Inter*net* Electronic Journal of Molecular Design

March 2002, Volume 1, Number 3, Pages 115–133

Editor: Ovidiu Ivanciuc

Special issue dedicated to Professor Alexandru T. Balaban on the occasion of the 70th birthday
Part 3

Guest Editor: Mircea V. Diudea

QSAR Carcinogenic Study of Methylated Polycyclic Aromatic Hydrocarbons Based on Topological Descriptors Derived from Distance Matrices and Correlation Weights of Local Graph Invariants

Damián J. G. Marino,[1] Pablo J. Peruzzo,[1] Eduardo A. Castro,[2] and Andrey A. Toropov[3]

[1] C.I.M.A., Departamento de Química, Facultad de Ciencias Exactas, UNLP, Calles 47 y 115, La Plata 1900, Argentina
[2] CEQUINOR, Departamento de Química, Facultad de Ciencias Exactas, UNLP, Calles 47 y 115, C.C. 962, La Plata 1900, Argentina
[3] Vostok Innovation Company, S. Azimstreet 4, 700047 Tashkent, Uzbekistan

Received: October 8, 2001; Revised: December 12, 2001; Accepted: January 14, 2002; Published: March 31, 2002

Citation of the article:
D. J. G. Marino, P. J. Peruzzo, E. A. Castro, and A. A. Toropov, QSAR Carcinogenic Study of Methylated Polycyclic Aromatic Hydrocarbons Based on Topological Descriptors Derived from Distance Matrices and Correlation Weights of Local Graph Invariants, *Internet Electron. J. Mol. Des.* **2002**, *1*, 115–133, http://www.biochempress.com.

Inter*net* Electronic Journal of Molecular Design

BioChem Press
http://www.biochempress.com

QSAR Carcinogenic Study of Methylated Polycyclic Aromatic Hydrocarbons Based on Topological Descriptors Derived from Distance Matrices and Correlation Weights of Local Graph Invariants[#]

Damián J. G. Marino,[1] Pablo J. Peruzzo,[1] Eduardo A. Castro,[2,*] and Andrey A. Toropov[3]

[1] C.I.M.A., Departamento de Química, Facultad de Ciencias Exactas, UNLP, Calles 47 y 115, La Plata 1900, Argentina
[2] CEQUINOR, Departamento de Química, Facultad de Ciencias Exactas, UNLP, Calles 47 y 115, C.C. 962, La Plata 1900, Argentina
[3] Vostok Innovation Company, S. Azimstreet 4, 700047 Tashkent, Uzbekistan

Received: October 8, 2001; Revised: December 12, 2001; Accepted: January 14, 2002; Published: March 31, 2002

Internet Electron. J. Mol. Des. **2002**, *1* (3), 115–133

Abstract

A quantitative structure–activity study for the carcinogenic activity of methylated polycyclic aromatic hydrocarbons is made on the basis of topological molecular descriptors derived from distance matrices and optimized correlation weights of local graph invariants. The multilinear regression equations allow us to predict correctly the carcinogenic activity of this set of compounds. Comparison with results derived from other theoretical studies show a quite satisfactory behavior of the present method. Some possible future extensions are pointed out.

Keywords. Methylated polycyclic aromatic hydrocarbons; topological index; quantitative structure–activity relationships; QSAR; distance matrix; correlation weights of local graph invariants.

1 INTRODUCTION

The biological responses to aromatic hydrocarbons are characterized by a high degree of selectivity. A simple feeding of a massive but tolerable amount of some of the polynuclear aromatic hydrocarbons (PAHs) to rats yields pathological lesions representing cancer and necrosis. While certain cells become malignant and massive areas of cellular death are found in other organs, the generality of the cells of the body are uninvolved and emerge quite uninjured from their encounter with the conjugated hydrocarbons [1]. The possibility that carcinogenic aromatic hydrocarbons

[#] Dedicated on the occasion of the 70th birthday to Professor Alexandru T. Balaban.

* Correspondence author; fax: 54–(0)221–425–9485; E–mail: castro@dalton.quimica.unlp.edu.ar.

might function at the molecular level by the formation of a charge transfer complex as a crucial event was first considered by Pullman and Pullman [2]. They concluded that there was no general relationship between charge transfer capability and carcinogenesis, although some rough correlations were observed in certain limited series of molecules.

PAHs represent a very interesting molecular set since they have a widely variable carcinogenic power, encompassing very strong carcinogens to rather inactive ones [3–5]. Reasons why some of these particularly resembling molecules possess carcinogenic activity while others do not have aroused a deep interest of the researchers and motivated intense efforts since the pioneering work of Cook [4]. During the last fifty years many theoretical approaches have been employed to test whether a given PAH possess or not carcinogenic activity [5]. In fact, theories using electronic and structural properties, statistical analysis, artificial intelligence, neural networks, and electronic indices methods were proposed and applied for this QSAR studies. Notwithstanding, practically all of them present some well define deficiencies and cannot describe the carcinogenic properties in satisfactory agreement with the available experimental data. One of the latest and more successful methodologies to group and identify PAHs carcinogenic activity is formulated in terms of very simple rules based on the concept of electronic local density of states ever specific molecular regions [5–8]. The authors have studied 81 methylated and non–methylated PAHs and their studies shoved that with the use of electronic indices with principal component analysis and artificial neural networks, it is possible to predict correctly the carcinogenic activity with a relative high accuracy, around 80%.

The aim of this study is to explore the possibility to ameliorate those predictions through the employment of suitable topological descriptors based on the detour matrix and correlation weighting of local graph invariants. This proposal arises from several previous results obtained in our laboratory that demonstrated improvements over predictions obtained from the usual distance matrices and rigid topological descriptors [9–17].

2 MATERIALS AND METHODS

2.1 Basic Definitions

A recent trend in mathematical chemistry, chemical graph theory, QSAR/QSPR studies as well as predictive toxicology is the employment of graph theoretical invariants for the characterization of structure and prediction of properties. Graph theoretic indices have been used for isomer discrimination and description of molecular frameworks [18–20], ordering of physicochemical properties, biomedical, and toxicological properties [21–31]. In order to judge the value of this sort of studies and the need to resort to results derived from graph theoretical invariants, one can mention, for example, that in environmental toxicology the Toxic Substances Control Act Inventory

has more than seventy six thousands entries and most of these chemicals do not have the experimental data necessary for their hazard assessment [32,33]. Besides, more than 15 million distinct chemical entities have been registered with the Chemical Abstract Service and the list is growing by nearly 775,000 per year. About 1,000 of these chemicals enter into use every year [34]. Few of these chemicals have the empirical data needed of risk assessment.

The Randić connectivity index, the higher order indices defined by Kier and Hall, and the Balaban index *J* are derived from the adjacency and distance matrices of molecular graphs [19,20]. Two types of graphs, namely hydrogen–suppressed graphs and hydrogen–filled graphs, are often used to model molecular structure in chemistry. While in the former only the non–hydrogen atoms are considered, in the latter all atoms are represented by graph vertices. We show in Figure 1 the hydrogen–filled graph and hydrogen–suppressed graphs (G_1 and G_2, respectively) of 2–methylbutane.

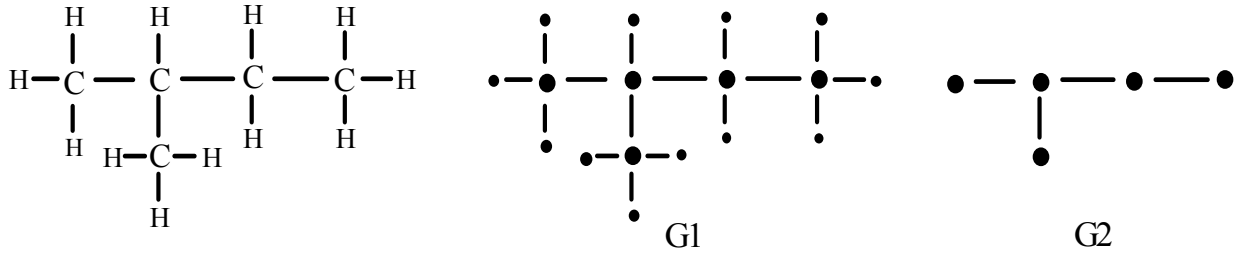


Figure 1. Hydrogen–suppressed and hydrogen–filled graph of methylbutane.

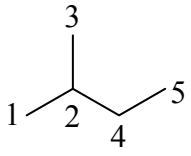
A graph *G* is a set of vertices (points or atoms, *V*) and edges (lines or bonds, *E*), which in terms of the mathematical language is written as

$$G = G(V, E) \tag{1}$$

G can be represented either by a geometrical or algebraic object, *i.e.* a matrix. The adjacency matrix **A** associated with a given graph *G* is:

$$\mathbf{A}_{ij} = \begin{Bmatrix} 1 \; for \; an \;\; adjacent \;\; pair \; of \;\; vertices \; v_i \; and \; v_j \\ 0 \; otherwise \end{Bmatrix} \tag{2}$$

In Figure 2 we represent the adjacency matrix **A** corresponding to methylbutane.

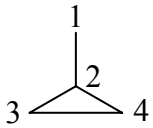


$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

Figure 2. Adjacency matrix for methylbutane.

Two important molecular graph matrices are the distance matrix **D** and the detour matrix **Δ**

[35,36]. Another closely related distance matrix is that proposed by Rücker and Rücker [37]. They argued that the detour matrix has zeros as the diagonal elements without a convincing reason. So, it is more logical to define the diagonal elements as the length of the longest path from vertex *i* to itself (*i.e.* the size of the longest cycle containing the vertex *i*). We present below examples for these three matrices for methyl–cyclopropane.



$$D = \begin{bmatrix} 0 & 1 & 2 & 2 \\ 1 & 0 & 1 & 1 \\ 2 & 1 & 0 & 1 \\ 2 & 1 & 1 & 0 \end{bmatrix} \quad \Delta = \begin{bmatrix} 0 & 1 & 3 & 3 \\ 1 & 0 & 2 & 2 \\ 3 & 2 & 0 & 2 \\ 3 & 2 & 2 & 0 \end{bmatrix} \quad \Delta^{*} = \begin{bmatrix} 0 & 1 & 3 & 3 \\ 1 & 3 & 2 & 2 \\ 3 & 2 & 3 & 2 \\ 3 & 2 & 2 & 3 \end{bmatrix}$$

The term topological index was proposed by Hosoya [38] for characterizing the topological nature of a graph. It is an integer quite easily obtained from a graph by the specified recipe. As pointed out before, there have been proposed more than one hundred different topological indices for chemical graphs [39–72]. Among them let us introduce here only those chosen in the present study.

2.1.1 The Wiener Index *W*

The Wiener index $W(w) = W(w,G)$ of a vertex– and edge–weighted graph G with N vertices is [60]:

$$W(w,G) = \sum_{i=1}^{N} \sum_{j=i}^{N} [\mathbf{D}(w,G)]_{ij} \tag{3}$$

where the distance matrix $\mathbf{D}(w)$ is computed with the weighting scheme w.

2.1.2 The Harary Index *H*

The reciprocal distance matrix $\mathbf{RD}(w) = \mathbf{RD}(w,G)$ of a vertex– and edge–weighted molecular graph G with N vertices is the square $N \times N$ symmetric matrix with real elements defined with the equation [58,59]:

$$[\mathbf{RD}(w,G)]_{ij} = \begin{cases} 1/[\mathbf{D}(w,G)]_{ij} & \text{if} \quad i \neq j \\ [\mathbf{D}(w,G)]_{ii} & \text{if} \quad i = j \end{cases} \tag{4}$$

where $[\mathbf{D}(w)]_{ij}$ is the graph distance between vertices v_i and v_j, $[\mathbf{D}(w)]_{ii}$ is the diagonal element corresponding to vertex v_i, and w is the weighting scheme used to compute the parameters *Vw* and *Ew*. The Harary index is the sum of the diagonal and upper–triangle elements from the reciprocal distance matrix [59,61]:

$$H = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=i}^{N} [\mathbf{RD}(w,G)]_{ij} \tag{5}$$

2.1.3 The Balaban Index *J*

The Balaban index (also called the average distance sum connectivity index) is a sort of Randić–type formula [62] applied to distance sums [63]:

$$J = \frac{M}{\mu + 1} \sum_{E(G)} \left(\mathbf{DS}_i \mathbf{DS}_j \right)^{-1/2} \tag{6}$$

where $\mathbf{DS}_i$ and $\mathbf{DS}_j$ denote the distance sums of the vertices v_i and v_j of an edge e_{ij} in the molecular graph G, M is the number of edges in the molecular graph, μ is the cyclomatic number, and the summation goes over all edges in the molecular graph, $E(G)$. The distance sum of the vertex v_i, $\mathbf{DS}_i$, is defined as the sum of the topological distances between vertex v_i and every vertex in the molecular graph, *i.e.* the sum over row i or column i in the **D** matrix:

$$\mathbf{DS}_i = \sum_{j=1}^{N} [\mathbf{D}]_{ij} = \sum_{j=1}^{N} [\mathbf{D}]_{ji} \tag{7}$$

2.1.4 The Molecular-topological Schultz index *MTI*

This index is defined as [64–66]:

$$MTI = \sum_{i=1}^{N} a_i \tag{8}$$

where a_i are elements of the row matrix

$$v(A + D) = (a_1, a_2, ..., a_n) \tag{9}$$

with v the valence row matrix, *i.e.* the vector of all the vertex degrees of a given graph. Thus, the 2–methylbutane valence row matrix is (see Figure 2) v = (1, 3, 1, 2, 1). As a natural extension of the definitions (3), (5), (6) and (8) based upon matrix **D**, we can define the associated indices **Wi**(**Δ**), **Wi**(**RΔ**), J(**Δ**) and MTI(**Δ**) on the basis of the detour matrix and **Wi**(**Δ***), **Wi**(**RΔ***), J(**Δ***) and MTI(**Δ***) indices computed from the Rücker matrix.

2.2 Variable Topological Descriptors

Previous results [72] show clearly that simple regression involving a single descriptor restricts regression analysis considerably. Many correlations, particularly when involving molecules of different size, need not be linear. But even if we have molecules of the same or similar size, a quadratic regression may result in a better description of the relationship than a simple linear model [73–83]. In general, one should test single or multiple regression analysis for quadratic dependence and, if warranted, for higher order polynomial relationships or other functional dependence.

Flexible topological descriptors based on the optimization of correlation weights of local graph invariants (OCWLGI) represent a valuable set of descriptors to use in QSAR/QSPR studies [9–17]. The inherent flexibility of these descriptors seems to be rather suitable to obtain satisfactory

predictions of the properties and activities under study. The basic outline of the optimization procedure is as follows [84,85]. The primary units of analysis are the atoms with their corresponding vertex degrees. Then, the graph invariants are formulated in the general form:

$$MD = f\{CW(a_i), CW(VD_i)\} \tag{10}$$

where *MD* is the molecular descriptor,

$$a_i = \sum_{j \text{ joined to } i} a_{ij} \tag{11}$$

$VD_{(i)}$ is the extended connectivity value of the *i*–th vertex, defined as

$$VD_i = \sum_{j \text{ joined to } i} a(f) \tag{12}$$

$CW(a_i)$ and $CW(VD_i)$ are the correlation weights corresponding to atom *i*. *CW* descriptors are calculated by means of an optimization procedure, *i.e.* they are determined in such a way to give the best correlation coefficient for the relationship under consideration by way of a trial an error procedure.

$$P = F(MD) \tag{13}$$

where *P* stands for the physical chemistry property or biological activity. There is complete freedom to choose the algebraic form of *f* and *F* functions. The most general polynomial form of the *F* function is

$$F = \sum_{k=0} A_k P^k \quad , \quad n \in N \tag{14}$$

while there are several possibilities to choose *f*. Some of the most simple equations for *MD* are

$$MD = \sum_{edge} \{CW(a_i) + CW(VD_i)\} \tag{15}$$

$$MD = \sum_{edge} \{CW(a_i) \times CW(VD_i)\} \tag{16}$$

$$MD = \prod_{edge} \{CW(a_i) + CW(VD_i)\} \tag{17}$$

$$MD = \prod_{edge} \{CW(a_i) \times CW(VD_i)\} \tag{18}$$

$$MD = \sum_{(i,j)} \{CW(a_i) \times CW(VD_i) + CW(a_j) \times CW(VD_j)\} \quad , \quad (i,j) \in E(G) \tag{19}$$

After computing *CW* values, one resort to relationship (13) to calculate the final correlation formula through a least squares procedure (*i.e.* to determine the optimum coefficients A_k for a molecular training set. Finally, the predictive capability of the method is tested with a different molecular set, *i.e.* the test set. In this study, *OCW* is based on hydrogen–filled chemical graphs, *i.e.* G_1 in Figure 1. As local graph invariants for calculations we have chosen the following ones:

$$VD0(i) = number\ of(i, j) \tag{20}$$

$$VD1(i) = \sum_{(i,j)} VD0(j) \tag{21}$$

$$VD2(i) = \sum_{(i,j)} VD1(j) \tag{22}$$

$$VD3(i) = \sum_{(i,j)} VD2(j) \tag{23}$$

where (*i*,*j*) is an edge. We can see that *VDX* (*X* = 1, 2, 3) is a modified version of the Morgan vertex degree, *i.e.* extended connectivity [86]. Another group of local graph invariants used to define the *DCW* descriptor are the following ones:

$$pt20(i) = number\ of(i, j, k) \tag{24}$$

where *j* and *k* represent non–hydrogen vertices and (*i*,*j*) and (*j*,*k*) are edges.

$$pt21(1) = \sum_{(i,j)} pt20(j) \tag{25}$$

$$pt22(1) = \sum_{(i,j)} pt21(j) \tag{26}$$

An example for computing these descriptors is presented for 6,12-dimethylanthanthrene (Figure 3) in Tables 1 and 2.

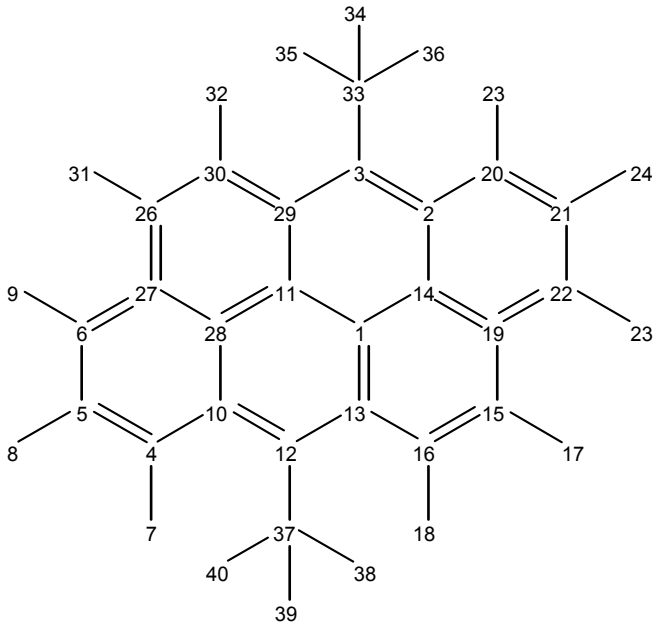


Figure 3. Molecular graph of 6,12-dimethylanthanthrene.

2.3 Carcinogenic Activity of Methylated PAHs

It is well know that PAHs are a class of molecules that can induce chemical carcinogenesis. In order to apply our approach based on topological molecular descriptors derived from the distance

and detour matrices and *OCWLGI* (optimized correlation weights of local graph invariants) to identify carcinogenic activity in methylated PAHs, we have chosen a set of 49 molecules studied by Galvão [5–8] to apply several methodologies in structure–activity studies of PAHs compounds.

Table 1. *CW*(*LGI*) and *CW*(*LGI*)´s Values for *LGI* = *VD2*, *pt21*

*VD*2		*pt*21		*VD*2	*CW*(*VD*2)	*pt*21	*CW*(*pt*21)
a_i	CW(a_i)	a_i	CW(a_i)	15	0.469	10	0.287
C	–0.637	C	–0.613	16	0.616	11	–0.100
H	–0.812	H	–0.725	17	1.147	12	1.315
*VD*2	*CW*(*VD*2)	*pt*21	*CW*(*pt*21)	18	2.083	13	1.169
5	–0.941	2	0.037	19	–0.497	14	2.798
6	0.375	3	0.363	20	–0.063	15	2.726
7	0.400	4	0.710	21	0.955	16	2.004
8	0.381	5	2.671	22	1.062	17	1.688
9	0.413	6	2.988	23	2.467		
10	0.836	7	3.007	24	2.407		
13	2.549	8	0.662	25	1.402		
14	0.025	9	0.500	26	1.963		

Table 2. *DCW* Descriptors Calculated for 6,12-Dimethylanthanthrene

Number	*VD*0	*VD*1	*VD*2	*VD*3	*pt*20	*pt*21	*pt*22	*CW*(a_i) to *VD*2	*CW*(*VD*2)	*CW*(a_i) to *pt*21	*CW*(*pt*21)
C1	3	9	26	72	6	17	45	–0.637	1.963	–0.613	1.688
C2	3	8	22	61	5	13	36	–0.637	1.062	–0.613	1.169
C3	3	7	22	54	4	12	33	–0.637	1.062	–0.613	1.315
C4	2	6	15	42	3	9	24	–0.637	0.469	–0.613	0.500
C5	2	5	14	34	2	8	19	–0.637	0.025	–0.613	0.662
C6	2	6	14	41	3	8	23	–0.637	0.025	–0.613	0.662
H7	1	2	6	15	2	3	9	–0.812	0.375	–0.725	0.363
H8	1	2	5	14	2	2	8	–0.812	–0.941	–0.725	0.037
H9	1	2	6	14	2	3	8	–0.812	0.375	–0.725	0.363
C10	3	8	22	61	5	13	36	–0.637	1.062	–0.613	1.169
C11	3	9	26	72	6	17	45	–0.637	1.963	–0.613	1.688
C12	3	7	22	54	4	12	33	–0.637	1.062	–0.613	1.315
C13	3	8	22	64	5	13	39	–0.637	1.062	–0.613	1.169
C14	3	9	24	69	6	15	42	–0.637	2.407	–0.613	2.726
C15	2	6	15	43	3	9	25	–0.637	0.469	–0.613	0.500
C16	2	6	16	43	3	10	25	–0.637	0.616	–0.613	0.287
H17	1	2	6	15	2	3	9	–0.812	0.375	–0.725	0.363
H18	1	2	6	16	2	3	10	–0.812	0.375	–0.725	0.363
C19	3	7	21	53	4	12	32	–0.637	0.955	–0.613	1.315
C20	2	6	15	42	3	9	24	–0.637	0.469	–0.613	0.500
C21	2	5	14	34	2	8	19	–0.637	0.025	–0.613	0.662
C22	2	6	14	41	3	8	23	–0.637	0.025	–0.613	0.662
H23	1	2	6	15	2	3	9	–0.812	0.375	–0.725	0.363
H24	1	2	5	14	2	2	8	–0.812	–0.941	–0.725	0.037
H25	1	2	6	14	2	3	8	–0.812	0.375	–0.725	0.363
C26	2	6	15	43	3	9	25	–0.637	0.469	–0.613	0.500
C27	3	7	21	53	4	12	32	–0.637	0.955	–0.613	1.315
C28	3	9	24	69	6	15	42	–0.637	2.407	–0.613	2.726
C29	3	8	22	64	5	13	39	–0.637	1.062	–0.613	1.169
C30	2	6	16	43	3	10	25	–0.637	0.616	–0.613	0.287
H31	1	2	6	15	2	3	9	–0.812	0.375	–0.725	0.363

Table 2. (Continued)

Number	*VD*0	*VD*1	*VD*2	*VD*3	*pt*20	*pt*21	*pt*22	$CW(a_i)$ to *VD*2	*CW*(*VD*2)	$CW(a_i)$ to *pt*21	*CW*(*pt*21)
H32	1	2	6	16	2	3	10	–0.812	0.375	–0.725	0.363
C33	1	6	10	40	2	7	18	–0.637	0.836	–0.613	3.007
H34	1	1	6	10	1	2	7	–0.812	0.375	–0.725	0.037
H35	1	1	6	10	1	2	7	–0.812	0.375	–0.725	0.037
H36	1	1	6	10	1	2	7	–0.812	0.375	–0.725	0.037
C37	1	6	10	40	2	7	18	–0.637	0.836	–0.613	3.007
H38	1	1	6	10	1	2	7	–0.812	0.375	–0.725	0.037
H39	1	1	6	10	1	2	7	–0.812	0.375	–0.725	0.037
H40	1	1	6	10	1	2	7	–0.812	0.375	–0.725	0.037
Σ*CW*								–28.028	25.270	–26.312	33.200
MD ecuation (15)								*DCW*(*VD*2)= –3.010		*DCW*(*pt*21)= 6.888	

Table 3. Carcinogenic Activity *CA* from Ref. [87] for 49 Methylated Polycyclic Aromatic Hydrocarbons[a]

Molecule	*CA*	Molecule	*CA*]
(1) 7,12–Dimethylbenz[*a*]anthracene	A	(26) 3–Methylbenzo[*c*]phenanthrene	I
(2) 6,12–Dimethylbenz[*a*]anthracene	A	(27) 6–Methylbenzo[*c*]phenanthrene	I
(3) 6,8,12–Trimethylbenz[*a*]anthracene	A	(28) 6–Methylbenz[*a*]anthracene	I
(4) 2–Methylbenz[*a*]pyrene	A	(29) 12–Methylbenz[*a*]anthracene	I
(5) 4–Methylbenzo[*a*]pyrene	A	(30) 6-Methylanthanthrene	I
(6) 11–Methylbenzo[*a*]pyrene	A	(31) 6,12-Dimethylanthanthrene	I
(7) 12–Methylbenzo[*a*]pyrene	A	(32) 1–Methylbenzo[*c*]phenanthrene	I
(8) 1–Methylbenzo[*a*]pyrene	A	(33) 2–Methylbenzo[*c*]phenanthrene	I
(9) 4,5–Dimethylbenzo[*a*]pyrene	A	(34) 10–Methylbenzo[*a*]pyrene	I
(10) 3–Methylbenzo[*a*]pyrene	A	(35) 6–Methylchrysene	I
(11) 1,2–Dimethylbenzo[*a*]pyrene	A	(36) 3–Methylbenz[*a*]anthracene	I
(12) 2,3–Dimethylbenzo[*a*]pyrene	A	(37) 1–Methylbenz[*a*]anthracene	I
(13) 3,12–Dimethylbenzo[*a*]pyrene	A	(38) 11–Methylbenz[*a*]anthracene	I
(14) 1,3–Dimethylbenzo[*a*]pyrene	A	(39) 9–Methylbenz[*a*]anthracene	I
(15) 1,4–Dimethylbenzo[*a*]pyrene	A	(40) 2–Methylbenz[*a*]anthracene	I
(16) 5–Methylbenzo[*c*]phenanthrene	A	(41) 5–Methylbenz[*a*]anthracene	I
(17) 5–Methylchrysene	A	(42) 8–Methylbenz[*a*]anthracene	I
(18) 6,8–Dimethylbenz[*a*]anthracene	A	(43) 2–Methylpyrene	I
(19) 7–Methylbenz[*a*]anthracene	A	(44) 4–Methylpyrene	I
(20) 5–Methylbenzo[*a*]pyrene	A	(45) 1–Methylpyrene	I
(21) 7–Methylbenzo[*a*]pyrene	A	(46) 7,10–Dimethylbenzo[*a*]pyrene	I
(22) 6–Methylbenzo[*a*]pyrene	A	(47) 6,10–Dimethylbenzo[*a*]pyrene	NA
(23) 1,6–Dimethylbenzo[*a*]pyrene	A	(48) 8–Methylbenzo[*a*]pyrene	NA
(24) 3,6–Dimethylbenzo[*a*]pyrene	A	(49) 9–Methylbenzo[*a*]pyrene	NA
(25) 4–Methylbenzo[*c*]phenanthrene	I		

[a] See Figure 4 for the molecular structures. The carcinogenic activity data are adapted from Cavaliere [87]. A and I refer to active and inactive, respectively. The carcinogenic activity of the last three molecules is not available (NA).

This particular choice is due to three main reasons: (*a*) this set is a suitable group of representative PAHs molecules; (*b*) there are experimental data available for 46 molecules and for the remaining three molecules there are not any data on carcinogenic activity, so that they make up an interesting enough subset to make predictions; (*c*) we can perform a direct comparison between our results, those derived by Galvão [5–8] and the experimental available data. Since there are some intriguing discrepancies between previous theoretical results and experimental data, they constitute a severe test conditions for our predictions. We display in Figure 4 the molecular structures of the 46 methylated PAHs and in Table 3 we list their IUPAC names together with the carcinogenic

activity, adapted from Cavaliere [87].

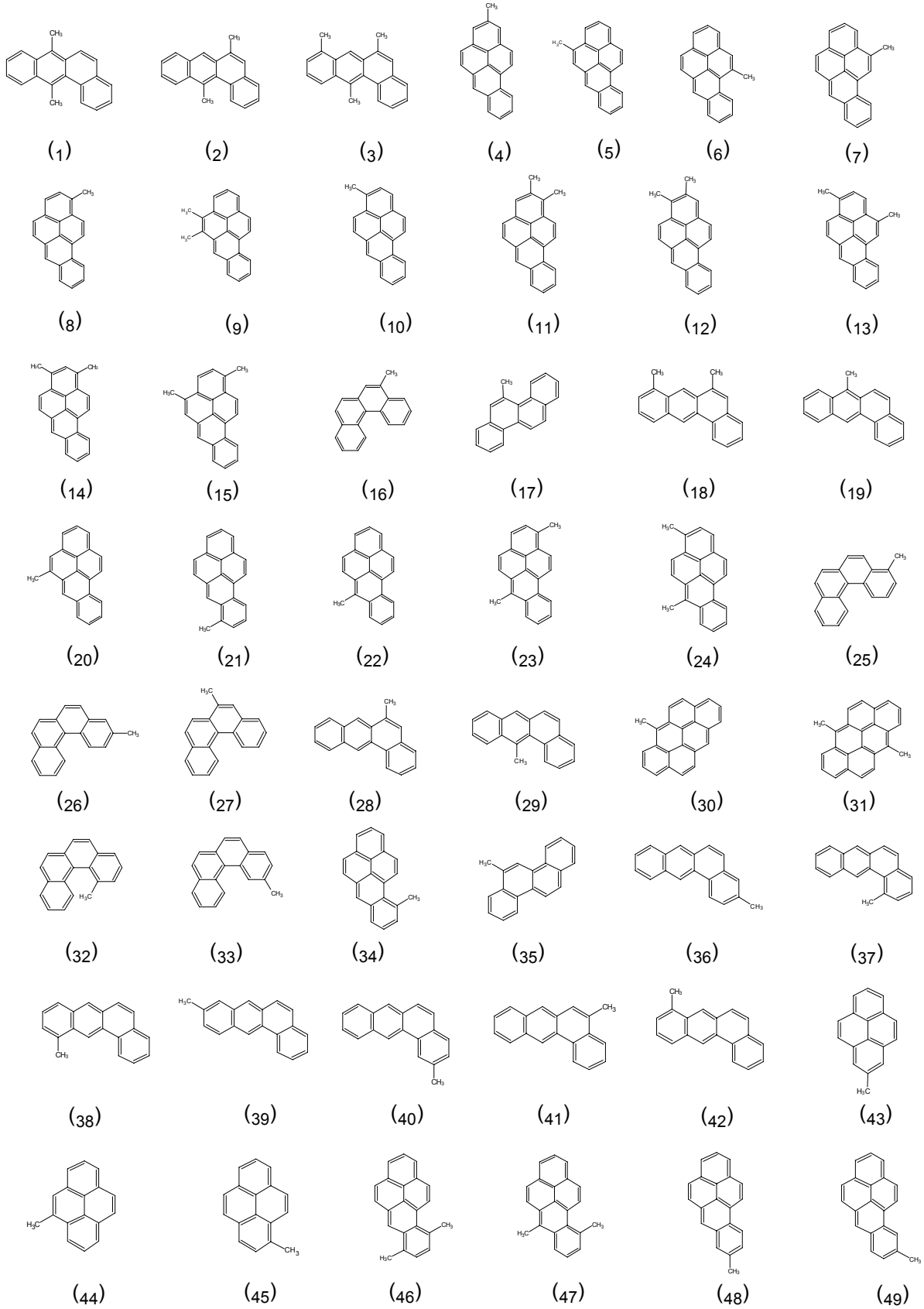


Figure 4. Molecular structures of 49 methylated polycyclic aromatic hydrocarbon (PAH) molecules.

We have tried several partitions of the whole set of 49 molecules into a training set and a test set and we have found that final results do not depend significantly the specific way to choose the members of each subset. Here we report results for the following choice. Training set: (**1**), (**5**), (**9**)–(**15**), (**20**)–(**28**), (**31**)–(**33**), (**35**), (**37**), (**38**), (**41**)–(**46**). Test set: (**2**)–(**4**), (**6**)–(**8**), (**16**)–(**19**), (**29**), (**30**), (**34**), (**36**), (**39**), (**40**). The carcinogenic activity scale proposed be Cavalieri [87] is as follows: extremely active, (+++++); very active, (++++); active, (+++); moderately active, (++); weakly active, (+); very weakly active, (+–); inactive, (–). In order to obtain a numerical scale for the carcinogenic activity of the methylated PAHs and thus to obtain true QSAR models, we propose the scale: extremely active, 5; very active, 4; active, 3; moderately active, 2; weakly active, 1; very weakly active, 0; inactive, –1.

3 RESULTS

The elements of both detour matrix and Rücker matrix have been computed with the algorithm presented by Toropov [88]. This procedure is based on the Monte Carlo technique and results are defined on the number of random walks starting from a given vertex *i* and arriving to another vertex *j*. We give in Table 4 the results of three running probes of the algorithm as a function of the number of random walks. Numerical data shows that the minimum number of random walks to yield stable values for the topological descriptors under consideration is 1500.

Table 4. Calculation of the Detour Matrix Descriptors of 6,12-Dimethylanthanthrene for 1500, 1000, 500, 100, and 50 Random Walks

Number of Walks	W(Δ)	H(Δ)	J(Δ)	MTI(Δ)	W(Δ^*)	H(Δ^*)	J(Δ^*)	MTI(Δ^*)
	5226	16.449	0.556	25262	5446	16.999	1.270	26382
1500	5226	16.449	0.556	25262	5446	16.999	1.270	26382
	5226	16.449	0.556	25262	5446	16.999	1.270	26382
	5226	16.449	0.556	25262	5446	16.999	1.270	26382
1000	5226	16.449	0.556	25262	5445	17.002	1.271	26376
	5226	16.449	0.556	25262	5445	17.002	1.271	26376
	5218	16.474	0.557	25216	5438	17.024	1.272	26336
500	5220	16.467	0.556	25230	5440	17.017	1.272	26350
	5224	16.454	0.556	25250	5444	17.004	1.271	26370
	5088	16.888	0.571	24584	5295	17.475	1.307	25642
100	5052	16.992	0.574	24428	5260	17.575	1.315	25488
	5084	16.912	0.571	24564	5290	17.502	1.308	25612
	4936	17.438	0.590	23802	5134	18.055	1.348	24812
50	4896	17.568	0.593	23688	5100	18.166	1.357	24720
	4926	17.426	0.590	23782	5117	18.078	1.352	24756

An illustrative calculation of the descriptors for 6,12-dimethylanthanthrene with the detour matrix on the basis of the 1500 random walks are given in Table 5, while the corresponding hydrogen–suppressed chemical graph is displayed in Figure 5.

Previous QSAR results have shown that there are three major outliers: 7,10–dimethylbenzo[*a*]pyrene (training set), 7–methybenz[*a*]anthracene (test set) and 10–methylbenzo–

[*a*]pyrene (test set).

Table 5. The Detour Matrix on 6,12-Dimethylanthanthrene

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	$DS_i(\Delta)$	$DS_i(\Delta^*)$
1	0	18	19	20	19	20	17	19	20	19	19	19	18	16	19	18	19	20	17	20	18	19	20	21	434	454
2	18	0	19	18	17	18	17	17	18	17	19	19	18	16	19	18	17	18	15	18	18	17	20	19	410	430
3	19	19	0	19	20	21	18	20	21	20	20	20	19	17	20	19	20	19	18	21	19	20	1	22	432	452
4	20	18	19	0	19	18	19	19	20	19	21	21	20	18	21	20	19	20	17	20	18	19	20	21	446	466
5	19	17	20	19	0	19	18	18	19	18	20	20	19	17	20	19	20	19	18	19	19	18	21	20	436	456
6	20	18	21	18	19	0	17	19	20	19	21	19	20	18	19	20	21	18	19	20	20	19	22	21	448	468
7	17	17	18	19	18	17	0	18	19	18	18	18	17	15	18	17	18	19	16	19	17	18	19	20	410	430
8	19	17	20	19	18	19	18	0	19	18	20	20	19	17	20	19	20	19	16	19	19	18	21	20	434	454
9	20	18	21	20	19	20	19	19	0	19	21	19	20	18	19	20	21	20	17	20	20	19	22	1	432	452
10	19	17	20	19	18	19	18	18	19	0	20	18	19	17	18	19	20	19	16	19	19	18	21	20	430	450
11	19	19	20	21	20	21	18	20	21	20	0	20	19	17	20	19	20	21	18	19	19	20	21	22	454	474
12	19	19	20	21	20	19	18	20	19	18	20	0	19	19	20	19	18	21	18	21	19	20	21	20	448	468
13	18	18	19	20	19	20	17	19	20	19	19	19	0	18	19	18	19	20	17	20	18	19	20	21	436	456
14	16	16	17	18	17	18	15	17	18	17	17	19	18	0	17	18	19	18	15	18	16	17	18	19	398	418
15	19	19	20	21	20	19	18	20	19	18	20	20	19	17	0	19	18	21	18	21	19	20	21	20	446	466
16	18	18	19	20	19	20	17	19	20	19	19	19	18	18	19	0	19	20	17	20	18	19	20	21	436	456
17	19	17	20	19	20	21	18	20	21	20	20	18	19	19	18	19	0	19	18	21	19	20	21	22	448	468
18	20	18	19	20	19	18	19	19	20	19	21	21	20	18	21	20	19	0	19	20	18	19	20	21	448	468
19	17	15	18	17	18	19	16	16	17	16	18	18	17	15	18	17	18	19	0	17	17	18	19	18	398	418
20	20	18	21	20	19	20	19	19	20	19	19	21	20	18	21	20	21	20	17	0	20	19	22	21	454	474
21	19	18	19	18	19	20	17	19	20	19	19	19	18	16	19	18	19	18	17	20	0	19	20	21	430	450
22	19	17	20	19	18	19	18	18	19	18	20	20	19	17	20	19	20	19	18	19	19	0	21	20	436	456
23	20	20	1	20	21	22	19	21	22	21	21	21	20	18	21	20	21	20	19	22	20	21	0	23	454	454
24	21	19	22	21	20	21	20	20	1	20	22	20	21	19	20	21	22	21	18	21	21	20	23	0	454	454

$a_i(\Delta)$ = (1038 981 1060 1084 1057 1090 981 1038 1060 1029 1084 1090 1058 948 1084 1057 10901090 948 1084 1029 1058 1112 1112) $MTI(\Delta)$ = 25262

$a_i(\Delta^*)$ = (1098 1041 1120 1124 1097 1130 1041 1098 1120 1089 1144 1130 1098 1008 1124 1097 1130 1130 1008 1144 1089 1098 1112 1112) $MTI(\Delta^*)$ = 26382

$Wi(\Delta)$ = 5226 $Wi(\Delta^*)$ = 5446 $J(\Delta)$ = 0.5555 $J(\Delta^*)$ = 1.2703

where Δ is the detour matrix, Δ^* is the Rücker matrix, $\mathbf{DS}_i(\Delta)$, $a_i(\Delta)$, $MTI(\Delta)$, $W(\Delta)$, $J(\Delta)$ are the distance sums, elements of the row matrix, Schultz index, Wiener index and Balaban index, respectively, on the basis of the detour matrix, and $\mathbf{DS}_i(\Delta^*)$, $a_i(\Delta^*)$, $MTI(\Delta^*)$, $W(\Delta^*)$ and $J(\Delta^*)$ are the same indices computed from the Rücker matrix.

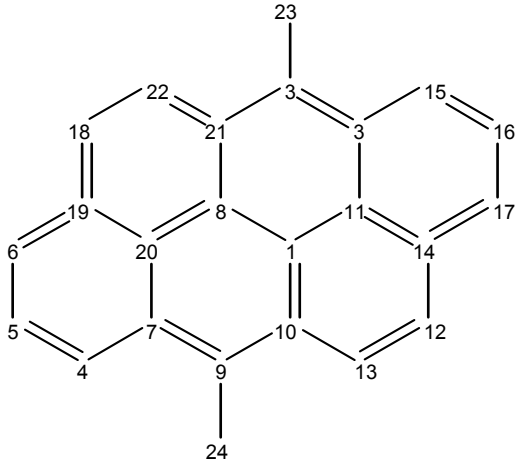


Figure 5. Labeling of vertices in the hydrogen-suppressed graph of 6,12-dimethylanthanthrene.

We have made a complete analysis of one, two, …, five–variables multivariate regression equations based on Wiener, Harary, Balaban and Schulz indices computed from the distance

matrices and the best linear fitting equation is:

$$CA = 1.3372W + 39.701J + 0.036MTI - 0.034W(\Delta) - 5.193H(\Delta^*) - 162.32674 \quad (27)$$

	N	*r*	*s*	F
Training set	30	0.8097	1.2109	23
Test set	16	0.8232	1.0749	15

In Table 6 we present the predicted results, the experimental data and the Galvão predictions.

Table 6. Several Results for Carcinogenic Activity

Chemical name	CA^a	$CA_{calc}{}^b$	CA^c
6,12–Dimethylbenz[*a*]anthracene	4.0	2.5	A
6,8,12–Trimethylbenz[*a*]anthracene	4.0	2.5	A
2–Methylbenzo[*a*]pyrene	4.0	2.8	A
11–Methylbenzo[*a*]pyrene	4.0	3.2	A
12–Methylbenzo[*a*]pyrene	4.0	2.9	A
1–Methylbenzo[*a*]pyrene	4.0	2.8	A
5–Methylbenzo[*c*]phenantrene	3.0	2.0	A
5–Methylchrysene	3.0	1.5	A
6,8–Dimethylbenz[*a*]anthracene	3.0	1.0	A
7–Methylbenz[*a*]anthracene	3.0	0.6	A
12–Methylbenz[*a*]anthracene	2.0	1.1	I
6–Methylanthanthrene	2.0	2.6	I
10–Methylbenzo[*a*]pyrene	1.0	3.0	I
3–Methylbenz[*a*]anthracene	–1.0	–1.2	I
9–Methylbenz[*a*]anthracene	–1.0	–1.2	I
2–Methylbenz[*a*]anthracene	–1.0	–0.9	I
6,10–Dimethylbenzo[*a*]pyrene		3.8	NA
8–Methylbenzo[*a*]pyrene		2.9	NA
9–Methylbenzo[*a*]pyrene		2.9	NA

[a] Carcinogenic activity scale proposed in this paper
[b] Carcinogenic activity calculated with equation (27)
[c] Carcinogenic activity adapted from Cavaliere [87]

The second half of our analysis was performed with the aid of the flexible descriptors defined before. We calculated three sets of OCWLGI for each molecular set (*i.e.* training and test sets) and made a complete computation of regression equations for one, and two variables fitting formulae at fist, second and third orders, as well as some fractional powers (*i.e.* 1/2 and 1/3). Some results are presented in Tables 7–10.

Table 7. Linear Regression $CA = a\mathrm{D} + b$

		Training set						Test set			
Probe	D	*a*	*b*	*n*	*r*	*s*	F	*n*	*r*	*s*	F
1	*pt*21	0.794	–3.487	30	0.9439	0.682	229	16	0.9618	0.584	173
2	*pt*21	0.813	–4.919	30	0.9425	0.689	223	16	0.9616	0.594	172
3	*pt*21	0.718	–3.863	30	0.9435	0.684	227	16	0.9616	0.587	172

Table 8. Quadratic regression $CA = a + b\mathrm{D} + c\mathrm{D}^2$

		Training set							Test set			
Probe	D	*n*	*a*	*b*	*c*	*r*	*s*	F	*n*	*r*	*s*	F
1	*pt*21	30	–4.5972	1.0868	–0.0227	0.9455	0.6963	236	16	0.9623	0.5597	188
2	*pt*21	30	–6.8333	1.3291	–0.0319	0.9454	0.6969	236	16	0.9633	0.5673	193
3	*pt*21	30	–4.8945	1.0240	–0.0199	0.9454	0.6969	236	16	0.9628	0.5629	190

In order to make a direct comparison with other theoretical results, we have adapted the carcinogenic activity data from the Cavaliere scale [87], in a similar fashion as done by Galvão [5].

Table 9. Cubic Regression $CA = a + b\mathrm{D} + c\mathrm{D}^2 + d\mathrm{D}^3$

		Training set								Test set			
Probe	D	*n*	*a*	*b*	*c*	*d*	*r*	s	F	*n*	*r*	*s*	F
1	*pt*21	30	0.8571	–1.8504	0.4741	–0.0257	0.9546	0.65	287	16	0.9748	0.44	264
2	*pt*21	30	5.0514	–3.5746	0.6033	–0.0261	0.9533	0.66	279	16	0.9742	0.43	279
3	*pt*21	30	1.7229	–2.0381	0.4094	–0.0187	0.9541	0.65	284	16	0.9732	0.44	269

Table 10. Multilinear Regression *CA* = *a VD2*.probe1 + *b pt21*.probex + *c*, Where *VD2*.probe1 Denotes Probe 1 for *VD2* and *pt21*.probrex Denotes Probex for *pt21*

	Training set							Test set			
probex	*n*	*a*	*b*	*c*	*r*	*s*	F	*n*	*r*	*s*	F
1	30	0.3841	0.4834	–0.1952	0.9472	0.6618	118	16	0.9671	0.48	101
2	30	0.4210	0.4637	–0.6939	0.9465	0.6657	116	16	0.9682	0.47	105
3	30	0.3910	0.4317	–0.3608	0.9469	0.6632	117	16	0.9674	0.48	102

Table 11. Theoretical and Experimental Carcinogenic Activity *CA* Results

Chemical name	CA^a	CA^b	CA^c	CA^d	CA^e
6,12–Dimethylbenz[*a*]anthracene	A	4.0	3.6	3.8	6.8
6,8,12–Trimethylbenz[*a*]anthracene	A	4.0	5.6	3.3	4.7
2–Methylbenzo[*a*]pyrene	A	4.0	4.0	4.0	4.0
11–Methylbenzo[*a*]pyrene	A	4.0	4.3	4.3	3.5
12–Methylbenzo[*a*]pyrene	A	4.0	3.5	3.8	3.5
1–Methylbenzo[*a*]pyrene	A	4.0	3.6	3.8	3.8
5–Methylbenzo[*c*]phenantrene	A	3.0	2.1	2.3	3.2
5–Methylchrysene	A	3.0	3.6	3.8	3.1
6,8–Dimethylbenz[*a*]anthracene	A	3.0	4.6	4.0	3.0
7–Methylbenz[*a*]anthracene	A	3.0	2.3	2.6	2.2
12–Methylbenz[*a*]anthracene	I	2.0	1.2	1.2	1.4
6–Methylanthanthrene	I	2.0	1.5	1.6	1.4
10–Methylbenzo[*a*]pyrene	I	1.0	1.6	1.7	1.8
3–Methylbenzo[*a*]anthracene	I	–1.0	–0.9	–1.0	–0.9
9–Methylbenzo[*a*]anthracene	I	–1.0	–0.8	–1.0	–0.8
2–Methylbenzo[*a*]anthracene	I	–1.0	–2.1	–1.0	–1.2
6,10–Dimethylbenzo[*a*]pyrene	NA	NA	2.4	2.7	3.0
8–Methylbenzo[*a*]pyrene	NA	NA	4.1	4.0	3.7
9–Methylbenzo[*a*]pyrene	NA	NA	2.3	2.6	2.3

[a] Carcinogenic activity adapted from Cavaliere [87]
[b] Carcinogenic activity scale proposed in this study
[c] Carcinogenic activity computed with the linear equation from Table 7 for probe 1
[d] Carcinogenic activity computed with the cubic equation from Table 9 for probe 1
[d] Carcinogenic activity computed with the multilinear equation from Table 10 for probe 1

4 DISCUSSION

The first point deserving a comment is the quite satisfactory agreement on the capabilities of both sets of topological descriptors. In fact, in general we can state that distance matrices and *OCWLGI* based descriptors yield very good regression equations and encouraging predictions. However, flexible descriptors are better than rigid ones. As a matter of fact, quite superior results are obtained from two–variables lineal relationships based on flexible descriptors than five–

variables linear regression equations based upon rigid topological descriptors.

Furthermore, the best Galvão results [5–8] obtained with the neural networks technique are quite similar to our predictions derived from the multilinear regression analysis and third–order equations. However, in order to judge in an appropriate manner these results, one must consider that our technique is a relatively simple and direct procedure to obtain final results, while previous methods are rather elaborate and not so direct. For the purpose to highlight the relative merits of the present approach, let us to discuss briefly 10 problematic molecules, among which we include the last three in Table 3 lacking experimental data on their carcinogenic activity. The main results are given in Table 12.

Table 12. Comparative Analysis Regarding the Results Reported by Galvão [5–8]

Chemical name	CA^a	CA^b	CA^c	CA^d	CA^e
4,5–dimethylbenzo[*a*]pyrene	A	A	I	3.9	4.1
5–methylbenzo[*c*]phenanthrene	A	A	I	2.3	3.2
5–methylchrysene	A	A	I	3.8	3.1
6,8–dimethylbenz[*a*]antracene	A	A	I	4.0	3.0
6–methylanthanthrene	I	I	A	1.6	1.4
6,12–dimethylanthanthrene	I	A	I	2.2	2.0
7,10–dimethylbenzo[*a*]pyrene	I	A	I	–0.3	0.01
6,10–dimethylbenzo[*a*]pyrene	NA	A	A	2.7	3.0
8–methylbenzo[*a*]pyrene	NA	A	I	4.0	3.7
9–methylbenzo[*a*]pyrene	NA	A	A	2.6	2.3

[a] Carcinogenic activity adapted from Cavaliere [87]
[b] ANN results from Galvão
[c] ANN results from Galvão
[d] Carcinogenic activity computed with the cubic equation from Table 9 for probe 1
[d] Carcinogenic activity computed with the multilinear equation from Table 10 for probe 1

Our predictions based on *OCWLGI* agree with experimental data and for those molecules without available experimental activities, our predictions are positive. These last predictions are in line with the Galvão overall analysis on their corresponding carcinogenic properties. However, Galvão methods based on neural networks and principal component analysis present some disagreements between theoretical predictions and experimental information.

5 CONCLUSIONS

The application of a quantitative structure–activity model based on topological descriptors derived from distance matrices and optimized correlation weights of local graph invariants to study the carcinogenic activity of PAHs allows us to obtain very good results when comparing available experimental data and theoretical predictions. The procedure is relatively straightforward and the quality of results is similar to those derived from more elaborate methods.

The suitable normalization of our numerical results leads to the definition of precise intervals of carcinogenic activities and it opens the possibility to extend the application to other molecular sets

presenting this sort of activity. This proposal is based on the quite satisfactory predictions for a particularly troublesome set of 10 PAHs and the agreement with previous theoretical results obtained from other methodologies grounded on quite different basis, such as electronic density of states over specific molecular regions.

Acknowledgment

We thank the valuable comments and suggestions made by the referees who have greatly contributed to the improvement of this paper.

5 REFERENCES

[1] C. Huggins, Cancer and Necrosis Induced Selectively by Hydrocarbons; in: *Horizons in Biochemistry*, Eds. Albert Szent–Györgi Dedicatory Volume, M. Kasha and B. Pullman, Academic Press, New York, 1962, pp. 497–511.

[2] A. Pullman and B. Pullman, Cancerisation par les Substances Chimiques et Structure Moleculaire, Masson, Paris, 1955.

[3] R. G. Harvey, New Synthetic Approaches to Polycyclic Aromatic Hydrocarbons and Their Carcinogenic Oxidized Metabolites: Derivatives of Benzo[*s*]picene, Benzo[*rst*]pentaphene, and Dibenzo[*b*,*def*] chrysene, *J. Org. Chem.* **2000**, *65*, 3952–3960.

[4] R. G. Harvey, Polycyclic Aromatic Hydrocarbons: Chemistry and Carcinogenesis, Cambridge University Press, Cabmridge, England, 1991.

[5] R. Vendrame, R. S. Braga, Y. Takahata, and D. S. Galvão, Structure–Carcinogenic Activity Relationship Studies of Polycyclic Aromatic Hydrocarbons (PAHs) with Pattern–Recognition Methods, *J. Mol. Struct.* (*Theochem*) **2001**, *539*, 253–265.

[6] R. S. Braga, P. M. V. B. Barone, and D. S. Galvão, Identifying Carcinogeic Activity of Methylated Polycyclic Aromatic Hydrocarbons (PAHs), *J. Mol. Struct.* (*Theochem*) **1999**, *464*, 257–266.

[7] R. Vendrame, R. S. Braga, Y. Takahata, and D. S. Galvão, Structure–Activity Relationship Studies of Carcinogenic Activity of Polycyclic Aromatic Hydrocarbons Using Calculated Molecular Descriptors with Principal Component Analysis and Neural Network Methods, *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1094–1104.

[8] P. M. V. B. Barone, A. Camilo, Jr., and D. S. Galvão, Theoretical Approach to Identify Carcinogenic Activity of Polycyclic Aromatic Hydrocarbons, *Phys. Rev. Lett.* **1996**, *77*, 1186–1189.

[9] A. Mercader, E. A. Castro, and A. A. Toropov, Maximum Topological Distances Based Indices as Molecular Descriptors for QSPR. 4– Modeling the Enthalpy of Formation of Hydrocarbons from Elements, *Int. J. Mol. Sci.* **2001**, *2*, 121–134.

[10] A. Mercader, E. A. Castro, and A. A. Toropov, Calculation of Total Molecular Electronic Energies from Correlation Weighting of Local Graph Invariants, *J. Mol. Model.* **2001**, *7*, 1–5.

[11] G. Krenkel, E. A. Castro, and A. A. Toropov, Improved Molecular Descriptors to Calculate Boiling Points Based on the Optimization of Correlation Weights of Local Graph Invariants, *J. Mol. Struct.* (*Theochem*) **2001**, *542*, 107–113.

[12] G. Krenkel, E. A. Castro, and A. A. Toropov, Improved Molecular Descriptors Based on the Optimization of Correlation Weights of Local Graph Invariants, *Int. J. Mol. Sci.* **2001**, *2*, 57–65.

[13] G. Krenkel, E. A. Castro, and A. A. Toropov, QSPR Modeling of the Enthalpy of Formation from Elements by Means of Correlation Weighting of Local Invariants of Atomic Molecular Graphs, *Chem. Phys. Lett.* **2000**, *330*, 612–623.

[14] E. A. Castro, M. Tueros, and A. A. Toropov, Maximum Topological Distances Based Indices as Molecular Descriptors for QSPR. 2– Application to Aromatic Hydrocarbons, *Comput. Chem.* **2000**, *25*, 571–576.

[15] M. Firpo, L. Gavernet, E. A. Castro, and A. A. Toropov, Maximum Topological Distances Based Indices as Molecular Descriptors for QSPR. 1– Application to Alkyl Benzenes Boiling Points, *J. Mol. Struct.* (*Theochem*) **2000**, *501–502*, 419–425.

[16] P. J. Peruzzo, D. J. G. Marino, E. A. Castro, and A. A. Toropov, Calculation of pK Values of Flavylium Salts from the Optimization of Correlation Weights of Local Graph Invariants, *J. Mol. Struct.* (*Theochem*) **2001**, *572*, 53–60.

[17] M. Tueros, E. A. Castro, and A. A. Toropov, Maximum Topological Distance Based Indices as Molecular Descriptors for QSPR. 3– Calculation of Hydrophobicity of Polyaromatic Hydrocarbons, *J. Mol. Model.* **2001**, *7*, 178–183.

[18] K. Balasubramanian and S. C. Basak, Characterization of Isospectral Graphs Using Graph Invariants and Derived Orthogonal Parameters, *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 367–373.
[19] A. T. Balaban, Highly Discriminating Distance–Based Topological Index, *Chem. Phys. Lett.* **1982**, *89*, 399–404.
[20] M. Randić and C. L. Wilkins, Graph Theoretical Approach to Recognition of Structural Similarity in Molecules, *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 31–37.
[21] S. C. Basak, G. D. Grunwald, and G. J. Niemi, Use of Graph Theoretic and Geometrical Molecular Descriptors in Structure–Activity Relationships; in: *From Chemical Topology to Three Dimensional Geometry*, Ed. A. T. Balaban, Plenum Press, New York, 1997, pp. 73–116.
[22] S. C. Basak, G. J. Niemi, and G. D. Veith, Optimal Characterization of Structure for Prediction of Properties, *J. Math. Chem.* **1990**, *4*, 185–205.
[23] S. C. Basak, V. R. Magnuson, G. J. Niemi, and R. R. Regal, Determining Structural Similarity of Chemicals Using Graph–Theoretic Indices, *Discrete Appl. Math.* **1988**, *19*, 17–44.
[24] S. C. Basak, S. Bertelsen, and G. D. Grunwald, Application of Graph Theoretical Parameters in Quantifying Molecular Similarity and Structure–Activity Studies, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 270–276.
[25] S. C. Basak, Information Theoretic Indices of Neighborhood Complexity and Their Applications; in: *Topological Indices and Related Descriptors in QSAR and QSPR*, Eds. J. Devillers and A. T. Balaban, Gordon and Breach Science Publishers, Amsterdam, 1999, pp 563-593.
[26] S. C. Basak, G. J. Niemi, and G. D. Veith, Recent Developments in the Characterization of Chemical Structure using Graphs Theoretic Indices; in: *Computational Chemical Graph Theory and Combinatories*, Ed. D. H. Rouvray, Nova, New York, 1990, pp. 235–277.
[27] S. C. Basak, Use of Molecular Complexity Indices in Predictive Pharmacology and Toxicology: A QSAR Approach, *Med. Sci. Res.* **1987**, *15*, 605–609.
[28] S. C. Basak, In Practical Applications of Quantitative Structure–Activity Relationships (QSAR) in Environmental Chemistry and Toxicology, Eds. W. Karcher and J. Devillers, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1990, pp.83–103.
[29] S. C. Basak and G. D. Grunwald, Molecular Similarity and Risk Assessment: Analog Selection and Property Estimation Using Graph Invariants, *SAR QSAR Environ. Res.* **1994**, *2*, 289–307.
[30] S. C. Basak and G. D. Grunwald, G. E. Host, G. J. Niemi and S. P. Bradbury, A Comparative Study of Molecular Similarity, Statistical, and Neural Methods for Predicting Toxic Modes of Action, *Environ. Toxicol. Chem.* **1998**, *17*, 1056–1064.
[31] S. C. Basak, S. Bertelsen, and G. D. Grunwald, Use of Graph Theoretic Parameters in Risk Assessment of Chemicals, *Toxicol. Lett.***1995**, *79*, 239–250.
[32] C. M. Auer, J. V. Nabhloz, and K. P. Baetcke, Mode of Action and the assessment of Chemical Hazards in the Presence of Limited Data: Use of Structure–Activity Relationships (SAR) under TSCA, Section 5, *Environ. Health Perspect.* **1990**, *87*, 183–197.
[33] National Research Council, *Toxicity Testing: Strategies to Determine Needs and Priorities*, National Academy Press, Washington, D. C. p. 382.
[34] J. C. Arcos, Structure–Activity Relationships: Criteria for Predicting the Carcinogenic Activity of Chemical Compounds, *Environ. Sci. Technol.* **1987**, *21*, 743–745.
[35] O. Ivanciuc and A. T. Balaban, Design of Topological Indices. Part 8. Path Matrices and Derived Molecular Graph Invariants, *MATCH* (*Commun. Math. Chem.*) **1994**, *30*, 141–152.
[36] O. Ivanciuc and T. Ivanciuc, Matrices and Structural Descriptors Computed from Molecular Graph Distances; in: *Topological Indices and Related Descriptors in QSAR and QSPR*, Eds. J. Devillers and A. T. Balaban, Gordon and Breach Science Publishers, Amsterdam, **1999**, pp 221–277.
[37] G. Rücker and C. Rücker, Symmetry–Aided Computation of the Detour Matrix in Chemistry, *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 710–714.
[38] Hosoya, A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons, *Bull. Chem. Soc. Jpn.* 1971, *44*, 2332–2339.
[39] D. H. Rouvray, The Origins of Chemical Graph Theory; in: *Chemical Graph Theory*, Eds. D. Bonchev and D. H. Rouvray, Abacus Press/Gordon Breach, New York, 1991.
[40] D. J. Klein and M. Randić, Resistance Distance, *J. Math. Chem.* **1993**, *12*, 81–95.
[41] M. V. Diudea and I. Gutman, Wiener–Type Topological Indices, *Croat. Chem. Acta* **1998**, *71*, 21–51.
[42] M. V. Diudea, Valencies of Property, *Croat. Chem. Acta* **1999**, *72*, 835–851.
[43] O. Ivanciuc, T. Ivanciuc, and M. V. Diudea, Molecular Graph Matrices and Derived Structural Descriptors, *SAR QSAR Environ. Res.* **1997**, *7*, 63–87.
[44] M. V. Diudea (Ed.), *QSPR/QSAR Studies by Molecular Descriptors*, Nova Science, Huntington, N.Y., 2001.
[45] A. T. Balaban (Ed.), *From Chemical Topology to Three-Dimensional Geometry*, Plenum, New York, 1997.
[46] O. Ivanciuc and A. T. Balaban, Graph Theory in Chemistry; in: *The Encyclopedia of Computational Chemistry*, Eds. P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, and P. R.

Schreiner, John Wiley & Sons, Chichester, 1998, pp. 1169–1190.
[47] A. T. Balaban and M. V. Diudea, Real Number Vertex Invariants: Regressive Distance Sums and Related Topological Indices, *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 421–428.
[48] O. Ivanciuc and A. T. Balaban, The Graph Description of Chemical Structures; in: *Topological Indices and Related Descriptors in QSAR and QSPR*, Eds. J. Devillers and A. T. Balaban, Gordon and Breach Science Publishers, Amsterdam, 1999, pp 59–167.
[49] O. Ivanciuc, T. Ivanciuc, and A. T. Balaban, Vertex– and Edge–Weighted Molecular Graphs and Derived Structural Descriptors; in: *Topological Indices and Related Descriptors in QSAR and QSPR*, Eds. J. Devillers and A. T. Balaban, Gordon and Breach Science Publishers, Amsterdam, 1999, pp 169–220.
[50] O. Ivanciuc, T. Ivanciuc, D. Cabrol–Bass, and A. T. Balaban, Comparison of Weighting Schemes for Molecular Graph Descriptors: Application in Quantitative Structure–Retention Relationship Models for Alkylphenols in Gas–Liquid Chromatography, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 732–743.
[51] A. T. Balaban, Chemical Graphs. Part 48. Topological Index J for Heteroatom–Containing Molecules Taking Into Account Periodicities of Element Properties, *MATCH* (*Commun. Math. Chem.*) **1986**, *21*, 115–122.
[52] O. Ivanciuc, T. Ivanciuc, and A. T. Balaban, The Complementary Distance Matrix, a New Molecular Graph Metric, *ACH - Models Chem.* **2000**, *137*, 57–82.
[53] O. Ivanciuc, T. Ivanciuc, D. Cabrol–Bass, and A. T. Balaban, Evaluation in Quantitative Structure–Property Relationship Models of Structural Descriptors Derived from Information–Theory Operators, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 631–643.
[54] A. T. Balaban, D. Mills, O. Ivanciuc, and S. C. Basak, Reverse Wiener Indices, *Croat. Chem. Acta* **2000**, *73*, 923–941.
[55] O. Ivanciuc, T. Ivanciuc, D. J. Klein, W. A. Seitz, and A. T. Balaban, Wiener Index Extension by Counting Even/Odd Graph Distances, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 536–549.
[56] O. Ivanciuc, T. Ivanciuc, D. J. Klein, W. A. Seitz, and A. T. Balaban, Quantitative Structure–Retention Relationships for Gas Chromatographic Retention Indices of Alkylbenzenes with Molecular Graph Descriptors, *SAR QSAR Environ. Res.* **2001**, *11*, 419–452.
[57] A. T. Balaban and T. S. Balaban, New Vertex Invariants and Topological Indices of Chemical Graphs Based on Information on Distances, *J. Math. Chem.* **1991**, *8*, 383–397.
[58] T. S. Balaban, P. A. Filip, and O. Ivanciuc. Computer Generation of Acyclic Graphs Based on Local Vertex Invariants and Topological Indices. Derived Canonical Labelling and Coding of Trees and Alkanes, *J. Math. Chem.***1992**, *11*, 79–105.
[59] O. Ivanciuc, T. S. Balaban, and A. T. Balaban, Design of Topological Indices. Part. 4. Reciprocal Distance Matrix, Related Local Vertex Invariants and Topological Indices, *J. Math. Chem.***1993**, *12*, 309–318.
[60] H. Wiener, Structural Determination of Paraffin Boiling Points, *J. Am. Chem. Soc.* **1947**, *69*, 17–20.
[61] D. Plavšić, S. Nikolić, N. Trinajstić, and Z. Mihalić, On the Harary Index for the Characterization of Chemical Graphs, *J. Math. Chem.* **1993**, *12*, 235–250.
[62] M. Randić, On Characterization of Molecular Branching, *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.
[63] A. T. Balaban, Topological Index based on Topological Distances in Molecular Graphs, *Pure Appl. Chem.* **1983**, *55*, 199–206.
[64] O. Mekenyan and S. C. Basak, *Topological Indices and Chemical Reactivity* in Graph Theoretical Approaches to Chemical Reactivity, Eds. D. Bonchev and O. Mekenyan, Kluwer/Academic Publishers, Dordrecht, 1994, pp.221–239.
[65] H. P. Schultz, Topological Organic Chemistry, *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 227–228.
[66] W. R. Mueller, K. Szymanski, J. V. Knop, and N. Trinajstić, An Algorithm for Construction of the Molecular Distance Matrix, *J. Chem. Inf. Comput. Sci.* **1987**, *8*, 170–173.
[67] M. Randić, Novel Graph Theoretical Approach to Heteroatoms in Quantitative Structure–Activity Relationships, *Chemom. Intell. Lab. Syst.* **1991**, *10*, 213–227.
[68] M. Randić, On Computation of Optimal Parameters for Multivariate Analysis of Structure–Property Relationship, *J. Comput. Chem.* **1991**, *12*, 970–980.
[69] M. Randić and S. C. Basak, Variable Molecular Descriptors; in: *Some Aspects of Mathematical Chemistry*, Eds. D. K. Sinha, S. C. Basak, R. K. Mohany and I. N. Busamallick, Visva–Bharati University Press, Santiniketan, India, 2000.
[70] M. Randić, P. J. Hansen, and P. C. Jurs, Search for Useful Graph Theoretical Invariants of Molecular Structure, *J. Chem. Inf. Comput. Sci.* **1998**, *28*, 60–68.
[71] O. Araujo and J. Rada, Randić Index and Lexicographic Order, *J. Math. Chem.* **2000**, *27*, 201–212.
[72] M. Randić, Chemical Structure… What is "She"? *J. Chem. Educ.* **1992**, *69*, 713–718.
[73] G. P. Romanelli, L. F. R. Cafferata, and E. A. Castro, Ameliorate QSPR Study of Alkyl–hydroperoxides, *Russ. J. Gen. Chem.* **2001**, *71*, 257–260.
[74] P. Duchowicz and E. A. Castro, Improved QSPR Prediction of Heats of Formation of Alkenes, *J. Korean Chem.*

Soc. **2000**, *44*, 501–506.
[75] P. Duchowicz and E. A. Castro, Calculation of Molecular Polarizability and Water Solubility of Alkanes and Alcohols, *J. Indian Chem. Soc*. **2001**, *78*, 192–196.
[76] G. P. Romanelli, L. F. R. Cafferata, and E. A. Castro, An Improved QSAR Study of Toxicity of Saturated Alcohols, *J. Mol. Struct.* (*Theochem*) **2000**, *504*, 261–265.
[77] P. Duchowicz and E. A. Castro, A Rather Simple Method to Calculate log P Values in QSAR/QSPR Studies, *Acta Chim. Slov*. **2000**, *47*, 281–292.
[78] G. P. Romanelli, L. F. R. Cafferata, and E. A. Castro, Application of Improved Quantitative Structure–Retention Relationships to Study the Molecular Mechanism of Stationary Phases in HPLC, *Khim. Fizika* **2000**, *19*, 105–109.
[79] M. Firpo, L. Gavernet, and E. A. Castro, Empirical Calculation of Vaporization Enthalpies at 298.15 K of the n–Alkanes from C_{21} to C_{28} and C_{30}, *Rom. J. Phys*. **1999**, *44*, 181–187.
[80] L. Gavernet, M. Firpo, and E. A. Castro, Heats of Formation of Alkenes Calculated from Simple Protocols, *Rev. Roum. Chim*. **1998**, *43*, 1079–1082.
[81] P. Duchowicz and E. A. Castro, Hydrocarbon Enthalpies of Formation from ab initio Calculations Improved through Bond Parameters, *J. Korean Chem. Soc*. **1999**, *43*, 621–627.
[82] M. Firpo, L. Gavernet, and E. A. Castro, An Improved Calculation of Bond–dissociation Energies of Organic Compounds based on the Concept of Stabilization Energies, *Polish J. Chem. Soc*. **1999**, *73*, 1041–1045.
[83] E. A. Castro, Improved Simple Protocols to Calculate Hydrocarbon Enthalpies of Formation from ab initio Total Energies, *Comput. Chem*. **1997**, *21*, 305–308.
[84] A. A. Toropov and A. P. Toropova, Modeling of Lipophilicity by Means of Correlation Weighting of Local Graph Invariants, *J. Mol. Struct.* (*Theochem*) **2001**, 538, 197–199.
[85] A. A. Toropov, A. P. Toropova, T. Ismailov, and D. Bonchev, 3D Weighting of Molecular Descriptors for QSPR/QSAR by the Method of Ideal Symmetry (MIS). 1. Application to Boiling Points of Alkanes, *J. Mol. Struct.* (*Theochem*) **1991**, *424*, 237–247.
[86] D. Bonchev, Overall Connectivity and Molecular Complexity, in *Topological Indices and Related Descriptors*, Eds. J. Devillers and A. T. Balaban, Gordon and Breach Science Publishers, Amsterdam, 1999.
[87] E. L. Cavaliere, E. G. Rogan, R. W. Roth, R. K. Saugier, and A. Hakan, *Chem.–Biol. Interact*. **1983**, *47*, 87–96.
[88] A. A. Toropov, A. P. Toropova, M. M. Ishankhodzhaeva, and N. A. Parpiev, QSPR Modeling of Stability Constants of Coordination Compounds by Optimization of Correlation Weights of Local Graph Invariants, *Russ. J. Inorg. Chem*. **2000**, *45*, 1057–1059.

Biographies

D. **J**. **G**. **Marino** is technical assistant at the Research Center of Environment Medium, La Plata University, and is devoted to study carcinogenic hydrocarbons via theoretical and experimental methods.

P. **J**. **Peruzzo** is technical assistant at the Research Center of Environment Medium, La Plata University and his main research interest is related to numerical methods to compute molecular descriptors in QSAR/QSPR field.

E. **A**. **Castro** is full professor at La Plata University and Superior Researcher in the National Research Council. He has published around 700 papers and is author of several books and chapters of books on Physical Chemistry, Science Education, Mathematical Chemistry and Science Management issues.

A. **A**. **Toropov** is employed at the Vostok Innovation Company, Tashkent, Uzbekistan, and has published several papers on QSAR/QSPR theory. His main scientific interest is related to flexible topological molecular descriptors.