

3D Connectivity Indices in QSPR/QSAR Studies

Ernesto Estrada* and Enrique Molina†

Faculty of Pharmacy, Department of Organic Chemistry, University of Santiago de Compostela,
Santiago de Compostela 15706, Spain

Received November 29, 2000

Topographic (3D) molecular connectivity indices based on molecular graphs weighted with quantum chemical parameters are used in QSPR and QSAR studies. These descriptors were compared to 2D connectivity indices (vertex and edge ones) and to quantum chemical descriptors in modeling partition coefficient ($\log P$) and antibacterial activity of 2-furylethylene derivatives. In describing $\log P$ the 3D connectivity indices produced a significant improvement (more than 29%) in the predictive capacity of the model compared to those derived with topological and quantum chemical descriptors. The best linear discriminant model for classifying antibacterial activity of these compounds was also obtained with the use of 3D connectivity indices. The global percent of good classification obtained with 3D and 2D connectivity as well as quantum chemical descriptors were 94.1, 91.2, and 88.2, respectively. In general, all these models predict correctly the antibacterial activity of a set of nine new 2-furylethylene derivatives. The best result is obtained with 3D connectivity indices that classified correctly 100% of these compounds versus 88.9% obtained with 2D connectivity or quantum chemical descriptors.

INTRODUCTION

Molecular connectivity indices were introduced 25 years ago by Randić, Kier, and Hall.^{1,2} These molecular descriptors have been the most successful topological indices in quantitative structure–property (QSPR) and quantitative structure–activity relationship (QSAR) studies.^{2–5} Recent success in rational combinatorial library design,^{6–8} virtual screening, and drug design^{9–11} have shown that these descriptors play an important role in the discovery of novel lead compounds for pharmacological uses.

The success of these indices in describing physicochemical, pharmacological, and toxicological properties of organic compounds proves that the graph-theoretical concepts and invariant in which they are based encoded an important amount of structural information.^{12–17} This information is encoded in the form of 2D structural features based on the connectivity among atoms in the molecules. It has been believed, and in some cases proved, that 2D descriptors contain important 3D structural information.^{18,19} However, as Bath et al. stated,²⁰ this information is not sufficient, and 3D descriptors are necessary for describing 3D features of molecules.

An important step forward in the development of graph-based molecular descriptors has been the definition of topographic descriptors. These kinds of molecular descriptors are based on molecular graphs with appropriate weights to account for 3D molecular features. The pioneering works in this direction were done by Milan Randić at the end of 1980s.^{21–23} In these works Randić proposed the use of topographic distance matrices, first based on graph embedded into a hexagonal lattice and then into a 3D diamond lattice. A further work of Bogdanov et al.²⁴ proposed a 3D distance

matrix to compute a 3D Wiener index. The other direction in the search of topographic indices was followed by Estrada who proposed the use of quantum chemical parameters as vertex and edge weights for building topographic adjacency matrices.^{25–27} This author proposed in 1993 the definition of topographic connectivity indices based on bond order weighted molecular graphs.²⁵ Further, this type of weighted graphs was used to define a topographic edge adjacency matrix (3D bond matrix) as well as a 3D bond connectivity index.²⁶ In a parallel approach Estrada introduced molecular graphs with vertices weighted by electron charge density on atoms to define a 3D adjacency matrix and a set of topographic connectivity descriptors.²⁷ In more recent years different authors have defined other types of topographic descriptors.^{28–31}

Up to now there are not extensive applications of topographic descriptors in QSPR/QSAR studies. Therefore, we propose here to start the study on the use of 3D (atom and bond) connectivity indices to describe quantitatively physicochemical and biological properties of organic compounds. This set of descriptors can be obtained from quantum chemical calculations in a straightforward way. So, they can be used as molecular descriptors in such studies that use quantum chemical parameters to describe molecular properties. We will compare the results obtained by using 3D connectivity indices to that obtained with 2D connectivity indices and with quantum chemical parameters.

THEORETICAL APPROACH

Three different theoretical approaches will be used here to describe the partition coefficient of a set of compounds. In general, these approaches pretend to find a linear model of the form

$$\text{property} = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n$$

in which x_i are topological (2D) or topographic (3D)

* Corresponding author fax: 34-981-594912; e-mail: estrada66@yahoo.com.

† On leave from the Department of Chemistry, Faculty of Chemical Engineering and Pharmacy, University of Camagüey, 74650 Camagüey, Cuba.

molecular connectivity or quantum chemical molecular descriptors. The first set of descriptors to be used here is formed by the 2D connectivity indices, which are defined as follows¹⁻³

$${}^h\chi_t = \sum_{s=1}^h (\delta_i \delta_j \dots \delta_{h+1})_s^{-0.5} \quad (1)$$

where ${}^h\chi_t$ is the connectivity index of type t and order h . The product is over the $h + 1$ vertex degrees in the subgraph having h edges, and the summation is carried out over all subgraphs of type t in the molecule. The different types of subgraphs studied in the molecular connectivity scheme are as follows: path, clusters, path-clusters, and chains, which are designed as p , C , pC , and Ch , respectively, according to their original definitions.

The valence connectivity indices are calculated in a similar way but using valence degrees instead of simple vertex degrees:^{2,3}

$${}^h\chi_t^v = \sum_{s=1}^h (\delta_i^v \delta_j^v \dots \delta_{h+1}^v)_s^{-0.5} \quad (2)$$

with the valence degree of an atom defined as follows

$$\delta_i^v = Z_i^v - h_i \quad \text{for second row elements} \quad (3)$$

$$\delta_i^v = (Z_i^v - h_i) / (Z_i - Z_i^v - 1) \quad \text{for the rest of the atoms} \quad (4)$$

where Z_i^v is the number of valence electrons, h_i is the number of hydrogen atoms bonded to the atom, and Z_i is the atomic number.

Bond connectivity indices have been defined as follows:³²⁻³⁵

$${}^h\epsilon_t = \sum_{s=1}^h (\delta(e_i) \delta(e_j) \dots \delta(e_h))_s^{-0.5} \quad (5)$$

where $\delta(e_i)$ is the bond degree, and the sum is over all edges of the subgraph t .

The topographic connectivity indices that will be used in the current work have been introduced by using the same graph invariant that for the topological connectivity indices but using quantum chemical properties as vertex and edge weights. These topographic descriptors can be classified in two different groups. The first is that of descriptors based on bond order-weighted molecular graphs. The other group is formed by descriptors based on electron charge density-weighted molecular graphs.

Bond order weighted vertex connectivity indices ${}^h\Omega_t$ are calculated from expression (1) but using bond order weighted vertex degrees instead of simple vertex degrees.²⁵ The bond order weighted vertex degrees are defined as follows²⁵

$$\delta_i(\rho) = \sum_j \rho_{ij} \quad (6)$$

where ρ_{ij} is the bond order or valence index of bond i - j , and the sum is over all non-hydrogen atoms adjacent to i .

The bond connectivity indices based on bond order weighted molecular graphs ${}^h\epsilon_t(\rho)$ are calculated from expres-

sion (5) but using the following definition of bond degrees²⁷

$$\delta(e_k) = \delta_i(\rho) + \delta_j(\rho) - 2\rho_{ij} \quad (7)$$

where the bond e_k is incident to vertices i and j .

Connectivity indices based on electron charge density ${}^h\Omega_t^c(q)$ are calculated from expression (1) but using the following vertex degrees²⁶

$$\delta_i(q) = q_i - h_i \quad (8)$$

where q_i is the electron charge density on atom i . A correction for hydrogen atoms is introduced in this scheme to calculate the indices ${}^h\Omega_t^c(q)$ according to the following approach

$$\delta_i^c(q) = q_i - \sum_j q_{hj} \quad (9)$$

where q_{hj} is the electron charge density of the j th hydrogen atom bonded to the atom i .

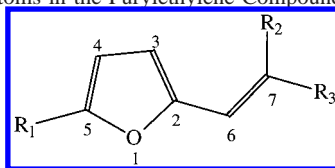
A series of quantum chemical molecular descriptors was also computed by following the procedure explained in the following section. This set of descriptors was calculated only for the atoms of the furylethylene framework, which is the structural pattern repeated in all compounds of this data set. It includes the electronic charge on atoms: the energy of the highest occupied molecular orbital (HOMO) and of the lowest unoccupied molecular orbital (LUMO) as well as superdelocalizability indices, such as electrophilic and nucleophilic superdelocalizabilities of σ and π electrons in the molecule. For the definition of these terms see ref 36. A resume of the symbols and names of the topological, topographic, and quantum chemical descriptors to be used in the current work is given in Table 1.

Computations and Data Set. Vertex and edge molecular connectivity indices of different types and orders were calculated for each compound in the data set by using the computer system MODEST for Windows.³⁷ 3D (topographic) connectivity indices were also calculated by using the same computer software. In computing these descriptors the computational strategy consisted of carrying out a full geometry optimization of the molecular structures of the furylethylene compounds by using the semiempirical quantum chemical method AM1.³⁸ This method was selected because we previously demonstrated that it reproduces the molecular geometry and conformational features of these compounds in an appropriate way.³⁹ After these calculations the output files were used as input for the system MODEST³⁷ in order to compute the topological, topographic, and quantum chemical descriptors (the software can be obtained free upon request to E. Estrada). In carrying out full geometry optimization with MOPAC 6.0⁴⁰ the keyword PRECISE was always used in the calculations in order to obtain better precision. Other two keywords used were VECTORS to generate the eigenvectors of the wave function and the word BONDS to obtain the bond orders and valencies of the correspondent bonds and atoms in the molecules. A total number of 41 topological, 55 topographic connectivity indices, and 51 quantum chemical descriptors were calculated according to the definitions given before and in Table 1.

A set of 34 2-furylethylene derivatives with different substituents at position 5 of the furan ring as well as at the β position of the exocyclic double bond were selected for

Table 1. Symbols for Topological, Topographic, and Quantum Chemical Descriptors and Their Definitions

${}^h\chi_p$	path connectivity index of order $h = 0-6$
${}^h\chi_c$	cluster connectivity index of order $h = 3-6$
${}^h\chi_{pC}$	path-cluster connectivity index of order $h = 4-6$
${}^h\chi_p^v$	valence path connectivity index of order $h = 0-6$
${}^h\chi_c^v$	valence cluster connectivity index of order $h = 3-6$
${}^h\chi_{pC}^v$	valence path-cluster connectivity index of order $h = 4-6$
${}^h\epsilon_p$	path bond connectivity index of order $h = 1-6$
${}^h\epsilon_c$	cluster bond connectivity index of order $h = 3-6$
${}^h\epsilon_{pC}$	path-cluster bond connectivity index of order $h = 4-6$
${}^h\Omega_p$	path bond-order-based topographic connectivity index of order $h = 0-6$
${}^h\Omega_c$	cluster bond-order-based topographic connectivity index of order $h = 3-6$
${}^h\Omega_{pC}$	path-cluster bond-order-based topographic connectivity index of order $h = 4-6$
${}^h\Omega_p(q)$	path charge-based topographic connectivity index of order $h = 0-6$
${}^h\Omega_c(q)$	cluster charge-based topographic connectivity index of order $h = 3-6$
${}^h\Omega_{pC}(q)$	path-cluster charge-based topographic connectivity index of order $h = 4-6$
${}^h\Omega_p^c(q)$	path hydrogen-corrected charge-based topographic connectivity index of order $h = 0-6$
${}^h\Omega_c^c(q)$	cluster hydrogen-corrected charge-based topographic connectivity index of order $h = 3-6$
${}^h\Omega_{pC}^c(q)$	path-cluster hydrogen-corrected charge-based topographic connectivity index of order $h = 4-6$
${}^h\epsilon_p(\rho)$	path bond-order-based topographic bond connectivity index of order $h = 1-6$
${}^h\epsilon_c(\rho)$	cluster bond-order-based topographic bond connectivity index of order $h = 3-6$
${}^h\epsilon_{pC}(\rho)$	path-cluster bond-order-based topographic bond connectivity index of order $h = 4-6$
$Q(i)$	electronic charge on atom i
$ES_\sigma(i)$	σ electrophilic superdeslocalizability on atom i
$ES_\pi(i)$	π electrophilic superdeslocalizability on atom i
$ES_T(i)$	total electrophilic superdeslocalizability on atom i
$NS_\sigma(i)$	σ nucleophilic superdeslocalizability on atom i
$NS_\pi(i)$	π nucleophilic superdeslocalizability on atom i
$NS_T(i)$	total nucleophilic superdeslocalizability on atom i
E_{HOMO}	energy of the highest occupied molecular orbital
E_{LUMO}	energy of the lowest unoccupied molecular orbital

Table 2. Chemical Structures and Enumeration of Atoms in the Furylethylene Compounds Used in This Study

no.	R ₁	R ₂	R ₃	no.	R ₁	R ₂	R ₃
1	H	NO ₂	COOCH ₃	18	NO ₂	H	CONHCH(CH ₃)C ₂ H ₅
2	CH ₃	NO ₂	COOCH ₃	19	NO ₂	H	CONHC(CH ₃) ₃
3	Br	NO ₂	COOCH ₃	20	NO ₂	H	CONHCH ₂ C(CH ₃) ₃
4	I	NO ₂	COOCH ₃	21	NO ₂	H	COOCH ₃
5	COOCH ₃	NO ₂	COOCH ₃	22	NO ₂	H	COOC ₂ H ₅
6	NO ₂	NO ₂	COOCH ₃	23	NO ₂	H	COO(CH ₂) ₂ CH ₃
7	NO ₂	COOC ₂ H ₅	COOC ₂ H ₅	24	NO ₂	H	COOCH(CH ₃) ₂
8	NO ₂	H	NO ₂	25	NO ₂	H	COO(CH ₂) ₃ CH ₃
9	H	H	NO ₂	26	NO ₂	H	COOCH ₂ CH(CH ₃) ₂
10	NO ₂	H	CONH ₂	27	NO ₂	H	COOCH(CH ₃)C ₂ H ₅
11	NO ₂	H	CONHCH ₃	28	NO ₂	H	COOC(CH ₃) ₃
12	NO ₂	H	CON(CH ₃) ₂	29	NO ₂	H	COO(CH ₂) ₄ CH ₃
13	NO ₂	H	CONHC ₂ H ₅	30	NO ₂	H	Br
14	NO ₂	H	CONH(CH ₂) ₂ CH ₃	31	NO ₂	H	CN
15	NO ₂	H	CONHCH(CH ₃) ₂	32	NO ₂	H	OCH ₃
16	NO ₂	H	CONH(CH ₂) ₃ CH ₃	33	NO ₂	H	H
17	NO ₂	H	CONHCH ₂ CH(CH ₃) ₂	34	NO ₂	CN	COOCH ₃

this study.⁴¹ The structure of these 34 furylethylene derivatives are given in Table 2. The 2-furylethylene compounds have been well-known as antimicrobials, antitumoral, and cytotoxic during many years.^{42,43} The recent interest in such chemicals emerged from the strong antibacterial/antifungal activity of some derivatives with applications in the industry as well as in veterinary and human medicine.⁴⁴⁻⁴⁶ It has been clear from structure-activity relationship studies that the lipophilicity and the nucleophilic addition of thiols groups of some enzymes to the exocyclic double bond of such compounds are critical for the development of their anti-

bacterial activity.⁴¹ The role of the different substituents in the mutagenic activity observed in some derivatives of 2-furylethylenes has also been studied.⁴⁷

The values of the partition coefficient *n*-octanol/water (log *P*) of these compounds have been experimentally determined and reported in the literature.⁴¹ On the other hand the antibacterial activity of these compounds was determined as the inverse of the concentration *C* that produces 50% of growth inhibition in *E. coli* at six different times and reported as log(1/*C*).⁴¹ We take the values at time 1 min of exposition of the chemical to bacteria as the measurement of activity

to be studied here. This antibacterial activity was used to classify furylethylenes in two groups. The group I of active compounds is composed of those compounds having values of $\log(1/C) \leq 3$, while the rest formed group II, of inactive, compounds.

PARTITION COEFFICIENTS

The partition coefficient *n*-octanol/water ($\log P$) has an important role in the understanding of the biological behavior of these 2-furylethylene derivatives.⁴¹ So, we will study this parameter to compare the possibilities of topographic descriptors in QSPR and will compare these results to those obtained by using topological and quantum chemical descriptors. Our objective here is not to find the best QSPR model for describing $\log P$ of this data set of compounds but to compare the possibilities of these three sets of descriptors in such QSPR study. For this reason, we will obtain models that include only one of such kind of descriptors, for instance one model with topological indices, another with topographic, and another with quantum chemical ones.

The best QSPR model by using the topological indices studied (vertex and edge connectivity indices) is given below together with the statistical parameters of the regression:

$$\log P = 0.223 - 1.087(^4\epsilon_p) + 0.763(^6\chi_{pC}) + 1.735(^1\chi_p^v) + 2.538(^5\chi_p^v) - 11.626(^6\chi_p^v) - 3.250(^5\chi_{pC}^v) - 0.447(HB1) \quad (10)$$

$$N = 34, R = 0.969, s = 0.199, RMSECV = 0.247, F(7,26) = 56.9$$

where *N* is the number of compounds, *R* is the correlation coefficient, *s* is the standard deviation of the regression, RMSECV is the root of the mean square error of the leave-one-out cross validation, and *F*(*n*₁,*n*₂) is the Fisher ratio with *n*₁ being the number of variables and *n*₂ = *N* - *n*₁ - 1. The term HB1 is a hydrogen bond number indicator that takes integer values depending on the number of hydrogen bonds, donor or acceptor, that an atom or group can form.⁴⁸ As can be seen this model explains almost 94% of the variance in $\log P$ for this data set of compounds. It includes one connectivity index, four valence connectivity indices and one bond connectivity index.

The use of topographic descriptors produces a model able to explain more than 96% of the variance in $\log P$. It is also clear from the leave-one-out cross validation experiment that the predictive ability of the topographic model represents an improvement of 29% respect to that obtained with topological descriptors. The model obtained with this kind of molecular descriptors is given below:

$$\log P = 0.125 + 0.708[^1\Omega_p(q)] - 0.549[^4\epsilon_p(\rho)] - 33.331[^5\epsilon_C(\rho)] + 44.730[^5\Omega_C^c(q)] + 0.730[^2\epsilon_p(\rho)] - 0.825[^3\Omega_C] - 0.590(HB1) \quad (11)$$

$$N = 34, R = 0.982, s = 0.155, RMSECV = 0.176, F(7,26) = 84.6$$

In contrast with model 10 this model includes four 3D edge connectivity indices, two 3D connectivity indices based on

electron charge density, and one connectivity index based on bond orders weighted molecular graphs.

When quantum chemical molecular descriptors were applied to describe this property, no one significant model from the statistical point of view could be obtained. By this reason, we used the Rogers and Cammarata approach⁴⁹ for describing the partition coefficient by using quantum chemical indices. This approach is based on the use of superdelocalizability and charge density of all atoms in the molecule. According to Rogers and Cammarata the partition coefficient *n*-octanol/water of a set of molecules can be described by the following quantitative expression⁴⁹

$$\log P = a_0 + b \sum_i ES_i + c \sum_i Q_i \quad (12)$$

where the sum of electrophilic superdelocalizabilities and the absolute values of atom charge densities is carried out over all atoms of the molecule. We also included here the term HB1 in order to improve the quality of the model so developed. The quantitative model so derived is given below:

$$\log P = 0.192 + 2.507 \left(\sum_i ES_i \right) + 0.393 \left(\sum_i Q_i \right) - 0.752(HB1) \quad (13)$$

$$N = 34, R = 0.906, s = 0.319, RMSECV = 0.37, F(3,30) = 45.5$$

This model is able to explain only 82% of the variance in the $\log P$ for the data set of furylethylenes studied, which can be considered as a poor result compared to those reported with the use of topological and topographic connectivity indices. However, this model contains only three variables contrary to the greater number of the previous ones. In fact, the quantum chemical parameters used in this model are good candidates to be combined with topological and topographic descriptors to derive a very relevant QSPR model for predicting $\log P$ of organic compounds.

In closing, we have proved that the inclusion of some electronic features (bond orders and electron charge densities) in the framework of the connectivity indices (vertex and edge) produces the best results for describing the partition coefficients *n*-octanol/water for the 34 furylethylenes studied. It has also been clear that the partition coefficient of this data set of 2-furylethylenes is well described by connectivity molecular descriptors in general and that the use of quantum chemical indices is not recommended in this case. The experimental values of $\log P$ as well as those calculated from the different models obtained here are given in Table 3. The linear correlation models are illustrated in Figure 1.

ANTIBACTERIAL ACTIVITY

Linear discriminant analysis will be used here to obtain three classification models of 2-furylethylene compounds according to their antibacterial activity. As we explained previously two groups were created taking as cutoff value for activity $\log(1/C) \leq 3$. Despite some compounds having $\log(1/C) > 3$ show some activity at longer times of exposition they will be considered as less active and will be maintained in group II because the main objective in this case is to develop compounds with great activity at short times of exposition. This criteria of selection for compounds with

Table 3. Experimental and Calculated Values of the Partition Coefficient *n*-Octanol/Water (log *P*) for the Furylethylenes Studied

no.	exp.	topol.	topog.	QC
1	1.879	1.894	1.955	1.836
2	2.439	2.482	2.398	2.239
3	2.739	2.753	2.748	2.405
4	2.999	2.905	2.898	2.510
5	1.869	1.763	1.930	1.976
6	1.599	1.619	1.550	1.679
7	2.504	2.703	2.640	2.706
8	1.303	1.191	1.338	1.456
9	1.583	1.453	1.783	1.583
10	0.649	0.433	0.300	0.180
11	0.984	0.999	1.091	1.076
12	0.819	1.160	0.870	2.149
13	1.386	1.583	1.423	1.482
14	1.860	2.311	1.941	1.858
15	1.803	1.966	2.084	1.906
16	2.356	2.168	2.332	2.240
17	2.225	2.493	2.526	2.241
18	2.284	2.384	2.383	2.277
19	2.333	2.316	2.316	2.346
20	2.605	2.382	2.575	2.618
21	1.652	1.347	1.585	1.830
22	2.098	1.984	1.947	2.126
23	2.673	2.733	2.459	2.504
24	2.641	2.484	2.666	2.592
25	2.827	2.726	2.837	2.902
26	3.135	3.052	3.034	2.902
27	3.091	3.018	2.952	2.943
28	3.060	2.994	3.002	3.029
29	3.404	3.227	3.252	3.266
30	2.447	2.510	2.469	2.132
31	1.050	1.365	1.258	1.344
32	1.591	1.510	1.500	1.711
33	1.611	1.738	1.515	1.590
34	1.488	1.309	1.424	1.504

great activity at short times of exposition will probably reduce or avoid some of the undesired effects of such compounds. In developing classification models the values of 1 and -1 were assigned to active and inactive compounds, respectively. To make the classification of compounds in both groups we preferred the use of the a posteriori probabilities instead of cutoff values. This is the probability that the respective case belongs to a particular group (active or inactive) and it is proportional to the Mahalanobis distance from that group centroid. When the probability of being active does not differ more than 5% from the probability of being inactive, the case is considered as not classified (NC).

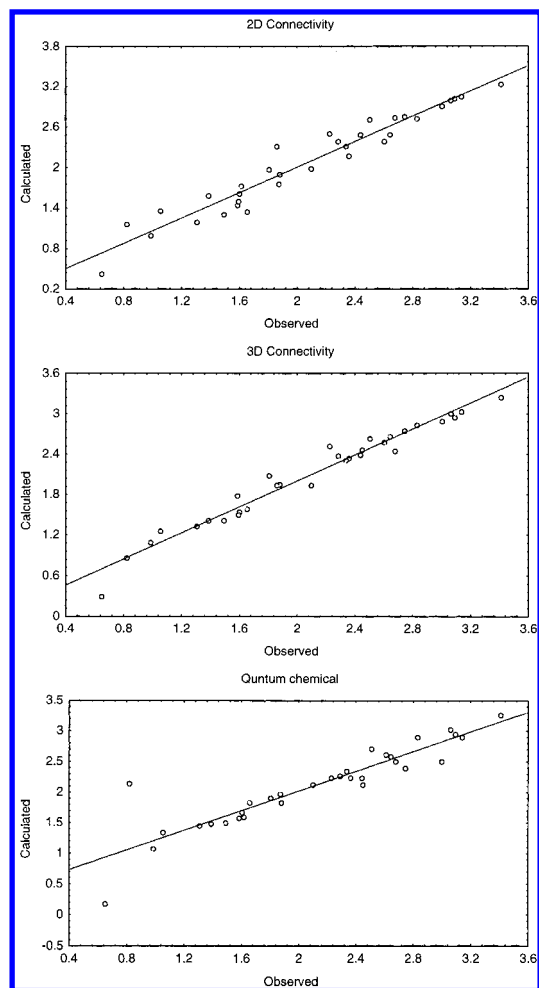
The first model derived here to classify 2-furylethylenes according to their antibacterial activity is obtained with topological connectivity indices:

$$\text{Act} = -5.417(\epsilon_p^5) + 3.073(\epsilon_{pC}^6) - 3.923(\chi_p^v) + 9.915(\chi_{pC}^v) + 1.947(\text{HB1}) + 2.458 \quad (14)$$

$$N = 34, \lambda = 0.43, F(5,28) = 7.7, D^2 = 5.7$$

The Wilks' λ statistic for the overall discrimination can takes values in the range of 0 (perfect discrimination) to 1 (no discrimination). The Mahalanobis square distance D^2 indicates the separation of the respective groups, showing whether the model possesses an appropriate discriminatory power for differentiating between the two respective groups.

This model classified correctly 92.9% of group I compounds and 90.0% of group II, giving a global percentage

**Figure 1.** Linear correlations of observed versus calculated log *P* according to the models obtained from 2D connectivity, 3D connectivity, and quantum chemical descriptors.

of good classification of 91.2%. The number of false active compounds, those inactive classified as active, represents almost 6% (2/34), while the number of false inactive, those active classified as inactive, is less than 3% (1/34).

The use of 3D connectivity indices to find a linear discriminant model for antibacterial 2-furylethylenes produces the following result:

$$\begin{aligned} \text{Act} = & -9.846[\epsilon_{pC}^4] + 1.559[\Omega_p^c(q)] - \\ & 29.219[\Omega_p^c(q)] + 119.998[\Omega_C^c(q)] + 2.335(\text{HB1}) - \\ & 4.963 \quad (15) \end{aligned}$$

$$N = 34, \lambda = 0.38, F(5,28) = 9.1, D^2 = 6.7$$

The improvement in the statistical parameters of this model compared to that using 2D connectivity indices is easily detected by the decrease in the Wilks' λ parameter and an increase in the Mahalanobis square distance. In consequence, model 15 classified correctly 92.9% of active compounds and 95.0% of inactive ones for a global good classification of 94.1%. The number of false inactives is the same as that from the previous model, but the number of false actives is improved to less than 3% instead of almost 6% from model 14.

The third model obtained here is with the use of quantum chemical parameters. The best classification model obtained

Table 4. Classification of 2-Furylethylene Derivatives as Antibacterial According to the Three Models Obtained with 2D and 3D Connectivity and Quantum Chemical Descriptors

compd	obsd.	2D conn.		3D conn.		quantum	
		class.	prob.	class.	prob.	class.	prob.
1	+	+	95.43	+	99.49	+	99.72
2	+	+	91.67	+	95.83	+	99.86
3	+	+	84.95	+	96.22	+	98.31
4	+	+	79.65	+	95.78	+	97.67
5	+	+	99.72	+	99.63	+	99.66
6	+	+	99.85	+	99.98	+	98.91
7	+	+	94.29	+	91.77	+	98.57
8	+	+	74.81	+	57.81	+	92.50
9	+	-	9.86	-	8.11	+	77.08
10	+	+	99.13	+	99.28	-	32.29
11	+	+	88.24	+	57.28	-	9.46
12	+	+	66.00	+	86.94	-	4.26
13	+	+	57.89	+	71.56	-	12.79
14	-	-	6.25	-	46.43	-	12.96
15	-	-	28.14	-	36.15	-	8.72
16	-	-	0.92	-	1.1	-	11.78
17	-	-	2.35	-	6.19	-	11.05
18	-	-	37.62	-	4.56	-	9.96
19	-	-	8.96	-	2.97	-	9.96
20	-	-	1.14	-	0.08	-	9.75
21	-	+	55.73	+	88.95	-	8.59
22	-	-	22.77	-	18.96	-	7.6
23	-	-	1.36	-	8.23	-	7.66
24	-	-	7.01	-	4.83	-	6.44
25	-	-	0.19	-	0.09	-	8.56
26	-	-	0.33	-	0.58	-	8.49
27	-	-	7.59	-	0.43	-	7.94
28	-	-	1.41	-	0.27	-	7.02
29	-	-	0.02	-	0.04	-	7.21
30	-	-	4.65	-	7.56	-	0.32
31	-	-	29.58	-	37.61	-	3.04
32	-	-	23.67	-	14.96	-	5.3
33	-	+	58.87	-	14.08	-	0.42
34	+	+	97.13	+	97.50	+	62.36

with these descriptors is given below:

$$\text{Act} = 90.28[ES_o(7)] - 2.08[NS_{\pi}(3)] - 24.92[Q(1)] + 91.33[Q(3)] + 1.44(HB1) + 14.65 \quad (16)$$

$$N = 34, \lambda = 0.44, F(5,28) = 7.1, D^2 = 5.2$$

This model has similar statistical parameters than that obtained with 2D connectivity indices. However, it shows the worst global classification results of the three models with a percentage of good classification of 88.2%. However, it is curious that this model classified correctly 100% of the compounds in group II, which gives not one false active compound. The main problem of this model is in the classification of compounds of group I where it fails in four cases, which gives 11.8% of false inactives. In Table 4 we give the classification of all 2-furylethylenes according to each model.

These models can be combined in different ways in order to predict the antibacterial activity of novel 2-furylethylene, or we can use a combination of topological, topographic, and quantum chemical descriptors to obtain one single model. However, here we will use these three models separately to predict the antibacterial activity of a series of nine new 2-furylethylenes synthesized in one of our laboratories. These compounds were synthesized by maintaining a NO₂ group at position R₃ and using Br or I at positions R₁ and/or R₂. All these compounds were shown to have antibacterial

Table 5. Predicted Antibacterial Activity of Novel R₁,R₂-Substituted 2-Furylethylenes (R₃ = NO₂) by Using Classification Models Obtained from 2D Connectivity, 3D Connectivity, and Quantum Chemical Descriptors^a

compd	R ₁	R ₂	2D conn.		3D conn.		quantum	
			class.	prob.	class.	prob.	class.	prob.
1	Br	Br	+	88.53	+	95.81	+	87.18
2	I	I	+	86.87	+	94.59	+	85.53
3	Br	H	+	59.01	+	65.00	+	54.95
4	H	Br	+	96.35	+	99.59	+	96.12
5	I	H	NC	50.07	+	53.00	NC	50.01
6	H	I	+	96.72	+	99.51	+	97.68
7	H	CH ₃	+	95.92	+	99.62	+	94.76
8	Br	CH ₃	+	86.38	+	96.06	+	84.27
9	I	CH ₃	+	81.75	+	95.74	+	79.29

^a NC stand for not classified compounds.

activity in different assays.^{44-46,50} The structures of these compounds as well as their classifications with the three models found here are given in Table 5.

As can be seen only one compound (number 5) is not classified as active by models 14 and 16. In this case both models give an a posteriori probability of active that does not differ from that of inactive by more than 5%. Consequently, we consider this compound as not-classified by these two models. However, model 15 obtained by using 3D connectivity indices classified correctly all nine compounds including compound 5.

CONCLUSIONS

We have shown here that the 3D connectivity indices based on molecular graphs with edges or vertices weighted with electronic parameters encode important structural information that is of great usefulness in QSPR and QSAR studies. The main advance of these descriptors is that they are based on the same invariants that their 2D analogues, which have been proved to contain important structural information. Thus, the applicability of 3D connectivity indices to any area of QSPR/QSAR in which 2D connectivity indices have been successful is ensured. The information added to the connectivity indices from quantum chemical parameters permits to account not only for 3D information but also for some of the electronic features responsible of the interactions of these molecules in different environments. In fact, we have to consider 3D connectivity indices as extensions of 2D indices in which additional structural information have been used.

We have compared 3D connectivity indices to their 2D analogues and to quantum chemical descriptors in describing two important properties of a set of compounds: the *n*-octanol/water partition coefficients and their antibacterial activity. In both examples we have shown that topographic descriptors produced somewhat better linear regressions than the other two set of descriptors. For instance, the improvement of 29% in the predictive ability of topographic descriptors versus 2D connectivity in describing partition coefficients is highly illustrative.

These 3D connectivity indices could be renamed as quantum-topological connectivity indices or electron-topological connectivity indices because their main features are those of the 2D connectivity indices that have been modified to account for more structural information, in this case coming from quantum-chemical parameters.

ACKNOWLEDGMENT

E.E. thanks the Ministerio de Educacion y Cultura, Spain for a contract as Foreign Young Research in Spain. E.M. thanks the Department of Organic Chemistry, University of Santiago for kind hospitality during his visit. We thank the Spanish Ministry of Science and Technology (PM99-0125) and the Xunta de Galicia (PGIDT00PX120317PR) for partial financial support.

REFERENCES AND NOTES

- (1) Randić, M. On characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.
- (2) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
- (3) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Analysis*; Research Studies Press: Letchworth, 1986.
- (4) Randić, M.; Trinajstić, N. Viewpoint 4–Comparative structure–property studies: the connectivity basis. *J. Mol. Struct. (Theochem)* **1993**, *284*, 209–221.
- (5) Hall, L. H.; Kier, L. B. Molecular connectivity chi indices for database analysis and structure–property modeling. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach: Amsterdam, The Netherlands, 1999; 307–360.
- (6) Zheng, W.; Jin-Chu, S.; Tropsha, A. Rational combinatorial library design. 1. Focus-2D: A new approach to the design of targeted combinatorial chemical libraries. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 251–258.
- (7) Jin-Chu, S.; Zheng, W.; Tropsha, A. Rational combinatorial library design. 2. Rational design of targeted combinatorial peptide libraries using chemical similarity and the inverse QSAR approaches. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 259–268.
- (8) Linusson, A.; Gottfries, J.; Lindgren, F.; Wold, S. Statistical molecular design of building blocks for combinatorial chemistry. *J. Med. Chem.* **2000**, *43*, 1320–1328.
- (9) De Julian-Ortiz, V.; Galvez, J.; Muñoz-Collado, C.; Garcia-Domenech, R.; Gimeno-Cardona, C. Virtual combinatorial síntesis and computational screening of new potential anti-herpes compounds. *J. Med. Chem.* **1999**, *42*, 3308–3314.
- (10) Galvez, J.; Garcia-Domenech, R.; De Julian-Ortiz, V.; Soler, R. Topological approach to drug design. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 272–284.
- (11) Grassy, G.; Calas, B.; Yasri, A.; Lana, R.; Woo, J.; Iyer, S.; Kaczorek, M.; Floc, R.; Buelow, R. Computer-assisted rational design of immunosuppressive compounds. *Nature Biotech.* **1998**, *16*, 748–752.
- (12) Altenburg, K. Eine bemerkung zu dem Randicshen “Molekularen Bindungs-Index (Molecular Connectivity Index)”. *Z. phys. Chem. Leipzig* **1980**, *261*, 389–393.
- (13) Kier, L. B.; Hall, L. H. Derivation and significance of valence molecular connectivity. *J. Pharm. Sci.* **1981**, *70*, 583–589.
- (14) Galvez, J. On a topological interpretation of electronic and vibrational molecular energies. *J. Mol. Struct. (Theochem)* **1998**, *429*, 255–264.
- (15) Pogliani, L. From molecular connectivity indices to semiempirical connectivity terms: Recent trends in graph theoretical descriptors. *Chem. Rev.* **2000**, *100*, 3827–3858.
- (16) Estrada, E. Connectivity polynomial and long-range contributions in the molecular connectivity model. *Chem. Phys. Lett.* **1999**, *312*, 556–560.
- (17) Kier, L. B.; Hall, L. H. Intermolecular accessibility: the meaning of molecular connectivity. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 784–791.
- (18) Petitjean, M. Applications of the radius-diameter diagram to the classification of topological and geometric shapes of chemical compounds. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 331–337.
- (19) Estrada, E.; Molina, E.; Perdomo-Lopez, I. Can 3D parameters be predicted from 2D descriptors? *J. Chem. Inf. Comput. Sci.* **2000**, submitted for publication.
- (20) Bath, P. A.; Poirrette, A. R.; Willett, P. The extent of the relationship between the graph-theoretical and the geometrical shape coefficient of chemical compounds. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 714–716.
- (21) Randić, M. Molecular topographic descriptors. In *MATCH/CHEM/COMP 1987. Proceedings of an International Course and Conference on the Interfaces between Mathematics, Chemistry and Computer Science, Dubrovnik, Yugoslavia, 22–26 June 1987*; Lacher, R. C., Ed.; Elsevier: Amsterdam, pp 101–108.
- (22) Randić, M. On characterization of three-dimensional structures. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1988**, *15*, 201–208.
- (23) Randić, M.; Jerman-Blazić, B.; Trinajstić, N. Development of 3-dimensional molecular descriptors. *Comput. Chem.* **1990**, *14*, 237–246.
- (24) Bogdanov, B.; Nikolić, S.; Trinajstić, N. On the three-dimensional Wiener number. *J. Math. Chem.* **1989**, *3*, 299–309.
- (25) Estrada, E.; Montero, L. A. Bond order weighted graphs in molecules as structure–property indices. *Mol. Eng.* **1993**, *2*, 363–373.
- (26) Estrada, E.; Ramirez, A. Edge adjacency relationships and molecular topographic descriptors. Definition and QSAR applications. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 837–843.
- (27) Estrada, E. Three-dimensional molecular descriptors based on electron charge density weighted graphs. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 708–713.
- (28) Randić, M.; Razinger, M. Molecular topographic indices. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 140–147.
- (29) Diudea, M. V.; Horvath, D.; Graovac, A. Molecular topology. 15. 3D distance matrixes and related topological indices. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 129–135.
- (30) Schultz, H. P.; Schultz, E. S.; Schultz, T. P. Topological organic chemistry. 9. Graph theory and molecular topological indices of stereoisomeric organic compounds. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 864–870.
- (31) Estrada, E. Characterization of 3D molecular structure. *Chem. Phys. Lett.* **2000**, *319*, 713–718.
- (32) Estrada, E. Edge adjacency relationships and a novel topological index related to molecular volume. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 31–33.
- (33) Estrada, E.; Guevara, N.; Gutman, I. Extension of edge connectivity index. Relationships to line graph indices and QSPR applications. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 428–431.
- (34) Estrada, E.; Rodriguez, L. Edge-connectivity indices in QSPR/QSAR studies. 1. Comparison to other topological indices in QSPR studies. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1037–1041.
- (35) Estrada, E. Edge-connectivity indices in QSPR/QSAR studies. 2. Accounting for long-range bond contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1042–1048.
- (36) Kikuchi, O. Systematic QSAR procedures with quantum chemical descriptors. *Quant. Struct.-Act. Relat.* **1987**, *6*, 179–184.
- (37) Rodriguez, L.; Estrada, E. MODEST (MOlecular DESign Tool) for Windows. Version 3.0; 1999–2000. This software is distributed free under request to E. Estrada: estrada66@yahoo.com.
- (38) Dewar, M. J. S.; Zoebish, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- (39) Estrada, E.; Gomez, M.; Castañedo, N.; Perez, C. Theoretical and experimental study on the structure of 1-(5-X-fur-2-yl)-2-nitro-2-Y-ethylenes. *J. Mol. Struct. (Theochem)* **1999**, *468*, 193–200.
- (40) Stewart, J. J. P. MOPAC 6.0, *Quantum Chemistry Program Exchange*; Indiana University, Bloomington, IN, Program 455, 1990.
- (41) Balaz, S.; Sturdik, E.; Rosenberg, M.; Augustin, J.; Skara, B. Kinetics of drug activities as influenced by their physicochemical properties: antibacterial effects of alkylating 2-furylethylenes. *J. Theor. Biol.* **1988**, *131*, 115–134.
- (42) Dore, J. Ch.; Viel, C. Antitumoral chemotherapy. X. Cytotoxic and antitumoral activity of β -nitrostyrenes and nitrovinyl derivatives. *Farmaco* **1975**, *30*, 81–109.
- (43) Sturdik, E.; Drobnica, L.; Balaz, S. Reactions of 2-furylethylenes with thiols in vivo. *Coll. Czech. Chem. Comm.* **1985**, *50*, 470–480.
- (44) Castañedo, N.; Goizueta, R.; Perez, J.; Gonzalez, J.; Silveira, E. Cuesta, M.; Martinez, A.; Lugo, E.; Estrada, E.; Carta, A.; Navia, O.; Delgado, M. Cuban patent 22446, 1994; European patent, application number 955000567; Canadian patent, application number 2,147,594; Japan patent, application number 222002; U.S. patent, application number 60008011.
- (45) Blondeau, J. M.; Castañedo, N.; Gonzalez, O.; Medina, R.; Silveira, E. In vitro evaluation of G-1: A novel antimicrobial compound. *Antimicrob. Agents Chemother.* **1999**, *11*, 16663–1669.
- (46) McCoy, W. F.; Thornburgh, S. Industrial antimicrobial uses for 2-(2-bromo-2-nitroethenyl)furan. U.S. patent 5,090,990, 1992.
- (47) Estrada, E. Structure-mutagenicity relationships in 2-furylethylene derivatives. A molecular orbital study of the role of nitro groups. *Mutat. Res.* **1998**, *420*, 67–75.
- (48) Yang, G.; Lien, E. J.; Guo, Z. Physical factors contributions to hydrophobic constant π . *Quant. Struct.-Act. Relat.* **1986**, *5*, 12–18.
- (49) Rogers, K. S.; Cammarata, A. Superdelocalizability and charge density. A correlation with partition coefficients. *J. Med. Chem.* **1969**, *12*, 692–693.
- (50) Internal Report of the Centro de Bioactivos Quimicos, Universidad Central de Las Villas. Santa Clara, Cuba.