dehydrogenase/reductase superfamily. Bioorg. Med. Chem. Lett. 14, 3933–3936.

Golbraikh, A., Tropsha, A., 2002. Beware of q2!. J. Mol. Gr. Model. 20, 269–276.

Gramatica, P., 2007. Principles of QSAR models validation: internal and external. QSAR Comb. Sci. 26, 694–701.

Hansch, C., Fujita, T., 1995. Classical and three-dimensional QSAR in agrochemistry. J. Am. Chem. Soc. 606, 318–329.

Hansch, C., Leo, A., 1995. Exploring QSAR. Fundamentals and Applications in Chemistry and Biology. American Chemical Society, Washington, D.C.

HyperChem. Hypercube Inc. ⟨http://www.hyper.com⟩.

Isaacs, N., Najem, T., 1988. The E1cb route for ester hydrolysis; volumes of activation as an additional criterion of mechanism. J. Chem. Soc. Perkin Trans. 2, 557–562.

Jarvis, W.R., 1992. Managing Diseases in Greenhouse Crops. APS Press, New York.

Kristan, K., Starčevic, S., Brunskole, M., Rižner, T.L., Gobec, S., 2006. Cinnamates and cinnamamides inhibit fungal 17β-hydroxysteroid dehydrogenase. Mole. Cell. Endocrinol. 248, 239–241.

Lee, E.-J., Kim, J.-R., Choi, D.-R., Ahn, Y.-J., 2008. Toxicity of cassia and cinnamon oil compounds and cinnamaldehyde-related compounds to *Sitophilus oryzae* (coleoptera: curculionidae). J. Ecol. Entomol. 101, 1960–1966.

Li, K., Foresee, L.N., Tunge, J.A., 2005. Trifluoroacetic acid-mediated hydroarylation: synthesis of dihydrocoumarins and dihydroquinolones. J. Org. Chem. 70,

2881–2883.

Mali, R., Papalkar, A., 2003. Useful synthesis of aryl cinnamates: synthesis of difengpin. J. Chem. Res. (S) 10, 603–604.

Mercader, A.G., Duchowicz, P.R., Fernández, F.M., Castro, E.A., 2010. Replacement method and enhanced replacement method versus the genetic algorithm approach for the selection of molecular descriptors in QSPR/QSAR theories. J. Chem. Inf. Model. 50, 1542–1548.

Mercader, A.G., Duchowicz, P.R., Fernández, F.M., Castro, E.A., 2011. Advances in the replacement and enhanced replacement method in QSAR and QSPR theories. J. Chem. Inf. Model. 51, 1575–1581.

Moghaddam, F.M., Ghaffarzadeh, M.G., Abdi-Oskoui, S.H., 1999. Tandem fries reaction-conjugate addition under microwave irradiation in dry media; one-pot synthesis of flavanones. J. Chem. Res. (S) 9, 574–576.

Palermo, V., Ruiz, D., Autino, J., Vázquez, P., Romanelli, G., 2012. Simple halogen-free synthesis of aryl cinnamates using Mo-Keggin heteropolyacids as catalyst. Pure Appl. Chem. 84, 529–540.

Pasquale, G., et al., 2012. Quantitative structure–activity relationships on chalcone derivatives: mosquito larvicidal studies. J. Agric. Food Chem. 60, 692–697.

Peterson, C.J., Tsao, R., Eggler, A.L., Coats, J.R., 2000. Insecticidal activity of cyanohydrin and monoterpenoid compounds. Molecules 5, 648–654.

Pinto, C.G.A., et al., 1999. Síntesis and molecular structure of 3-(2-Benzyloxy-6-hydroxyphenyl)-5-styrylpyrazoles. Reaction of 2-Styrylchromones and hydrazine hydrate. Tetrahedron 55, 10187–10200.

Pratim Roy, P., Paul, S., Mitra, I., Roy, K., 2009. On two novel parameters for validation of predictive QSAR models. Molecules 14, 1660–1701.

Puzyn, T., Leszczynski, J., Cronin, M.T.D. (Eds.), 2010. Recent advances in QSAR studies: methods and applications. Challenges and Advances in Computational Chemistry and Physics. Springer Science & Business Media B.V., Netherlands.

Rücker, C., Rücker, G., Meringer, M., 2007. Y-randomization and its variants in QSPR/QSAR. J. Chem. Inf. Model. 47, 2345–2357.

Romanelli, G., Ruiz, D., Vázquez, P., Thomas, H., Autino, J., 2010a. Preyssler heteropolyacid H14[NaP5W29MoO110]: a heterogeneous, green and recyclable catalyst used for the protection of functional groups in organic synthesis. Chem. Eng. J. 161, 355–362.

Romanelli, G.P., et al., 2010b. Sustainable synthesis of flavonoid derivatives, QSAR study and insecticidal activity against the fall armyworm, Spodoptera frugiperda (Lep.: *Noctuidae*). J. Agric. Food Chem. 58, 6290–6295.

Ruiz, D., et al., 2008. Direct esterification of cinnamic acids with phenols and imidoalcohols: a simple, heteropolyacid-catalyzed procedure. Arkivoc 12, 269–276.

Ruiz, D., 2009. Síntesis Sustentable de Fenilpropanoides: Cinamatos de Arilo, 4-Fenilcumarinas, Dihidrocumarinas y Dihidroangelicinas (Doctoral thesis). UNLP (Argentina), La Plata, pp. 52–62, and references cited herein.

Shankaran, K., Sloan, C.P., Snieckus, V., 1985. Synthetic connections to the aromatic directed metalation reaction. Radical-induced cyclization to substituted benzofurans, benzopyrans, and furopyridines. Tetrahedron Lett. 26, 6001–6004.

Shu, L., Hongjun, Y., 2013. An overview on synthetic methods of alkyl cinnamates. Eur. Chem. Bull. 2, 76–77.

Sinha, A., Sharma, A., Swaroop, A., Kumar, V., 2007. An efficient chemoselective strategy for the preparation of (E)-cinnamic esters from cinnamaldehydes using heterogeneous catalyst and DDQ. Tetrahedron 63, 1000–1007, and references cited herein.

Speck-Planche, A., Kleandrova, V.V., Rojas-Vargas, J.A., 2011c. QSAR model toward the rational design of new agrochemical fungicides with a defined resistance risk using substructural descriptors. Mol. Divers. 15, 901–909.

Speck-Planche, A., Kleandrova, V.V., Scotti, M.T., 2012b. Fragment-based approach for the in silico discovery of multi-target insecticides. Chemom. Intell. Lab. Syst. 111, 39–45.

Speck-Planche, A., Cordeiro, M.N.D.S., Guilarte-Montero, L., Yera-Bueno, R., 2011a. Current computational approaches towards the rational design of new insecticidal agents. Curr. Comput. Aided Drug Des. 7, 304–314.

Speck-Planche, A., Kleandrova, V.V., Luan, F., Cordeiro, M.N.D.S., 2012a. Predicting multiple ecotoxicological profiles in agrochemical fungicides: a multi-species chemoinformatic approach. Ecotoxicol. Environ. Saf. 80, 308–313.

Speck-Planche, A., Guilarte-Montero, L., Yera-Bueno, R., Rojas-Vargas, J.A., Garcia-Lopez, A., Uriarte, E., Molina-Perez, E., 2011b. Rational design of new agrochemical fungicides using substructural descriptors. Pest Manag. Sci. 67, 438–445.

Sunnerheim, K., et al., 2007. Quantitative structure–activity relationships of Pine Weevil antifeedants, a multivariate approach. J. Agric. Food Chem. 55, 9365–9372.

Tawata, A., et al., 1996. Synthesis and antifungal activity of cinnamic acid esters. Biosci. Biotech. Biochem. 60, 909–910.

Todeschini, R., Consonni, V., 2009. Molecular Descriptors for Chemoinformatics (Methods and Principles in Medicinal Chemistry). Wiley-VCH, Weinheim, p. 41.

Tute, M.S., 1971. History and Objectives of Quantitative Drug Design in Advances in Drug Research. Academic Press, London.

Womack, E., McWhirter, J., 1955. Phenyl cinnamate. Org. Synth. Coll. 3, 714–715.

Zhao, Y., Wang, J., Ballevre, O., Luo, H., Zhang, W., 2011. Antihypertensive effects and mechanisms of chlorogenic acids. Hypertens. Res. 35, 370–374.