Chemical Papers
https://doi.org/10.1007/s11696-026-05153-1

ORIGINAL PAPER

Classification and QSAR of the anticancer activity of (*E*)-stilbenes

Francisco Torrens[1] · **Adela León**[2] · **Juan A. Castillo-Garit**[3] · **Gloria Castellano**[4]

Received: 14 September 2024 / Accepted: 30 May 2026

Abstract
In the present report 29 (*E*)-stilbenes are clustered by using a procedure based on artificial intelligence. The objective is to predict cytotoxicity (anticancer activities) of them and other similar stilbenes in nine cell lines. We make a periodic classification of stilbenes by using the information entropy theory to select the most active classes. The structures of seven different classes are obtained. The most active Class 1 is located at the bottom right of the periodic classification. We provide new compounds that would also have high activity because they are in the bottom-right groups. Moreover, we relate stilbenes' cytotoxicity in the cell lines to their physical and chemical properties by QSAR and PCA. The scores plot separates clusters that contain classes in the periodic system. The results of the nine QSAR models are good with r^2 greater than 0.57 and show the repetition of most variables for all the nine cell lines. The leave-*m*-out cross-validation determines the good robustness for the cell lines ($q^2 > 0.33$). These results agree with the loading plot and suggest the importance of these descriptors in further studies of other cell lines.

Keywords Clustering · Cytotoxicity · Equipartition conjecture · Information entropy · Leave-many-out cross validation · Periodic table · Principal component analysis · Robustness

Introduction

Stilbenes are a line of phenolic vegetable secondary metabolites, which have been the topic of exhaustive investigation because of their significant pharmacological properties. Stilbenes are originated either via the phenylpropanoid path or via the polyketide path (Dubrovina and Kiselev 2017).

Each stilbene is made of two phenol moieties connected by a C2 link (*cf.* Fig. 1). Structurally they are separated into *Z*-type and *E*-type based on the configuration of their central double bond. They can endure *Z*–*E* isomerization altering the general configuration and decreasing the bioactivity. The essential stilbene structure is extensive in plants, and different species-precise replacement designs exist. In a plant, stilbenes seem to function as guards against a number of biotic and abiotic pressures (Hammerbacher et al. 2011). They show bioactivity, which is significant for the whole bionetwork. Besides, stilbenes show antibacterial and fungistatic properties, and some of them show helpful effects on the cardiovascular system (Chrzascik 2009).

A lot of research has lately been published on resveratrol, a phenolic (*E*)-stilbene, which has been proposed as a possible chemopreventive molecule, grounded on its arresting inhibitory property on cellular events related to tumor initiation, promotion, and progression, as well as showed in vitro growth inhibition in several human tumor cell lines (Frémont 2000). Notwithstanding, resveratrol hurts from little water solubility, degradation, and weak bioavailability (Bohara et al. 2022). In mammals, resveratrol is significantly metabolized and quickly removed, as well as therefore it

✉ Francisco Torrens
torrens@uv.es

✉ Gloria Castellano
gloria.castellano@ucv.es

[1] Institut Universitari de Ciència Molecular, Universitat de València, Edifici d'Instituts de Paterna, P. O. Box 22085, 46071 València, Spain

[2] Escuela de Doctorado, Universidad Católica de Valencia San Vicente Mártir, 46008 València, Spain

[3] Instituto Universitario de Investigación y Desarrollo Tecnológico (IDT), Universidad Tecnológica Metropolitana, Santiago, Chile

[4] Research Group Structure-Bioactivity Relationship of Organic Compounds, Facultad de Veterinaria y Ciencias Experimentales, Universidad Católica de Valencia San Vicente Mártir, Guillem de Castro-94, 46001 València, Spain

Published online: 16 June 2026

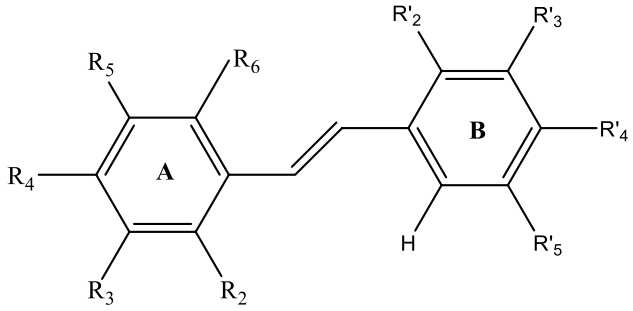


Fig. 1 The basic structure of the stilbene derivatives

shows a weak bioavailability (Chimento et al. 2019; Pannu and Bhatnagar 2019).

Polymethoxystilbenes and connected molecules are commonly antiproliferative substances with antiangiogenic property (Morris et al. 2015), as well as they produce inhibition of vascular endothelial growth factor (VEGF), nuclear factor κB (Basini et al. (2010) and the multidrug resistance of cancer cells (Holwell et al. 2002; Lion et al. 2005; Ma et al. 2019; Mazue et al. 2010; Roberti et al. 2003; Shnyder et al. 2003; Valdameri et al. 2012). They relate to biomembrane models, experience a distinct metabolic change, and have a greater bioavailability than resveratrol.

To verify the function of the number and locations of the methoxyl functional groups, double-bond configuration, or the extra hydroxyl groups on antitumor properties, and to describe the structure–activity relationships (SARs), many polymethoxylated resveratrol derivatives have been synthesized and verified. The SARs analyses, carried out on methoxylated stilbenes, disclosed that 3,5-dimethoxyl and 3,4,5-trimethoxyl patterns are significant for the pro-apoptotic property (Chalal et al. 2014; De Filippis et al. 2027; Yang et al. 2017; Zhang et al. 2014).

Roslie et al. (2012) published in what way 3,5-dibenzyloxy-4'-hydroxystilbene provokes initial caspase-9 activation throughout apoptosis, in human K562 chronic myelogenous leukemia cells. Martí-Centelles et al. (2015) informed the inhibitory influence of cytotoxic (*E*)-stilbenes associated with resveratrol on the expression of the VEGF, hTERT and c-Myc genes. Sirerol et al. (2016) analyzed the function of natural stilbenes in the inhibition of tumors. Almosnid et al. (2016) reported the in vitro anticancer influence of a couple of new oligostilbenes, (*Z*)- and (*E*)-suffruticosol D, isolated from *Paeonia suffruticosa* seeds. Mikstacka et al. (2018) published the cytotoxic, tubulin-interfering and proapoptotic properties of resveratrol-equivalent 4'-methylthio-(*E*)-stilbenes. Aja et al. (2020) informed the searching of natural stilbene oligomers from *Vitis vinifera* for antitumor properties, on human hepatocellular carcinoma cells. Zielińska-Przyjemska et al. (2020) reported the influence of methoxyl stilbenes on the feasibility and stimulation of cell cycle arrest and apoptosis, on human myeloid leukemia cells. A significant function in the discovery of the leaders is participated by the control of the target that they operate, and the influence of in what way a specific molecule will bind to an enzyme or DNA. For instance, it is accepted that stilbenes are efficient inhibitors of human topoisomerases (Hwangbo et al. 2012; Liu et al. 2022; Yamada et al. 2006), one of the important cell cycle enzymes (D'yakonov et al. 2017), which inhibition unavoidably causes cancer cell death. Distinct cell lines cover diverse quantities of topoisomerases.

Csuk et al. (2012) synthesized a set of (*E*)-stilbenes and searched for anticancer properties in a group of eight human tumor cell lines by colorimetry, with the sulforhodamine B assay. These molecules are motivators of apoptosis, and some of them present great anticancer property in opposition of the cell lines. Cell death is activated by apoptosis as shown by a dye-exclusion examination, extra dual acridine orange/ethidium bromide tests and DNA-laddering trials. Additional tests are undecided whether these molecules show important tubulin-binding activity and can operate as antiangiogenic agents (Kimura et al. 2008; Mustafa et al. 2019; Tron et al. 2006). Some of these molecules show great cytotoxicity.

In earlier publications, we reported results of the periodic classification of 66 stilbenes into a scheme of their antioxidant properties, by utilizing a technique founded on information entropy theory in artificial intelligence (Castellano et al. 2014), as well as a periodic classification of congeneric antileukemic *S*–alkylcysteine ketones and quantitative SAR (QSAR) of homologous ones (Castellano et al. 2021). We provided importance to the structural conditions for antioxidant activity or tumor strength. Now our research group intend, from the data got by Csuk et al. (2012), to do an analysis to understand the effect of the substituents in the cycles of the stilbenes on antitumor properties, in the eight cell lines. Similarly, to create a periodic classification of the stilbenes consistent with the antitumor activities. We have chosen the QSAR model to quantify the importance of physicochemical and structural properties, and to see which ones have the greatest influence on anticancer activity in each cell line.

In the present report a family of 29 (*E*)-stilbenes was grouped in accordance with their similar activity. This work deals merely with (*E*)-stilbenes. The purpose of the work was to forecast the cytotoxicity (antitumor property) of the stilbenes in a collection of nine cell lines. For this purpose, we linked their cytotoxicity (pIC_{50}) in eight human tumor cell lines: 18A2 (melanoma), A253 (head), A549 (lung), A2780 (ovarian), DLD1 (colon), 850C (anaplastic thyroid), MCF7 (breast) and LIPO (liposarcoma), as well as non-malignant mouse fibroblasts (NiH3T3) to the physicochemical properties of the molecules, by QSAR and principal

component analysis (PCA). Furthermore, we created a periodic classification of the stilbenes into a scheme of properties by utilizing a technique founded on information entropy theory, in artificial intelligence. Besides, to fit QSAR models (Poddar et al. 2023) of the cytotoxicity information for computing the very significant data. As cytotoxicity was offered for diverse cell lines it was important to examine if the descriptors were recurrent, which would be of interest in additional works with different cell lines. Lastly the leave-many-out cross-validation fixed the robustness of the models. The conclusions that we draw from this work will aid us to learn the derivatives of combretastatin A4 (*Z*)-stilbenes (which are surely more active but less stable, difficult to study and not available in the databases), since we will have a better information of the influence of the substituents on the antitumor properties of the overall stilbenes.

The structure of the work follows (*cf*. Table 1). First, a family of 29 (*E*)-stilbenes is selected, owing to their obtainability of experimental information of cytotoxicity in a collection of eight human tumor and one non-malignant cell lines. Furthermore, it is chosen by the accessibility of a series of properties in the ChEMBL databank. The gathering of stilbenes is given, incorporating the hierarchical vector of properties elected to define the antitumor properties (*cf*. "Program MolClas for molecular classification based on the equipartition conjecture of entropy production" section). The vector of properties contains six structural features in reducing rank regarding the antitumor properties. Formerly the stilbenes experimental information of cytotoxicity is offered in the nine lines. Next the partial correlation plot of the vector of properties proposes us to perform a molecular classification of the stilbenes, founded on the vector of properties. This classification lets us to comprehend the dissimilar structural features of the stilbenes in the diverse classes. The classification entropy permits us to obtain several classifications with unlike number of classes. However, the Equipartition Conjecture of Entropy Production allows us to pick out the best classification. The discovery that there are *holes* in the periodic classification advises us to suggest novel extremely functioning molecules. By utilizing the antitumor properties and the ChEMBL information, the principal component analysis lets us to distinct three clusters of stilbenes and selected groups of variables. The QSAR models allow us to quantify the SARs to choose the finest variables defining the cytotoxicity of the stilbenes. Then an examination of the sign of the correlations permits us to distinguish the cell lines. The applicability domain is verified by utilizing Williams diagrams. The expected *vs*. observed charts show up the separate residuals. The leave-many-out cross-validation offers the robustness of the fits. Lastly the external validation designates the applicability of the QSAR models for external stilbenes that are not comprised in the fits.

Table 1 Set of methods followed to classify (*E*)-stilbenes

Step	Method
1	Choose a set of 29 (*E*)-stilbenes based on cytotoxicity in nine cell lines
2	Adapt the set to the availability of properties in the ChEMBL database
3	Assign the hierarchical vector of properties describing their anticancer activities
4	Calculate the partial correlation diagram from the vector of properties
5	Perform molecular classifications based on the vector of properties by entropy (information theory)
6	Select the number of classes by the Equipartition Conjecture of Entropy Production
7	Perform a principal component analysis with anticancer activities and ChEMBL data to separate clusters of stilbenes and groups of variables
8	Obtain QSAR models to quantify previous SAR models
9	Carry out an analysis of the sign of the correlations of the models
10	Obtain the applicability domain of the models with Williams plots
11	Carry out experimental *vs*. observed plots to analyze the individual residuals
12	Carry out the leave-many-out cross-validation to test the robustness of the models
13	Carry out external validation to analyze the applicability for external stilbenes

The selection of the methods follows. As the vector of properties contains dichotomic *0* or *1* variables, the molecular classification of the stilbenes is performed with a method based on the information entropy to manage this kind of variables. As some properties are more important than some others, they are ordered in a hierarchical vector of properties with decaying importance. Each posterior property has half the importance of the anterior one in describing the SAR models. The classification entropy causes a combinatorial explosion of classifications with different numbers of classes. However, the Equipartition Conjecture of Entropy Production permits us to choose the finest classification. The principal component analysis is utilized to split the clusters of stilbenes and the sets of variables, since both antitumor properties and ChEMBL information are continuous variables. This technique allows to replace the original set of variables with another set of orthogonal variables (principal components), sorted according to decreasing explained variance. Therefore, the new variables can be truncated causing minimum error. Consequently, we reduce the initial variables to a smaller number of principal components. The QSAR is used to choose the finest variables defining the antitumor properties, since both variables and properties are continuous variables. The QSAR multilinear regression by least squares guarantees that the variance of the fit

is negligible, compared to that of the original data. Then the fitted model can be used instead of the original data.

Results and discussion

Table 2 gives the stilbene derivatives studied in this work and their hierarchical vector of proprieties, obtained by a computational method explained in "Materials and methods" section. Table 3 collects the experimental data of cytotoxicity (IC_{50}) obtained by Csuk et al. (2012) in the set of eight cancer cell lines: 518A2 (melanoma), A253 (head), A549 (lung), A2780 (ovarian), DLD1 (colon), 850C (anaplastic thyroid), MCF7 (breast) and LIPO (liposarcoma), as well as non-malignant mouse fibroblasts (NiH3T3). A non-malignant mouse fibroblast cell line was chosen because there are no published data on non-malignant human cell lines. Furthermore, we relied on an article in which all cell lines were studied using the same method and conditions. The IC_{50} values have been softened as $pIC_{50} = -\log IC_{50}$ (IC_{50} in mol L^{-1}) as it is usual in QSAR studies.

GraphCor partial correlation diagram

The methods applied in this subsection are to choose a set of 29 (*E*)-stilbenes based on cytotoxicity in nine cell lines, to adapt the set to the availability of properties in the ChEMBL database, to assign the hierarchical vector of properties describing their anticancer activities and to calculate the partial correlation diagram from the vector of properties.

The matrix of Pearson correlation coefficients has been calculated between each pair of vector properties $<i_1,i_2,i_3,i_4,i_5,i_6>$ (*cf.* Table 2 and "Materials and methods" section), for the 29 stilbene derivatives. The Pearson inter-correlations are computed for the partial correlation diagram, which contains high partial correlations ($r \geq 0.75$), medium partial correlations ($0.50 \leq r < 0.75$), low partial correlations ($0.25 \leq r < 0.50$) and *zero* partial correlations ($r < 0.25$). The pairs of compounds with high partial correlation show a similar vector property.

With the Equipartition Conjecture of Entropy Production, the partial correlations matrix (*cf.* Fig. 2, *left*) contains 163 high, 33 medium, 45 low and 165 *zero* partial correlations.

Many partial correlations are high. Red lines, representing high partial correlations, link stilbenes with the greatest

Table 2 Stilbenes with their vector of properties

No.	Name	R_2	R_3	R_4	R_5	R_6	R'_2	R'_3	R'_4	R'_5	Vector properties
1	(*E*)-6-Fluoro-3-hydroxy-2',4,4'-trimethoxystilbene	H	OH	OMe	H	F	OMe	H	OMe	H	111111
2	(*E*)-4',6-Difluoro-3-hydroxy-3',4,5'-trimethoxystilbene	H	OH	OMe	H	F	H	OMe	F	OMe	111110
3	(*E*)-3-Hydroxy-3',4,5'-trimethoxystilbene	H	OH	OMe	H	H	H	OMe	H	OMe	111011
4	(*E*)-4'-Fluoro-3-hydroxy-3',4,5'-trimethoxystilbene	H	OH	OMe	H	H	H	OMe	F	OMe	111010
5	(*E*)-3-Hydroxy-2',4,5'-trimethoxystilbene	H	OH	OMe	H	H	OMe	H	H	OMe	111011
6	(*E*)-3-Hydroxy-2',4,4'-trimethoxystilbene	H	OH	OMe	H	H	OMe	H	OMe	H	111011
7	(*E*)-4'-Fluoro-3',4,5'-trimethoxystilbene	H	H	OMe	H	H	H	OMe	F	OMe	111010
8	(*E*)-6-Fluoro-3-hydroxy-2',4,5'-trimethoxystilbene	H	OH	OMe	H	F	OMe	H	H	OMe	110001
9	(*E*)-4-Hydroxy-2',3,5,5'-tetramethoxystilbene	H	OMe	OH	OMe	H	OMe	H	H	OMe	110001
10	(*E*)-2',4,5'-Trimethoxystilbene	H	H	OMe	H	H	OMe	H	H	OMe	100011
11	(*E*)-3,3',4,5'-Tetramethoxystilbene	H	OMe	OMe	H	H	H	OMe	H	OMe	100011
12	(*E*)-2',4,4'-Trimethoxystilbene	H	H	OMe	H	H	OMe	H	OMe	H	100011
13	(*E*)-3,4',5-Trihydroxy-3',5'dimethoxystilbene	H	OH	H	OH	H	H	OMe	OH	OMe	110011
14	(*E*)-4'-Fluoro-4-hydroxy-3,3',5,5'-tetramethoxystilbene	H	OMe	OH	OMe	H	H	OMe	F	OMe	110010
15	(*E*)-2,5-Dihydroxy-3'-methoxystilbene	OH	H	H	OH	H	H	OMe	H	H	000001
16	(*E*)-2,4',5-Trihydroxy-3'-methoxy stilbene	H	OH	H	OH	H	H	OMe	OH	H	000001
17	(*E*)-2,2',4'-Trimethoxystilbene	OMe	H	H	H	H	OMe	H	OMe	H	000001
18	2,5-Dihydroxy-4'-methoxystilbene	OH	H	H	OH	H	H	H	OMe	H	000001
19	(*E*)-3',5'-Dimethoxy-4'-fluoro-4-hydroxystilbene	H	H	OH	H	H	H	OMe	F	OMe	000010
20	(*E*)-3',5'-Dimethoxy-4-hydroxystilbene	H	H	OH	H	H	H	OMe	H	OMe	000011
21	(*E*)-4-Hydroxy-3,3',5,5'-tetramethoxystilbene	H	OMe	OH	OMe	H	H	OMe	H	OMe	000011
22	(*E*)-4-Hydroxy-4'-fluoro-3,3',5'-trimethoxystilbene	H	OMe	OH	H	H	H	OMe	F	OMe	000010
23	(*E*)-4'-Fluoro-3,3',4,5'-tetramethoxystilbene	H	OMe	OMe	H	H	H	OMe	F	OMe	000010
24	(*E*)-2,2',5'-Trihydroxystilbene	OH	H	H	H	H	OH	H	H	OH	000001
25	(*E*)-2',5'-Dimethoxy-4-hydroxystilbene	H	H	OH	H	H	OMe	H	H	OMe	000001
26	(*E*)-4-Hydroxy-3',4',5'-trifluorostilbene	H	H	OH	H	H	H	F	F	F	000001
27	(*E*)-2',4'-Dimethoxy-3-hydroxystilbene	H	OH	H	H	H	OMe	H	OMe	H	000011
28	(*E*)-2,4',5-Trihydroxystilbene	OH	H	H	OH	H	H	H	OH	H	000001
29	(*E*)-2,3',5-Trihydroxystilbene	OH	H	H	OH	H	H	OH	H	H	000001

Table 3 Experimental data of cytotoxicity IC_{50} (μM) in eight cancer and one non-malignant cell lines

Entry	518A2	850C	A253	A549	A2780	DLD1	LIPO	NiH3T3	MCF7
1	0.03	0.03	0.03	0.03	0.01	0.04	0.03	0.05	0.06
2	0.2	0.13	0.48	0.18	0.11	0.2	0.21	0.19	0.22
3	0.22	0.21	0.83	0.25	0.15	0.21	0.4	0.21	0.51
4	0.52	0.7	0.52	0.67	0.64	0.85	0.61	0.92	0.39
5	0.72	0.86	0.80	0.96	0.87	0.91	0.65	1.54	0.54
6	1.33	1.33	1.53	1.8	1.27	2.00	1.41	2.08	1.64
7	1.49	2.02	1.79	2.28	2.06	2.22	1.56	2.55	0.89
8	1.74	2.26	1.34	2.21	2.01	2.26	1.98	2.66	2.09
9	2.11	2.62	1.34	2.35	1.40	2.73	2.12	1.87	0.89
10	2.47	1.93	2.19	3.65	1.93	2.65	3.28	3.01	1.50
11	2.81	2.46	2.35	3.04	2.06	2.86	2.86	3.68	1.89
12	3.67	3.30	3.64	5.14	3.21	4.74	4.98	4.90	2.34
13	13.05	13.67	8.93	8.90	10.23	12.09	15.29	10.78	13.00
14	14.89	19.75	13.65	7.17	10.84	14.05	14.15	10.57	9.87
15	15.13	14.53	11.28	8.11	15.87	25.70	12.45	7.16	7.50
16	15.40	13.68	12.78	11.41	12.99	29.43	15.55	9.46	13.00
17	15.53	11.95	9.10	12.75	16.15	9.30	17.01	12.35	16.41
18	15.99	22.94	14.33	5.28	11.66	16.42	21.98	4.51	11.58
19	16.49	13.35	9.30	14.31	11.84	12.43	19.87	11.53	5.80
20	17.39	15.18	17.33	15.86	15.56	16.45	14.05	12.19	14.11
21	17.47	17.42	14.43	18.96	25.96	17.65	15.61	15.00	7.61
22	17.73	18.16	12.87	14.74	29.94	19.42	13.16	14.96	13.16
23	18.04	17.72	15.83	22.16	23.92	14.95	16.51	6.76	11.20
24	22.00	24.45	14.53	19.01	12.42	18.11	24.07	12.38	13.18
25	22.03	24.76	17.25	22.46	18.89	26.83	21.28	7.45	7.25
26	22.45	21.79	19.75	12.24	4.15	19.28	23.63	8.86	12.17
27	25.32	22.49	16.76	21.61	28.05	16.52	26.57	19.56	9.75
28	28.36	18.70	15.16	12.66	16.62	22.91	17.37	9.59	14.17
29	28.47	23.01	10.76	10.00	16.68	21.68	18.26	10.52	8.03

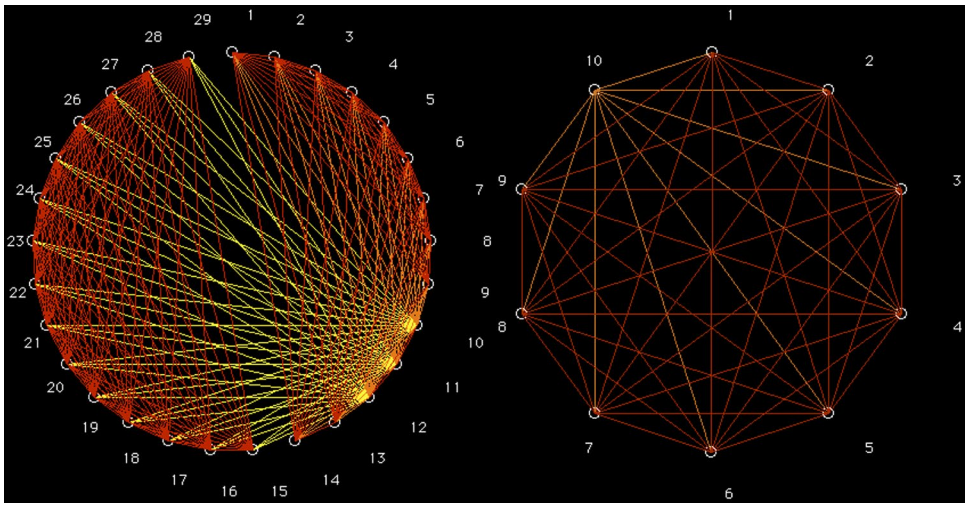


Fig. 2 Partial correlation diagram: high (*red*), medium (*orange*) and low (*yellow*) correlations (*left*), and detail (*right*)

anticancer activities, because the most active compound (Entry **1**) is taken as reference molecule with vector properties <111111>.

For the anticancer activities the partial correlation diagrams (Fig. 2) show that, on going from the whole set of 29 (*left*) to the 10 most active (*right*) stilbenes, the percentage of high correlations increases from 40 to 80%, and the percentage of medium correlations augments from 8 to 20%, while both low and *zero* correlations disappear.

MolClas molecular classification based on the equipartition conjecture of entropy production

The methods applied in this subsection are to perform molecular classifications based on the vector of properties by entropy (information theory) and to select the number of classes by the Equipartition Conjecture of Entropy Production.

The *grouping rule* is the following: *Two molecules are assigned to the same class if* r≥b*, where* b *is the classification level*. A comparative analysis of the molecular dataset, from 29 classes (each compound in its own class) to one class (containing all compounds), by the method of information entropy theory, matching $<i_1,i_2,i_3,i_4,i_5,i_6>$ and classification at level b (C_b), is calculated for anticancer activities (Csuk et al. 2012) and summarized in Table 4.

Table 4 displays the periodic classification of the stilbenes. The most active classes appear located at the bottom (greater i_5 and i_6) as well as, especially, at the right side (greater i_1, i_2, i_3 and i_4) of the table. Therefore, the most active Class 1 is located at the bottom right, and the least potent Class 7 is placed at the top left.

Table 4 lists the structures of the different classes by anticancer activity. For instance, the most active Class 1 presents a hydroxyl group at position R_3, a methoxyl group at site R_4 and a fluorine atom located at R_6 in ring A, as well as two methoxyl groups in *meta*, and a fluorine or hydrogen atom at R'_4 in ring B. For some choices of the substituents we have no experimental data. For instance, for the periodic classification in seven classes that we used we have some *holes* for the vector of properties 000100, 000101, 000110, 000111, etc. However, as in this vector the properties are in decreasing order of importance, we consider the most important stilbenes those with a structure close to the reference molecule 111111, *e.g.*, 111101, with one F atom in position 6, and one hydroxyl or methoxyl group in position 3 in *ortho* with one methoxyl in ring A, as well as two methoxyl groups in *meta* or *para*, and one F atom in ring B. We propose these substituents for a future study of (*Z*)-stilbenes like derivatives of combretastatin A4.

Table 5 reports the structures of the different classes by anticancer activity from the highest to the lowest order. The structural features that contribute to high cytotoxicity against human cancer cell lines, without causing cytotoxicity against non-malignant cells NiH3T3, are an OH group in position 3, a CH_3O group in position 4 and an F atom in position 6 in ring A, as well as an F or H atom in position 4', and two CH_3O groups in *meta* in ring B.

The information entropy and classification level b for different numbers of classes are listed in Table 6.

The combinatorial explosion of classifications (Table 6) is solved by the Equipartition Conjecture of Entropy Production. The best classification is that which is closest to the equipartition line (trend line) in the h–b plot (*cf*. Fig. 3): the classification with seven classes. Thus, the grouping rule for the classification level $0.94 \le b \le 0.96$ with the related entropy $h(\mathbf{R}_b) = 19.5243$ bits allows for the seven classes (Table 6), which are grouped from Class 1 (vector property 11111X, X = 0, 1) to Class 7 (vector property 000001) in Table 4. The 29 stilbenes are comprised in the seven groups in this classification.

Table 4 sums up the periodic classification of the activities by utilizing the technique founded on the information entropy theory, in artificial intelligence. The former four features were chosen to mean the group or column, and the latter two characteristics were utilized to designate the period or row in the table of periodic classification. Stilbenes in identical group show alike activities. Besides, molecules in the equal period too present all-out similarity. In this work the stilbenes are linked to the experimental information of cytotoxicity for all the nine lines, obtained from the technical literature (Csuk et al. 2012). The antitumor properties enlarge on going right across a period and enhance when falling away in a group. Those stilbenes with the utmost potency (Class 1, compounds **1** and **2**) are grouped into identical class, according to compounds with the moiety of Fig. 4.

Furthermore, stilbenes with enough potency (Classes 2–4) are clustered into additional groupings. Lastly the groups with the smallest antitumor properties are stilbenes situated at the left side of the table (Classes 6–7). Generally, the calculations correspond to Fig. 2, since couples of molecules in identical class with equal vector properties $<i_1,i_2,i_3,i_4,i_5,i_6>$ present red lines, signifying high partial correlations, *e.g.*, the pair (**1**, **2**) in Class 1.

Principal component analysis for the classification of the stilbene compounds

The method applied in this subsection is to perform a principal component analysis with anticancer activities and

Table 4 Periodic classification of the stilbene derivatives by information entropy method

$i_5,i_6/i_1,i_2,i_3,i_4$	0000	1000	1100	1110	1111
01	Class 7 15,16,17,18,24,25,26,28,29		Class 4 8,9		
1X (X = 0, 1)	Class 6 19,22,23 20,21,27	Class 5 10,11,12	Class 3 14 13	Class 2 4,7 3,5,6	Class 1 2 1

Table 5 Structures of the different classes by anticancer activity from the highest to the lowest order

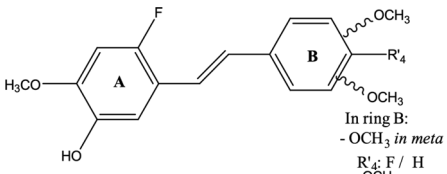
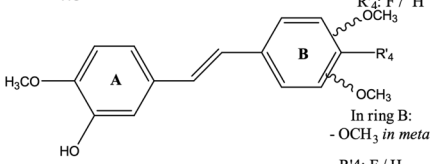
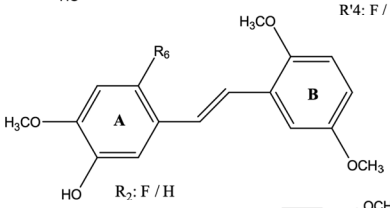
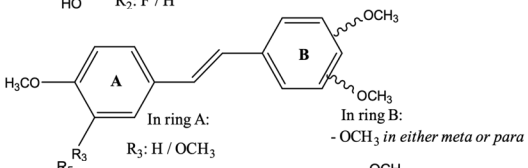
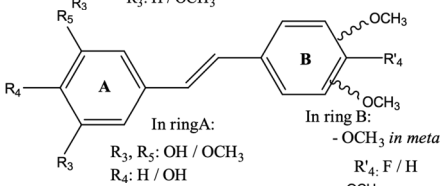
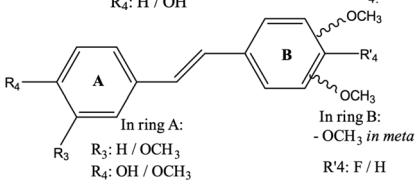
Vector properties	Class	Structures
11111X (X = 0, 1)	1	
11101X (X = 0, 1)	2	
11001X	3	
110001	4	
10001X	5	
000010	6	
000001	7	Absence of more than 2 OMe in ring B with OH/OMe in ring A

Table 6 Entropy and classification level *b* for different numbers of classes

Level *b*	Entropy *h* (bits)	No. classes
1.00	287.0306	29
0.98	45.1948	11
0.96	19.5243	7
0.93	10.1888	5
0.90	6.7947	4
0.76	3.8735	3
0.54	1.5463	2
0.04	0.4301	1
0.01	0.0805	1

ChEMBL data, to separate clusters of stilbenes and groups of variables.

Once getting the classification of the nine pIC_{50} of the stilbenes, on the one hand a principal component analysis (PCA) principal component (PC)1–PC2 scores plot was completed (*cf.* Fig. 5) with the cytotoxicity for all the nine lines (Table 3), and three clusters were found. The PCA was carried out to decrease the number of original variables to a smaller set of principal components (PCs), to find a global change of the molecules and to recognize behavioral models. The total variance explained by PC1 and PC2 is 98.35% (PC1 50.28%, PC2 48.07%).

The PCA PC1–PC2 scores plot (Fig. 5) permits splitting all the pairs of clusters by straight lines, guaranteeing a suitable grouping into three clusters. Notwithstanding it is unlikely to divide Cluster 2 into Classes 2, 3 and 4, as well as to rip Cluster 3 into Classes 6 and 7 by straight lines. Consequently, the grouping into clusters is more important than that into classes.

Cluster 1 (Fig. 5) matches with the greatest potent molecules (**1**–**3**; PC1 > –0.5, $PC_2 > 2$) with the structure of Fig. 6. This structure is in accordance with our classification (Table 4 and Fig. 3).

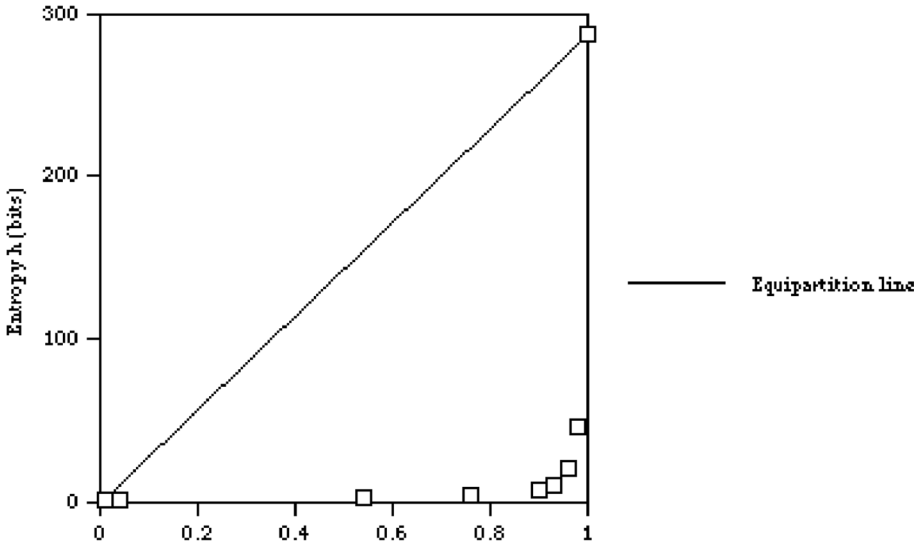


Fig. 3 Entropy *h* versus level *b* for stilbenes classifications

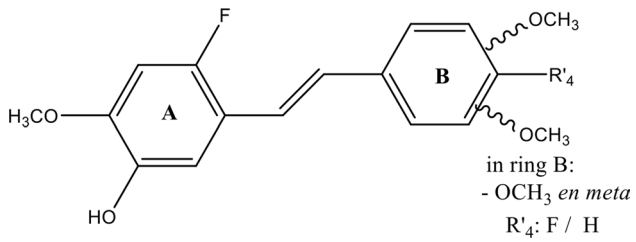


Fig. 4 Structure of stilbenes with the greatest activity

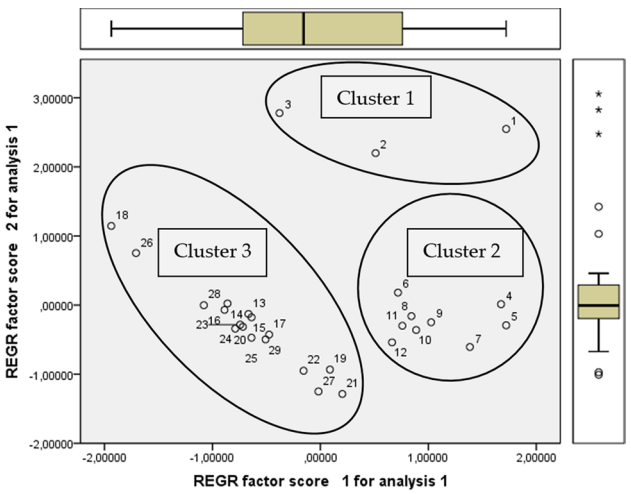


Fig. 5 Scores plot for 29 stilbenes with the cytotoxicity for all the nine lines

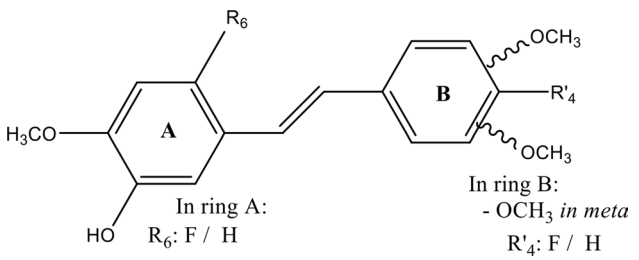


Fig. 6 Structure of the most active stilbenes

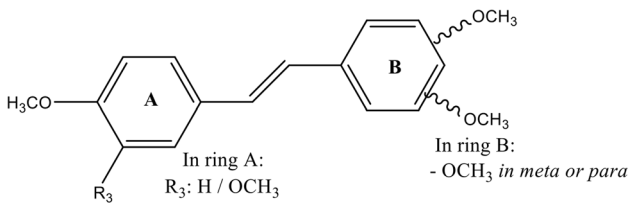


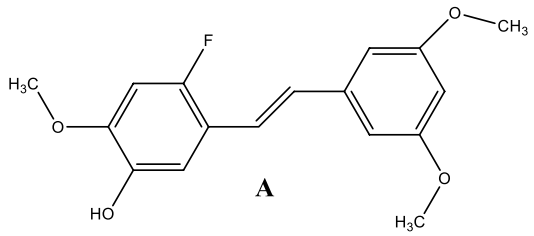
Fig. 7 Structure of the stilbenes with intermediate bioactivity

Cluster 2 (Fig. 5) corresponds to the compounds of intermediate bioactivity (**4**–**12**, Classes 2, 3 and 4; $PC_1 > 0.5$, $PC_2 < 0.5$), with the structure of Fig. 7.

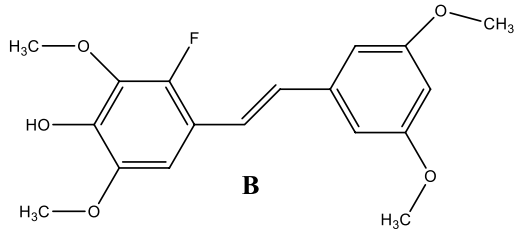
Cluster 3 (Fig. 5) presents the least active compounds (**13**–**29**, Classes 5, 6 and 7; $PC_1 < 0.5$, $PC_2 < 1.5$).

The finding that there are *holes* in the periodic classification of (*E*)-stilbenes (Table 4) and the three clusters observed in the PCA scores plot (Fig. 5) suggests us to propose new highly active compounds.

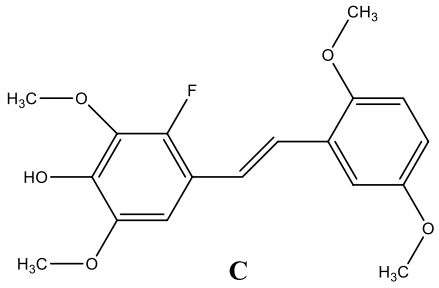
In Fig. 8, we propose new highly active compounds that are not described in the ChEMBL database and that have not been used in this study. Compounds **A** and **B** have the same vector of properties <111111> and would be found in



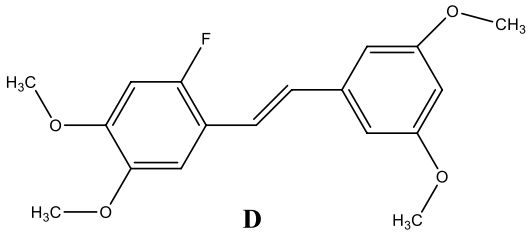
6-Fluoro-3-hydroxy-3',4,5'-trimethoxystilbene



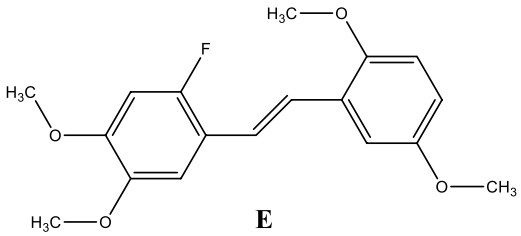
6-Fluoro-4-hydroxy-3,3',5,5'-tetramethoxystilbene



6-Fluoro-4-hydroxy-2',3,5,5'-tetramethoxystilbene



6-Fluoro-3,3',4,5'-tetramethoxystilbene



6-Fluoro-2',3,4,5'-tetramethoxystilbene

Fig. 8 Proposed active stilbenes not described in ChEMBL and not used in this study

Class 1 in Table 4, together with compounds **1** and **2**, which are the most active used in the study. Compound **C** with the vector of properties <111011> would appear in Class 2 in Table 4, as well as compounds **D** and **E** with the vector of properties <110011> would appear in Class 3 in Table 4. It should be noticed that the most active classes appear below and to the right in the periodic table of classification. All the structures are not available as natural compounds and are suggested here for future synthesis.

On the other hand, we describe in a loading plot (*cf.* Fig. 9) the behavior of the 23 variables taken from database ChEMBL (*cf.* Supplementary Material Table S1), the energy (kcal/mol) obtained by software calculation with MMFF94-Chem3D minimization and the cytotoxicity for all the nine lines. The properties furthest from the origin (0.0, 0.0) are the most important for describing PCs, and those closest to the origin are the least important.

The PCA PC1-PC2 loading plot (Fig. 9) displays the cytotoxicity for all the nine lines, placed together with positive loadings for both PC1 and PC2, which suggests that they would depend in part on the same variables.

Principal component PC1 (37.29% of the total variance) shows positive loading mainly with acd_logp, acd logd (pH 7), acd logd (pH 5), ACD/KOC (pH 7), ACD/KOC (pH 5), ACD/BCF (pH 7), ACD/BCF (pH 5), molar volume (MV), free rotations (FR), polarizability, molar refraction, number of methoxyl groups, fluorine number, boiling point (BP), enthalpy of vaporization (EV), Energy and hydrogen bond acceptor (HBA), as well as pIC_{50} values of the cytotoxicity for all the nine lines, while negative loading with surface tension (ST), density, refraction index (IR), Hydroxyl number, hydrogen bond donor (HBD), flash point (FP) and polar surface (PS).

On the other hand PC2 (35.09% of the total variance) shows positive loading, mainly with acd_logp, acd logd (pH 7), acd logd (pH 5), ACD/KOC (pH 7), ACD/KOC (pH 5), ACD/BCF (pH 7), ACD/BCF (pH 5), molar volume (MV), free rotations (FR), polarizability, molar refraction, number

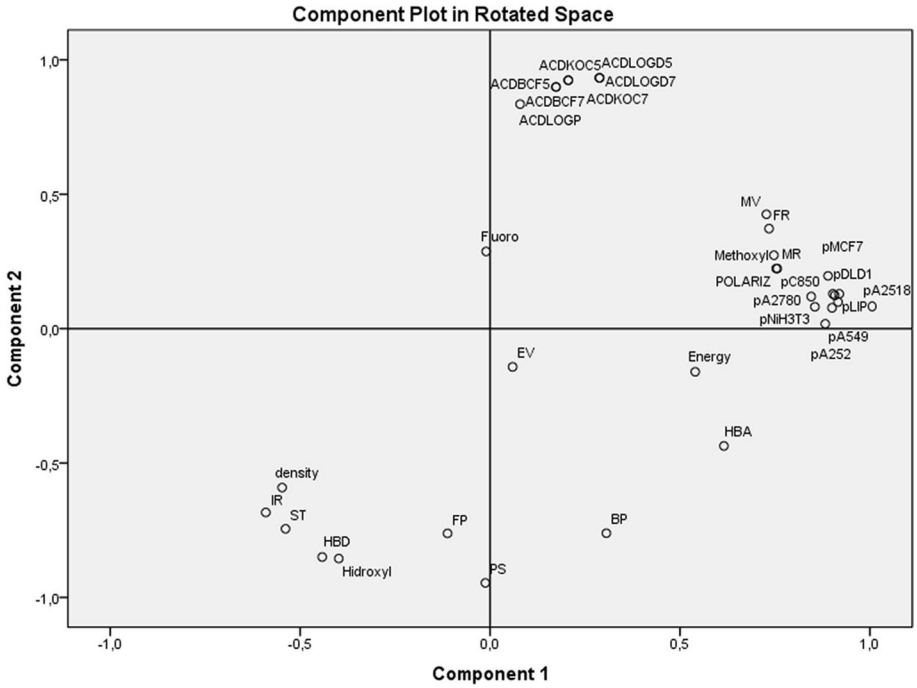


Fig. 9 Loading plot for 24 variables and cytotoxicity for all the nine lines for stilbenes

of methoxyl groups, fluorine number and the pIC_{50} values of the cytotoxicity for all the nine lines, while negative loading with surface tension (ST), density, refraction index (IR), hydroxyl number, hydrogen bond donor (HBD), flash point (FP), polar surface (PS), enthalpy of vaporization (EV), Energy, hydrogen bond acceptor (HBA) and boiling point (BP).

Moreover, pIC_{50} values show positive loading with MV, FR, MR, number of methoxyl groups, polarizability, acd_logp, acd logd (pH 7), acd logd (pH 5), ACD/KOC (pH 7), ACD/KOC (pH 5), ACD/BCF (pH 7) and ACD/BCF (pH 5), while negative loading with surface tension (ST), density, refraction index (IR), Hydroxyl number, hydrogen bond donor (HBD), flash point (FP) and polar surface (PS).

The finding that the loadings plot shows the cytotoxicity for all the nine lines placed together suggests us to carry out a QSAR study, to test if they depend in part on the same variables.

Quantitative structure–activity relationships for stilbenes

The methods applied in this subsection are to obtain QSAR models to quantify previous SAR models, to carry out an analysis of the sign of the correlations of the models, to obtain the applicability domain of the models with Williams plots, to carry out experimental *vs*. observed plots to analyze the individual residuals and to carry out the leave-one-out cross-validation to test the robustness of the models.

A multiple linear regression model approach was adopted to determine the quantitative importance of the combined presence of some of the 24 descriptors: acd_logp, acd logd (pH 7), acd logd (pH 5), ACD/KOC (pH 7), ACD/KOC (pH 5), ACD/BCF (pH 7), ACD/BCF (pH 5), molar volume (MV), free rotations (FR), polarizability, molar refraction, number of methoxyl groups, fluorine number, boiling point (BP), enthalpy of vaporization (EV), Energy, hydrogen bond acceptor (HBA), Surface tension (ST), density, refraction index (IR), Hydroxyl number, hydrogen bond donor (HBD), flash point (FP) and polar surface (PS), as well as five easily derived. These properties, except the energy that was calculated for each compound, were taken from the ChEMBL database to explain such pIC_{50} of the cytotoxicity for all the nine lines: 518A2 (melanoma), 850C (anaplastic thyroid), A253 (head), A549, (lung), DLD1 (colon), LIPO (liposarcoma), MCF7 (breast), A2780 (ovarian) and non-malignant mouse fibroblasts (NiH3T3). The fits were checked with the statistics correlation coefficient r, the coefficient of determination r^2, the adjusted r^2, the standard deviation s, Fisher's ratio F, as well as cross-validated correlation coefficient q and coefficient of determination q^2 in leave-1-out. We have included Gramatica's measures r^2, adjusted r^2, mean absolute error (MAE), root mean square error (RMSE), Y-scrambling determination coefficient r^2Yscr., q^2 and Y-scrambling cross-validated determination coefficient q^2Yscr. (Gramatica 2013). We selected the models with a high statistical signification but having as few parameters as possible. To avoid overfitting or chance correlation, the ratio between the cases and the variables in the equation should be between four and five as minimum (Castillo-Garit et al. 2012, 2015). In this work models with a ratio (number of cases)/(number of variables) lower than four were rejected. The equations

of the models between the homologous series of compounds and the properties follow:

$$\begin{aligned}
\mathrm{pIC_{50}A518A2} &= -(7\pm4) + (6.4\pm2.4)\ \text{Density}\ (\mathrm{g/cm^3}) \\
&\quad + (0.55\pm0.14)\,(\text{Methoxyl})^2 - (0.12\pm0.03)\,(\text{Methoxyl})^3 \\
&\quad + (0.07\pm0.03)\,(\text{acd_logp})^2 + (0.012\pm0.005)\ \text{Energy} \\
&\quad (\text{kcal/mol}) + (24\pm15)\,(\text{MV}/1000)^2\ (\mathrm{cm^6}) \\
N = 29 &\quad r = 0.806 \quad r^2 = 0.650 \quad \text{Adjusted } r^2 = 0.554 \\
&\quad \text{MAE} = 0.3500 \quad \text{RMSE} = 0.4563 r^2\text{Yscr.} = 0.2175 \\
&\quad s = 0.524 \quad F = 6.802 \quad q = 0.669 \quad q^2 = 0.448 \\
&\quad q^2/r^2 = 68.9\% \quad q^2\text{Yscr.} = -0.6034
\end{aligned} \tag{1}$$

$$\begin{aligned}
\mathrm{pIC_{50}C850} &= -(15\pm6) + (8\pm3)\,\text{Density}\,(\mathrm{g/cm^3}) + (0.8\pm0.3)\,(\text{acd_logp}) \\
&\quad + (0.53\pm0.14)\ (\text{Methoxyl})^2 - (0.12\pm0.03)\ (\text{Methoxyl})^3 \\
&\quad + (0.14\pm0.09)\ (\text{Polarizability})\ (\mathrm{cm^3}) - (0.010\pm0.005) \\
&\quad (\text{FP})\ (^\circ\text{C}) + (0.011\pm0.005)\ (\text{BP})\ (^\circ\text{C}) \\
N = 29 &\quad r = 0.811 \quad r^2 = 0.658 \quad \text{Adjusted } r^2 = 0.543 \\
&\quad \text{MAE} = 0.3350 \quad \text{RMSE} = 0.4489 \quad r^2\text{Yscr.} = 0.2524 \\
&\quad s = 0.528 \quad F = 5.756 \quad q = 0.575 \quad q^2 = 0.331 \\
&\quad q^2/r^2 = 50.3\% \quad q^2\text{Yscr.} = -0.7302
\end{aligned} \tag{2}$$

$$\begin{aligned}
\mathrm{pIC_{50}A253} &= -(7\pm5) + (5.8\pm2.4)\ \text{Density}\,(\mathrm{g/cm^3}) \\
&\quad + (0.48\pm0.14)\,(\text{Methoxyl})^2 - (0.10\pm0.03)\,(\text{Methoxyl})^3 \\
&\quad + (0.07\pm0.03)\,(\text{acd_logp})^2 + (0.009\pm0.005) \\
&\quad \text{Energy}\,(\text{kcal/mol}) + (0.04\pm0.03)\,(\text{MR})\,(\mathrm{cm^3}) \\
N = 29 &\quad r = 0.757 \quad r^2 = 0.573 \quad \text{Adjusted } r^2 = 0.456 \\
&\quad \text{MAE} = 0.3233 \quad \text{RMSE} = 0.4340 \quad r^2\text{Yscr.} = 0.2140 \\
&\quad s = 0.498 \quad F = 4.908 \quad q = 0.601 \quad q^2 = 0.361 \\
&\quad q^2/r^2 = 63.0\% \quad q^2\text{Yscr.} = -0.6274
\end{aligned} \tag{3}$$

$$\begin{aligned}
\mathrm{pIC_{50}A549} &= -(11\pm5) + (9\pm3)\,\text{Density}\,(\mathrm{g/cm^3}) \\
&\quad + (0.46\pm0.14)\,(\text{Methoxyl})^2 - (0.11\pm0.03)\,(\text{Methoxyl})^3 \\
&\quad + (0.07\pm0.04)\,(\text{MR})\,(\mathrm{cm^3}) + (0.014\pm0.005) \\
&\quad \text{Energy}\,(\text{kcal/mol}) - (0.011\pm0.004)\,(\text{FP})\,(^\circ\text{C}) \\
N = 29 &\quad r = 0.777 \quad r^2 = 0.604 \quad \text{Adjusted } r^2 = 0.495 \\
&\quad \text{MAE} = 0.3304 \quad \text{RMSE} = 0.4371 \quad r^2\text{Yscr.} = 0.2144 \\
&\quad s = 0.502 \quad F = 5.577 \quad q = 0.658 \quad q^2 = 0.433 \\
&\quad q^2/r^2 = 71.7\% \quad q^2\text{Yscr.} = -0.5143
\end{aligned} \tag{4}$$

$$\begin{aligned}
\mathrm{pIC_{50}DLD1} &= -(6\pm3) + (7.0\pm2.2)\,\text{Density}\,(\mathrm{g/cm^3}) \\
&\quad + (0.48\pm0.13)\,(\text{Methoxyl})^2 - (0.12\pm0.03)\,(\text{Methoxyl})^3 \\
&\quad + (0.013\pm0.005)\,\text{Energy}\,(\text{kcal/mol}) - (0.013\pm0.004)\,(\text{FP})\,(^\circ\text{C}) \\
&\quad + (0.007\pm0.004)\ (\text{BP})\,(^\circ\text{C}) + (100\pm40)\,(\text{MV}/1000)^3\,(\mathrm{cm^9}) \\
N = 29 &\quad r = 0.834 \quad r^2 = 0.696 \quad \text{Adjusted } r^2 = 0.593 \\
&\quad \text{MAE} = 0.2882 \quad \text{RMSE} = 0.4017 \quad r^2\text{Yscr.} = 0.2515 \\
&\quad s = 0.472 \quad F = 6.835 \quad q = 0.738 \quad q^2 = 0.545 \\
&\quad q^2/r^2 = 78.3\% \quad q^2\text{Yscr.} = -0.8215
\end{aligned} \tag{5}$$

$$\begin{aligned}
\mathrm{pIC_{50}LIPO} &= -(11\pm4) + (8.0\pm2.2)\,\text{Density}\,(\mathrm{g/cm^3}) \\
&\quad + (0.57\pm0.13)\,(\text{Methoxyl})^2 - (0.13\pm0.03)\,(\text{Methoxyl})^3 \\
&\quad + (0.010\pm0.005)\,\text{Energy}\,(\text{kcal/mol}) - (0.007\pm0.004)\,(\text{FP})\,(^\circ\text{C}) \\
&\quad + (1.1\pm0.3)\,(\text{acd_logp}) + (0.7\pm0.3)\ (\text{HBA}) \\
N = 29 &\quad r = 0.840 \quad r^2 = 0.706 \quad \text{Adjusted } r^2 = 0.607 \\
&\quad \text{MAE} = 0.3107 \quad \text{RMSE} = 0.4011 \quad r^2\text{Yscr.} = 0.2503 \\
&\quad s = 0.471 \quad F = 7.179 \quad q = 0.730 \quad q^2 = 0.533 \\
&\quad q^2/r^2 = 75.5\% \quad q^2\text{Yscr.} = -0.5337
\end{aligned} \tag{6}$$

$$\begin{aligned}
\mathrm{pIC_{50}NiH3T3} &= -(7\pm3) + (7.4\pm2.0)\,\text{Density}\,(\mathrm{g/cm^3}) \\
&\quad + (0.40\pm0.12)\,(\text{Methoxyl})^2 - (0.10\pm0.03)\,(\text{Methoxyl})^3 \\
&\quad + (0.012\pm0.004)\,\text{Energy}\,(\text{kcal/mol}) \\
&\quad - (0.010\pm0.004)\,(\text{FP})\,(^\circ\text{C}) \\
&\quad + (0.12\pm0.04)\,(\text{acd_logp})^2 + (0.71\pm0.22)\,(\text{HBA}) \\
N = 29 &\quad r = 0.807 \quad r^2 = 0.652 \quad \text{Adjusted } r^2 = 0.536 \\
&\quad \text{MAE} = 0.2760 \quad \text{RMSE} = 0.3645 \quad r^2\text{Yscr.} = 0.2524 \\
&\quad s = 0.428 \quad F = 5.614 \quad q = 0.649 \quad q^2 = 0.421 \\
&\quad q^2/r^2 = 64.6\% \quad q^2\text{Yscr.} = -0.5032
\end{aligned} \tag{7}$$

$$\begin{aligned}
\mathrm{pIC_{50}MCF7} &= -(7\pm4) + (5.2\pm1.8)\,\text{Density}\,(\mathrm{g/cm^3}) \\
&\quad + (0.48\pm0.11)\,(\text{Methoxyl})^2 - (0.10\pm0.03)\,(\text{Methoxyl})^3 \\
&\quad + (0.008\pm0.004)\,\text{Energy}\,(\text{kcal/mol}) \\
&\quad + (0.9\pm0.3)\,(\text{acd_logp}) + (0.28\pm0.19)\,(\text{HBA}) \\
N = 29 &\quad r = 0.820 \quad r^2 = 0.673 \quad \text{Adjusted } r^2 = 0.584 \\
&\quad \text{MAE} = 0.2864 \quad \text{RMSE} = 0.3703 \quad r^2\text{Yscr.} = 0.2167 \\
&\quad s = 0.425 \quad F = 7.540 \quad q = 0.719 \quad q^2 = 0.517 \\
&\quad q^2/r^2 = 76.8\% \quad q^2\text{Yscr.} = -0.4229
\end{aligned} \tag{8}$$

$$\begin{aligned}
\mathrm{pIC_{50}A278} &= -(11\pm4) + (10\pm3)\,\text{Density}\,(\mathrm{g/cm^3}) \\
&\quad + (0.57\pm0.15)\,(\text{Methoxyl})^2 - (0.13\pm0.03)\,(\text{Methoxyl})^3 \\
&\quad + (0.015\pm0.006)\,\text{Energy}\,(\text{kcal/mol}) - (0.014\pm0.005)\,(\text{FP})\,(^\circ\text{C}) \\
&\quad + (0.14\pm0.06)\,(\text{acd_logp})^2 + (0.8\pm0.3)\,(\text{HBA}) \\
N = 29 &\quad r = 0.810 \quad r^2 = 0.656 \quad \text{Adjusted } r^2 = 0.541 \\
&\quad \text{MAE} = 0.3594 \quad \text{RMSE} = 0.4807 \quad r^2\text{Yscr.} = 0.2487 \\
&\quad s = 0.565 \quad F = 5.710 \quad q = 0.678 \quad q^2 = 0.460 \\
&\quad q^2/r^2 = 70.1\% \quad q^2\text{Yscr.} = -0.5090
\end{aligned} \tag{9}$$

In the QSARs for stilbenes all the nine fits [Eqs. (1)–(9)] are good, corresponding to coefficients of determination r^2 greater than 0.57, standard derivations lesser than 0.565 and Fisher's ratios greater than 4.908.

Overall, the independent variables showing positive correlations are Density, Methoxyl^2, acd_logp^2, Energy, MV^2, acd_logp, Polarizability, BP, MR, MV^3 and HBA, while the descriptor with negative associations results Methoxyl^3, followed by FP. The repetition of many variables in most fits agrees with the result of the loading plot, showing the cytotoxicity for all the nine lines placed together, and suggests the importance of considering the inclusion of the 13 (or a least seven) descriptors in forthcoming studies of other cell lines.

Table 7 shows the sign of the positive and negative correlations of $\mathrm{pIC_{50}}$ of the nine dependent variables (cytotoxicity), with the independent variables. In all the cases all the symbols of the descriptors are the same for each different cytotoxicity. The variables that contribute to high cytotoxicity against cancer cell lines in Eqs. (1)–(6), (8) and (9) without causing cytotoxicity against non-malignant cells in Eq. (7) are acd_logp, MV^2, polarizability, BP, MR and MV^3.

The good statistics for all the nine fits suggest us to carry out Williams plots to test the applicability domain of the models.

The applicability domain of all the nine models (1)–(9) is analyzed by Williams plots, which are the chart of cross-validated standardized residuals *vs*. leverage (Hat diagonal) values (k). As an example, Fig. 10 shows the Williams plot

Table 7 Sign of the positive and negative correlations of the pIC_{50} values with the independent variables

	pIC_{50} A518A2	pIC_{50} C850	pIC_{50} A253	pIC_{50} A549	pIC_{50} DLD1	pIC_{50} LIPO	pIC_{50} NiH3T3	pIC_{50} MCF7	pIC_{50} A278
Density	+	+	+	+	+	+	+	+	+
$Methoxyl^2$	+	+	+	+	+	+	+	+	+
$Methoxyl^3$	–	–	–	–	–	–	–	–	–
acd_logp^2	+		+				+		+
acd_logp		+				+		+	
$(MV/1000)^2$	+								
Energy	+		+	+	+	+	+	+	+
Polarizability		+							
BP		+			+				
MR			+	+					
$(MV/1000)^3$					+				
HBA						+	+	+	+
FP		–		–	–	–	–		–

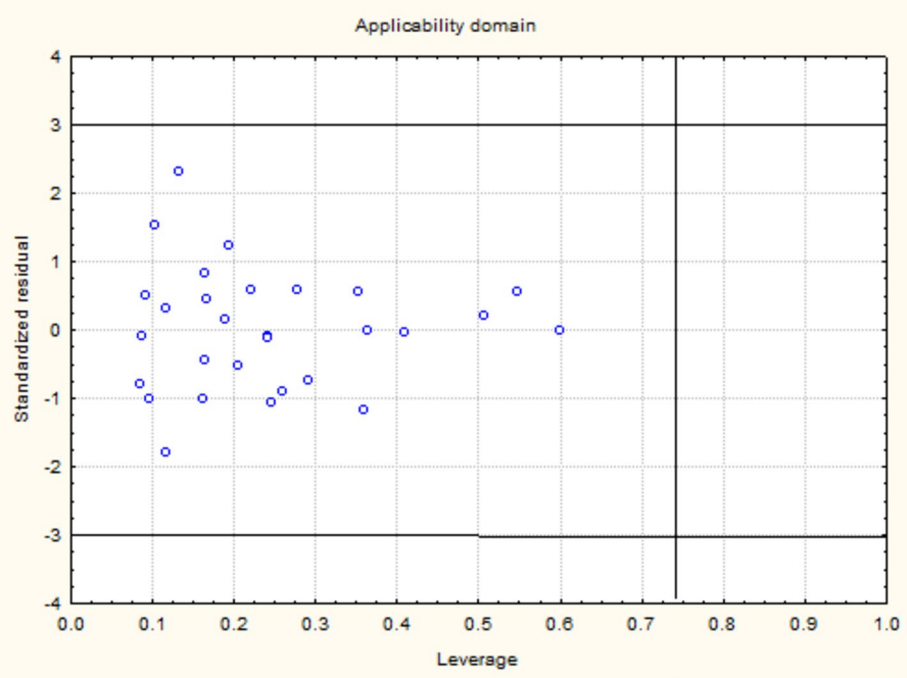


Fig. 10 Williams plot for the graphic display of outliers for the response (on *Y*-axis: standardized residuals > 3 *s*) or for the structure (on *X*-axis: highest Hat value > *h** cut-off line), in the pIC_{50} MCF7 regression model [Eq. (8)]

for the pIC_{50} MCF7 regression model [Eq. (8)], with no response outlier (cross-validated standardized residual > 3 *s*) and no structurally influential chemical ($h > h^*$). In Eqs. (1)–(5) there is no response outlier, and the structurally influential chemical is compound **23** (*cf.* Supplementary Material Figure S1). In Eqs. (6)–(9) there is no response outlier and no structurally influential chemical.

The good statistics for the applicability domains (Williams plots) of all the nine models suggests us to plot the expected *vs*. observed activities to test, especially, compound **23**.

The scatterplots of all the nine expected *vs*. observed activities are carried out. As an example, Fig. 11 displays the plot for the pIC_{50} MCF7 regression model [Eq. (8)], with all the points (including compound **23**) close to the main diagonal ($y = x$). In Eqs. (1)–(9) the points are close to the main diagonal (*cf.* Supplementary Material Figure S2). Compound **23** is close to the diagonal so its predictions must be taken for granted, and it cannot be considered an outlier, in agreement with the fact that it is not a response outlier (cross-validated standardized residual > 3 *s*) in Williams plots (Fig. 10 and Supplementary Material Figure S1).

The good statistics for the scatterplots of predicted *vs*. observed activities, for all the nine models, suggests us to plot the regression standardized residuals expected *vs*. observed activities to test, especially, compound **23**.

The scatterplots of all the nine regression standardized residuals expected *vs*. observed activities are carried out. As an example, Fig. 12 displays the plot for the pIC_{50} MCF7 regression model [Eq. (8)], with all the points (including compound **23**) close to the main diagonal ($y = x$). In Eqs. (1)–(9) the points are close to the main diagonal (*cf.*

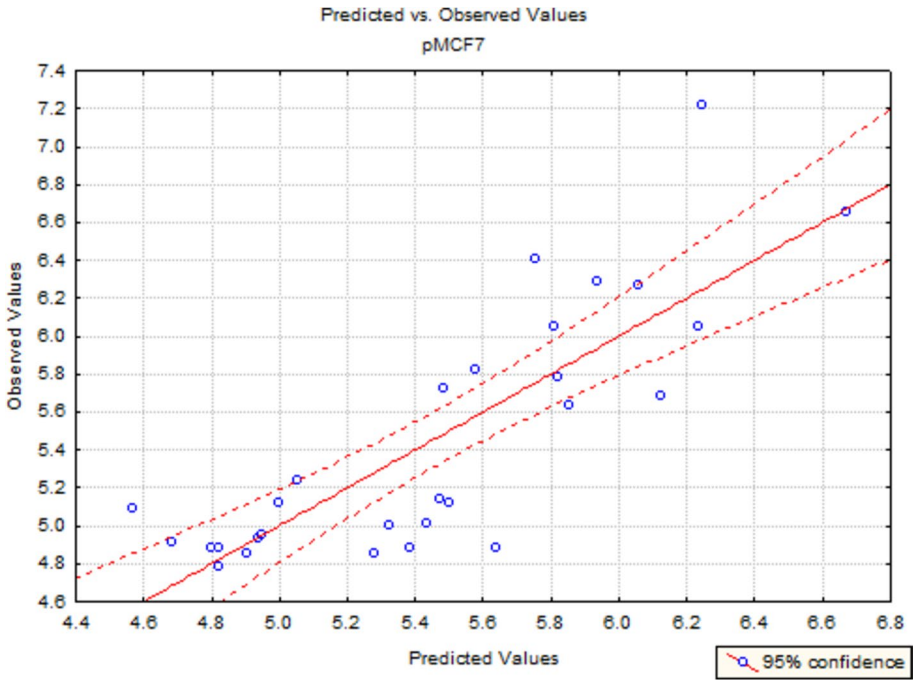


Fig. 11 Scatterplot of predicted *vs*. observed activities for the pIC_{50} MCF7 model [Eq. (8)]

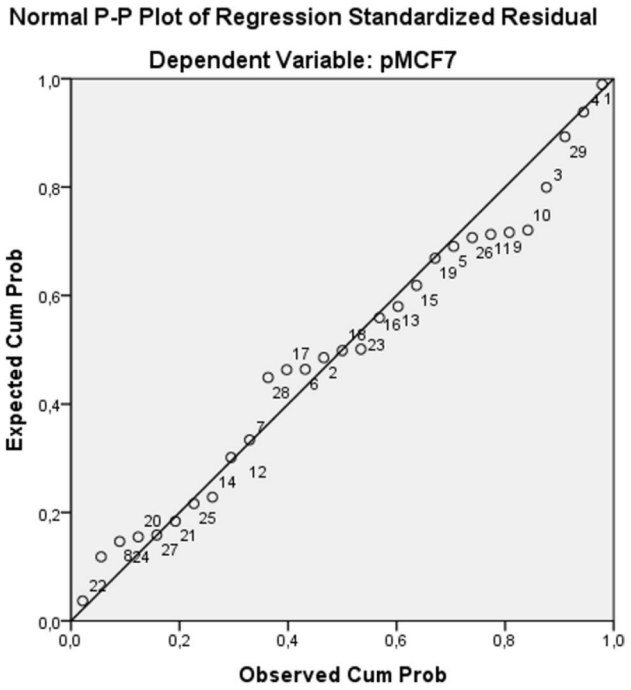


Fig. 12 Scatterplot of regression standardized residuals predicted *vs*. observed activities for the pIC_{50} MCF7 model [Eq. (8)]

Supplementary Material Figure S3). Compound **23** is close to the diagonal so its predictions must be taken for granted, and it cannot be considered an outlier, in agreement with the fact that it is not a response outlier in Williams plots and predicted *vs*. observed activities (Figs. 10 and 11, as well as Supplementary Material Figures S1 and S2).

The good statistics for the nine QSAR models for the set of 29 stilbenes suggest us to perform a cross-validation, to test the robustness of the equations for external stilbenes.

Leave-many-out cross-validation for QSAR robustness

The method applied in this subsection is to carry out the leave-many-out (leave-*m*-out) cross-validation (CV), to test the robustness of the models.

Leave-*m*-out ($1 \le m \le 13$) cross-validated coefficients of determination r_{cv}^2 calculated for stilbenes [$q^2 = r_{cv}^2$ ($m = 1$), *cf*. Table 8] indicate that r_{cv}^2 decays with *m*, except for pIC_{50} A518A2 and pIC_{50} C850 [Eqs. (1) and (2)], which denotes possible outliers.

Table 8 shows eight out of nine models [Eqs. (1) and (3)–(9)], with $q^2 = r_{cv}^2$ ($m = 1$) greater than 0.36 and good robustness for stilbenes not included in the fits. The different CV q^2 values indicate diverse robustness of models (1)–(9). For models (5), (6) and (8) q^2 is above 0.5, and the best CV results were obtained. Models (1), (4), (7) and (9) are good with CV values above 0.4. Finally, models (2) and (3) present a lower quality of CV fit but the CV value is above 0.3.

The good statistics for the cross-validation for the models suggests us to carry out an external validation, with a set of compounds that have not been included in the fits.

External validation for QSAR models

The method applied in this subsection is to carry out external validation to analyze the applicability for external stilbenes.

Table 8 Cross-validated coefficient of determination *in* leave-*m*-out for stilbenes

m	Equation (1)	Equation (2)	Equation (3)	Equation (4)	Equation (5)	Equation (6)	Equation (7)	Equation (8)	Equation(9)
1	0.448	0.331	0.361	0.433	0.545	0.533	0.421	0.517	0.460
2	0.448	0.332	0.361	0.432	0.543	0.531	0.420	0.517	0.458
3	0.449	0.332	0.361	0.430	0.540	0.530	0.420	0.516	0.457
4	0.449	0.333	0.360	0.429	0.539	0.527	0.419	0.514	0.456
5	0.449	0.334	0.360	0.427	0.536	0.526	0.419	0.513	0.454
6	0.450	0.335	0.359	0.425	0.533	0.524	0.417	0.511	0.453
7	0.450	0.335	0.359	0.423	0.529	0.523	0.416	0.508	0.452
8	0.452	0.335	0.358	0.421	0.524	0.520	0.416	0.507	0.450
9	0.452	0.335	0.356	0.419	0.518	0.518	0.415	0.506	0.448
10	0.453	0.334	0.354	0.416	0.513	0.516	0.415	0.503	0.446
11	0.453	0.333	0.353	0.412	0.507	0.514	0.413	0.500	0.444
12	0.452	–	–	0.410	–	–	0.413	0.497	0.441
13	0.450	–	–	–	–	–	–	0.494	–

Table 9 Experimental and predicted values for the external validation set

Compound name	pIC_{50} NiH3T3		pIC_{50} MCF7		pIC_{50} DLD1		pIC_{50} A253		pIC_{50} A278	
	Exp.[a]	Pred.[b]	Exp.[a]	Pred.[b]	Exp.[a]	Pred.[b]	Exp.[a]	Pred.[b]	Exp.[a]	Pred.[b]
(*E*)-2′,5′-Dihydroxy-2-methoxystilbene	4.954	5.105	4.828	5.083	4.619	4.558	4.660	4.884	4.799	4.928
(*E*)-3′,5′-Dimethoxy-4′-fluoro-3-hydroxystilbene	4.928	5.254	5.245	5.269	4.755	5.346	4.683	5.206	4.550	5.241
(*E*)-6-Fluoro-4-methoxy-2′,3,5′-trihydroxystilbene	4.677	6.187	4.973	5.465	4.745	5.917	5.005	5.353	4.980	6.310
(*E*)-3′,5′-Dimethoxy-4′-fluoro-2-hydroxystilbene	4.830	5.953	4.887	5.852	4.596	5.488	–	–	–	–
(*E*)-3′,4,5′-Trihydroxystilbene	4.616	4.250	–	–	–	–	–	–	–	–

[a]Experimental value; [b]Predicted value for the correspondent model

Once the models were internally validated and their robustness, well established external validation can be performed. This is the best way to prove the ability of the models to predict new compounds. The procedure is performed by applying the equation of each model obtained with the training set, to a set of compounds that have never been used in the development of the model. Sometimes like in this case it is difficult to find new compounds with experimental reports for the desired activities. More difficult even when we are analyzing several activities in parallel. However, we were able to find a little set of compounds with experimental reports for several activities modeled in this work (activities with only one or two compounds were not taken into account). The predictions for five of the nine activities of the models, for these compounds and their experimental values, are shown in Table 9. As can be seen the predicted values are very close to the experimental values of the external set. This permits us to say that our models have good predictive power and can be used to predict the anticancer activities of other (*E*)-stilbenes.

Although the models predict a greater variability in the activities for compounds similar in the experiment, the purpose of our work is not to predict the most active substances, but to differentiate between active and inactive molecules to obtain SAR and QSAR rules of activity. The prediction methods developed here will also work on glycosyl-substituted stilbenes like isorhapontin, otherwise prenylated or geranylated stilbenes, which have diverse pharmacological activities. The glycosylation of bioactive compounds can enhance their water solubility, physicochemical stability, intestinal absorption and biological half-life, as well as improve their bio- and pharmacological properties. Therefore, the cytotoxicity (anticancer activities) can be predicted, but the final effect should also depend on the bioavailability and time of residence in the organism. We have not experimental data on anticancer activity in these cell lines and, since their ChEMBL properties will be different because of their structural characteristics, it is logical to think that our QSAR relationships would not be suitable for glycosylated compounds. However, we could be sure that ingested glycosylated stilbenes would be effective as anticancer agents, since they would be hydrolyzed in the intestine and transferred to cells, being as bioavailable as the methoxylated ones.

The limitation of this article is that the most active stilbenes are derivatives of combretastatin A4 (*Z*)-stilbenes, while the data of anticancer activity are available for the (*E*)-stilbenes studied here. Moreover, the knowledge of the effect of the substituents on the anticancer activity of the (*E*)-stilbenes would give us a better knowledge of the effect of the substituents on the anticancer activity of the stilbenes, in general, as a first step for a forthcoming study on the (*Z*)-stilbenes.

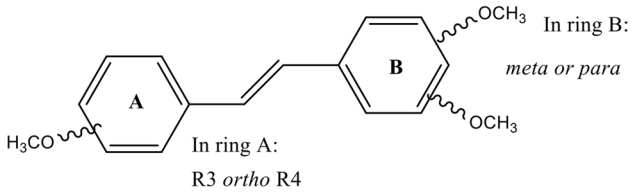


Fig. 13 Structural moiety corresponding to component i_1 of the vector of properties

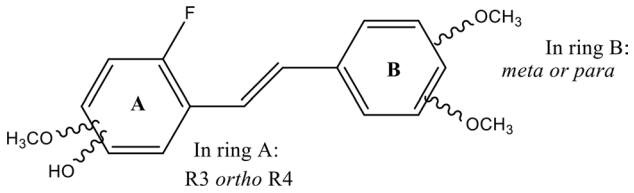


Fig. 14 Structural moiety corresponding to component i_2 of the vector of properties

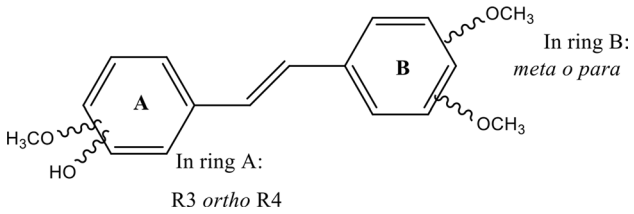


Fig. 15 Structure moiety corresponding to component i_3 of the vector of properties

Materials and methods

Program MolClas for molecular classification based on the equipartition conjecture of entropy production

The computational method is the same as we successfully applied to the classification of polyphenolic compounds (Castellano et al. 2014). The first step in quantifying the concept of similarity for stilbenes is to list the most important moieties, with respect to the anticancer activities of such molecules. Furthermore, the vector of properties $i=(i_1, i_2 \ldots i_k \ldots)$ should be associated with each stilbene i, whose components correspond to a number of characteristic groups in the molecule, in a hierarchical order according to the expected importance of their anticarcinogenic potency. The components $\boldsymbol{i_k}$ are either *1* or *0*, according to the experimental evidence of anticancer powers to structural variations in the stilbene. Index $\boldsymbol{i_1}=1$ indicates the presence of 2 OMe groups in *meta* or *para* in ring B, and 1 OMe in R_3/R_4 in *ortho* with OH/OMe in ring A (*cf.* Fig. 13).

The index $\boldsymbol{i_2}=1$ denotes the occurrence of F in R_6 and OH/OMe in R_3 in *ortho* with OMe in ring A, as well as the presence of 2 OMe groups in *meta* or *para* in ring B (*cf.* Fig. 14).

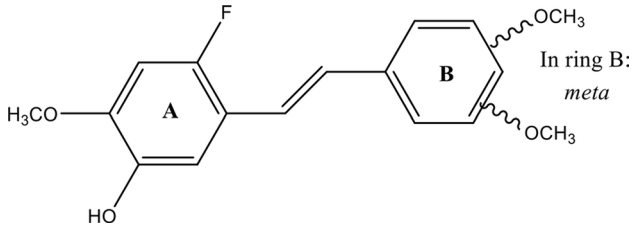


Fig. 16 Structural moiety corresponding to component i_4 of the vector of properties

The index $\boldsymbol{i_3}=1$ indicates the presence of 2 OMe groups in *meta* or *para* in ring B, and OH in R_3/R_4 in *ortho* with OMe in ring A (*cf.* Fig. 15).

The index $\boldsymbol{i_4}=1$ denotes the presence of F in R_6 and OH in R_3 in *ortho* with OMe in ring A, as well as the occurrence of 2 OMe groups in *meta* in ring B (*cf.* Fig. 16).

The index $\boldsymbol{i_5}=1$ means the absence of F in ring B, and the index $\boldsymbol{i_6}=1$ indicates the absence of more than 2 OMe in ring B with OH/OMe in ring A. The structures and vectors of properties of the 29 studied stilbenes are listed in Table 2.

The similarity matrix $\mathbf{R}=[r_{ij}]$ between two stilbenes $i=(i_1, i_2 \ldots i_k \ldots)$ and $j=(j_1, j_2 \ldots j_k \ldots)$ is defined by the correlation coefficient between them, modified by six weights $a_k=0.5$, 0.25, 0.125, etc. The grouping rule is: i *and* j *are assigned to the same class if* $r_{ij} \geq b$ *(classification level)*. Learning procedures similar to those found in stochastic methods are implemented as follows (White 1989). Consider a given partition into classes as *good* or ideal from practical observations. This corresponds to a reference similarity matrix $\mathbf{S}=[s_{ij}]$ obtained for an arbitrary number of properties. Next consider the same set of species as in the *good* classification and the actual properties.

The distance between any partition **R** and the *good* partition **S** is given by their corresponding matrix elements (Iordache 2011, 2012; Kullback 1959):

$$D = -\sum_{ij}(1-r_{ij})\ln\frac{1-r_{ij}}{1-s_{ij}} - \sum_{ij} r_{ij}\ln\frac{r_{ij}}{s_{ij}} \quad \forall 0 \leqslant r_{ij}, s_{ij} \leqslant 1 \tag{10}$$

From each classification level *b*, the corresponding entropy *h* is calculated. To decide the best classification an entropy-classification level diagram can be obtained. According to the Tondeur-Kvalen Equipartition Conjecture, the best classification is that in which the entropy production is most uniformly distributed, *i.e.*, a fitting of *h* vs. *b* intersects the *h*-*b* plot in the best *b* value (Tondeur and Kvaalen 1987):

$$h = h_{\max} \cdot b \tag{11}$$

We have written program MolClas for molecular classification. It performs cluster analyses (CAs) with IMSL subroutine CLINK (IMSL 1989) and punches the difference matrix

in format NEXUS (.nex), for standard programs (PAUP, MacClade and Splits-Tree).

Program GraphCor for the partial correlation diagram

The partial correlation diagram displays pairs of similar molecules with the high partial correlations ($|r| \geq 0.75$) in red color, pairs of molecules with the medium partial correlations ($0.50 \leq |r| < 0.75$) in orange and pairs of molecules with the low partial correlations ($0.25 \leq |r| < 0.50$) in yellow. Programs MolClas and GraphCor are available from the authors at Internet (torrens@uv.es) and are free for academics.

Statistical analysis

The molecular descriptors were taken from the ChEMBL (vs. 31) database (Mendez et al 2019). Energy was calculated with ChemBioOffice Ultra (vs. 21.0.0) and MarvinSketch (vs. 19.25). Energy is obtained by software calculation with MMFF94-Chem3D minimization. Merck molecular force field 1994 (MMFF94) provides good accuracy across a range of organic and *drug-like* molecules. The core parameterization was provided by high-quality quantum calculations, rather than experimental data, across 500 test molecular systems. The method includes parameters for a wide range of atom types (C, H, N, O, F, Si, P, S, Cl, Br and I), as well as ions (Fe^{2+}, Fe^{3+}, F^-, Cl^-, Br^-, Li^+, Na^+, K^+, Zn^{2+}, Ca^{2+}, Cu^+, Cu^{2+} and Mg^{2+}). MMFF94 performs well at optimizing geometries, bond lengths, angles, etc., as well as includes electrostatic and H-bonding effects.

The multiple linear regression models and principal components analysis (PCA) were carried out with Statistica (vs. 14.0.1), IBM SPSS (vs. 28.0.1.1), GNU PSPP (vs. 1.6.2), Knowledge Miner for Excel and Microsoft Office 365 (Excel) for Windows 11 Pro OS. The computed statistics are the number of points N, the Pearson correlation coefficient r, the coefficient of determination r^2, the adjusted r^2, the standard derivation s and the Fisher's ratio F. In addition we have used Gramatica's measures r^2, adjusted r^2, MAE, RMSE, Y-scrambling determination coefficient, q^2 and Y-scrambling cross-validated determination coefficient.

The cross-validated correlation coefficients r_{cv} and coefficients of determination r_{cv}^2 were worked out with the leave-many-out (LMO) program (Besalú 2001). The LMO selected the best set of descriptors according to the criterion of maximization of r_{cv}^2. The cross-validation was employed to determine the robustness of the models, which were compared and validated by taking into account r_{cv}^2 (q^2, etc.).

Conclusion

From the present results and discussion, the following conclusions can be drawn:

1. We obtained a periodic classification of stilbenes by using information entropy theory. The cytotoxicity for all the nine lines increases on going right through a period and augments when descending along a group. The stilbenes in the same group present similar properties. The most active Class 1 and Class 2 are characterized by the presence of two Methoxyl (OMe) groups in *meta* or *para* in ring B, as well as a Hydroxyl and an OMe in *ortho* in R_3/R_4 positions in ring A. The presence of Fluorine in R_6 sit in ring A augments the anticancer activities. The occurrence of Fluorine or more than two OMe groups in ring B decreases the activities.
2. The PC1-PC2 scores plot separates stilbenes in three groups. The grouping in clusters is more fundamental than that in classes. Cluster 1 includes Class 1 and corresponds to the most active stilbenes. Cluster 2 comprises Classes 2–4 and corresponds to high and intermediate activities, as well as Cluster 3 contains Classes 5–7 with the least activities. These results are shown to be consistent.
3. The PC1-PC2 loading plot shows the cytotoxicity for all the nine lines located together with positive loadings for both PC1 and PC2, suggesting that they would depend in part on the same variables. This is confirmed in the QSAR results.
4. Nine models are provided for fitting different cytotoxicity data *vs.* different cell lines. All the nine QSAR regressions are good in terms of correlation coefficient ($r^2 > 0.57$), standard derivation and Fisher's ratio. The most important descriptors are Density, $Methoxyl^2$, Energy, acd_logp^2 and HBA for positive correlation, as well as $Methoxyl^3$ and FP for negative association. The repetition of most independent variables for different cell lines agrees with the result of the loading plot, and we suggest the importance of including the selected set of seven variables in further studies of additional cell lines.
5. Leave-*m*-out cross-validated correlations for all the nine correlations are good ($q^2 > 0.33$), with good robustness for stilbenes not included in the fits. The different q^2 values indicate diverse robustness of models (1)–(9). For models (5), (6) and (8) q^2 is above 0.5, and the best cross-validated results were obtained.
6. We have proposed five new highly active compounds that are not described in the ChEMBL database and that have not been used in this study: Two of them would be found within Class 1, which is the most active used in the study, a third would appear in Class 2, and two

compounds would appear in Class 3. These structures are not available as natural compounds and are suggested here for future synthesis.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11696-026-05153-1.

Acknowledgements The authors acknowledge Dr. Besalú for providing them his full-linear leave-many-out program before publication. F.T. thanks project developed within the framework of the own program of the Vice-Rectorate for Research of the UV, Special Actions call, file UV-INV-AE-4235306. G.C. thanks funding from Universidad Católica de Valencia San Vicente Mártir.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Aja I, Ruiz-Larrea MB, Courtois A, Krisa S, Richard T, Ruiz-Sanz JI (2020) Screening of natural stilbene oligomers from *Vitis vinifera* for anticancer activity on human hepatocellular carcinoma cells. Antioxidants 9:469

Almosnid NM, Gao Y, He C, Park HS, Altman E (2016) *In vitro* antitumor effects of two novel oligostilbenes, *cis* and *trans*-suffruticosol D, isolated from *Paeonia suffruticosa* seeds. Int J Oncol 48:646–656

Basini G, Tringali C, Baioni L, Bussolati S, Spatafora C, Grasselli F (2010) Biological effects on granulosa cells of hydroxylated and methylated resveratrol analogues. Mol Nutr Food Res 54:S236–S243

Besalú E (2001) Fast computation of cross-validated properties in full linear leave-many-out procedures. J Math Chem 29(3):191–203

Bohara RA, Tabassum N, Singh MP, Gigli G, Ragusa A, Leporatti S (2022) Recent overview of resveratrol's beneficial effects and its nanodelivery systems. Molecules 27:5154

Castellano G, Lara A, Torrens F (2014) Classification of stilbenoid compounds by entropy of artificial intelligence. Phytochemistry 97:62–69

Castellano G, León A, Torrens F (2021) Classification of congeneric and QSAR of homologous antileukemic *S*–alkylcysteine ketones. Molecules 26:235–241

Castillo-Garit JA, del Toro-Cortés O, Kouznetsov VV, Puentes CO, Romero Bohórquez AR, Vega MC, Rolón M, Escario JA, Gómez-Barrio A, Marrero-Ponce Y, Torrens F, Abad C (2012) Identification *in silico* and *in vitro* of novel trypanosomicidal drug-like compounds. Chem Biol Drug des 80:38–45

Castillo-Garit JA, del Toro-Cortés O, Vega MC, Rolón M, Rojas de Arias A, Casañola-Martin GM, Escario JA, Gómez-Barrio A, Marrero-Ponce Y, Torrens F, Abad C (2015) Bond-based bilinear indices for computational discovery of novel trypanosomicidal drug-like compounds through virtual screening. Eur J Med Chem 96:238–244

Chalal M, Delmas D, Meunier P, Latruffe N, Vervandier-Fasseur D (2014) Inhibition of cancer derived cell lines proliferation by synthesized hydroxylated stilbenes and new ferrocenyl-stilbene analogs: comparison with resveratrol. Molecules 19:7850–7868

Chimento A, De Amicis F, Sirianni R, Sinicropi MS, Puoci F, Casaburi I, Saturnino C, Pezzi V (2019) Progress to improve oral bioavailability and beneficial effects of resveratrol. Int J Mol Sci 20:1381

Chrzascik I (2009) Analysis of biologically active stilbene derivatives. Crit Rev Anal Chem 39:70–80

Csuk R, Albert S, Siewert B, Schwarz S (2012) Synthesis and biological evaluation of novel (*E*)-stilbene-based antitumor agents. Eur J Med Chem 54:669–678

De Filippis B, Ammazzalorso A, Fantacuzzi M, Giampietro L, Maccallini C, Amoroso R (2027) Anticancer activity of stilbene-based derivatives. ChemMedChem 12:558–570

Dubrovina AS, Kiselev KV (2017) Regulation of stilbene biosynthesis in plants. Planta 246:597–623

D'yakonov VA, Dzhemileva LU, Dzhemilev UM (2017) Chapter 2: advances in the chemistry of natural and semisynthetic topoisomerase I/II inhibitors. Stud Nat Prod Chem 54:21–86

Frémont L (2000) Biological effects of resveratrol. Life Sci 66:663–673

Gramatica P (2013) On the development and validation of QSAR models. In: Reisfeld B, Mayeno A (eds) Computational toxicology, methods in molecular biology No. 930. Humana, Totowa, pp 499–526

Hammerbacher A, Ralph SG, Bohlmann J, Fenning TM, Gershenzon J, Schmidt A (2011) Biosynthesis of the major tetrahydroxystilbenes in spruce, astringin and isorhapontin, proceeds via resveratrol and is enhanced by fungal infection. Plant Physiol 157:876–890

Holwell SE, Cooper PA, Thompson MJ, Pettit GR, Lippert JW III, Martin SW, Bibby MC (2002) Anti-tumor and anti-vascular effects of the novel tubulin-binding agent combretastatin A-1 phosphate. Anticancer Res 22:3933–3940

Hwangbo K, Zheng MS, Kim YJ, Im JY, Lee CS, Woo MH, Jahng Y, Chang HW, Son JK (2012) Inhibition of DNA topoisomerases I and II of compounds from *Reynoutria japonica*. Arch Pharm Res 35:1583–1589

IMSL (1989) Integrated Mathematical Statistical Library (IMSL). IMSL, Houston

Iordache O (2011) Modeling multi-level systems. Springer, Berlin

Iordache O (2012) Self-evolvable systems: machine learning in social media. Springer, Berlin

Kimura Y, Sumiyoshi M, Baba K (2008) Antitumor activities of synthetic and natural stilbenes through antiangiogenic action. Cancer Sci 99:2083–2096

Kullback S (1959) Information theory and statistics. Wiley, New York

Lion CJ, Matthews CS, Stevens MFG, Westwell AD (2005) Synthesis, antitumor evaluation, and apoptosis-inducing activity of hydroxylated (*E*)-stilbenes. J Med Chem 48:1292–1295

Liu JC, Chen B, Yang JL, Weng JQ, Yu Q, Hu DX (2022) Design, synthesis and cytotoxicity of thiazole-based stilbene analogs as novel DNA topoisomerase IB inhibitors. Molecules 27:1009–1011

Ma Z, Zhang X, Xu L, Liu D, Di S, Li W, Jiao Z, Hongmei Z, Xiaofei L, Jing H (2019) Pterostilbene: mechanisms of its action as oncostatic agent in cell models and *in vivo* studies. Pharmacol Res 145:104265-1–104312

Martí-Centelles R, Falomir E, Murga J, Carda M, Marco JA (2015) Inhibitory effect of cytotoxic stilbenes related to resveratrol on the expression of the VEGF, hTERT and c-Myc genes. Eur J Med Chem 103:488–496

Mazue F, Colin D, Gobbo J, Wegner M, Rescifina A, Spatafora V, Fasseur D, Delmas D, Meunier P, Tringali C, Latruffe N (2010) Structural determinants of resveratrol for cell proliferation inhibition potency: experimental and docking studies of new analogs. Eur J Med Chem 45:2972–2980

Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Felix E, Magariños MP, Mosquera JF, Mutowo P, Nowotka M, Gordillo-Marañon M, Hunter F, Junco L, Mugumbate G, Rodriguez-Lopez M, Atkinson F, Bosc N, Radoux CJ, Segura-Cabrera A, Hersey A, Leach AR (2019) ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res 47:D930–D940

Mikstacka R, Zielińska-Przyjemska M, Dutkiewicz Z, Cichocki M, Stefański T, Kaczmarek M, Baer-Dubowska W (2018) Cytotoxic, tubulin-interfering and proapoptotic activities of 4'-methylthio-*trans*-stilbene derivatives, analogues of *trans*-resveratrol. Cytotechnology 70:1349–1362

Morris VL, Toseef T, Nazumudeen FB, Rivoira C, Spatafora C, Tringali C, Rotenberg SA (2015) Anti-tumor properties of *cis*-resveratrol methylated analogs in metastatic mouse melanoma cells. Mol Cell Biochem 402:83–91

Mustafa M, Anwar S, Elgamal F, Ahmed ER, Aly OM (2019) Potent combretastatin A-4 analogs containing 1,2,4-triazole: synthesis, antiproliferative, anti-tubulin activity, and docking study. Eur J Med Chem 183:111697-1–111712

Pannu N, Bhatnagar A (2019) Resveratrol: from enhanced biosynthesis and bioavailability to multitargeting chronic diseases. Biomed Pharmacother 109:2237–2251

Poddar A, Pal R, Rong C, Chattaraj PK (2023) A conceptual DFT and information-theoretic approach towards QSAR modeling in polychlorobiphenyls. J Math Chem 61:1143–1164

Roberti M, Pizzirani D, Simoni D, Rondanin R, Baruchello R, Bonora C, Buscemi F, Grimaudo S, Tolomeo M (2003) Synthesis and biological evaluation of resveratrol and analogues as apoptosis-inducing agents. J Med Chem 46:3546–3554

Roslie H, Chan KM, Rajab NF, Velu SS, Kadir SAIASA, Bunyamin I, Weber JFF, Thomas NF, Majeed ABA, Myatt G, Inayat-Hussain SH (2012) 3,5-Dibenzyloxy-4'-hydroxystilbene induces early caspase-9 activation during apoptosis in human K562 chronic myelogenous leukemia cells. J Toxicol Sci 37:13–21

Shnyder SD, Cooper PA, Pettit GR, Lippert JW III, Bibby MC (2003) Combretastatin A-1 phosphate potentiates the antitumour activity of cisplatin in a murine adenocarcinoma model. Anticancer Res 23:1619–1623

Sirerol JA, Rodríguez ML, Mena S, Asensi MA, Estrela JM, Ortega AL (2016) Role of natural stilbenes in the prevention of cancer. Oxid Med Cell Longev 2016:3128951

Tondeur D, Kvaalen E (1987) Equipartition of entropy production: an optimality criterion for transfer and separation processes. Int J Cheminf Chem Eng 26:50–56

Tron GC, Pirali T, Sorba G, Pagliai F, Busacca S, Genazzani AA (2006) Medicinal chemistry of combretastatin A4: present and future directions. J Med Chem 49:3033–3044

Valdameri G, Pereira Rangel L, Spatafora C, Guitton JE, Gauthier V, Arnaud O, Ferreira-Pereira A, Falson P, Winnischofer SMB, Rocha MEM, Tringali CM, Di Pietro A (2012) Methoxy stilbenes as potent, specific, untransported, and noncytotoxic inhibitors of breast cancer resistance protein. ACS Chem Biol 7:322–330

White H (1989) Neural network learning and statistics. AI Expert 4:48–52

Yamada M, Hayashi K, Ikeda S, Tsutsui K, Tsutsui K, Ito T, Iinuma M, Nozaki H (2006) Inhibitory activity of plant stilbene oligomers against DNA topoisomerase II. Biol Pharm Bull 29:1504–1507

Yang SC, Tseng CH, Wang PW, Lu PL, Weng YH, Yen FL, Fang JY (2017) Pterostilbene, a methoxylated resveratrol derivative, efficiently eradicates planktonic, biofilm, and intracellular MRSA by topical application. Front Microbiol 8:1103–1111

Zhang Y, Shen M, Cui S, Hou T (2014) Synthesis and antiproliferative evaluation of 2-hydroxylated (*E*)-stilbenes. Bioorg Med Chem Lett 23:5470–5472

Zielińska-Przyjemska M, Kaczmarek M, Krajka-Kuźniak V, Wierzchowski M, Baer-Dubowska W (2020) Effect of methoxy stilbenes—analogs of resveratrol—on the viability and induction of cell cycle arrest and apoptosis in human myeloid leukemia cells. Mol Cell Biochem 474:113–123