What Are the Limits of Applicability for Graph Theoretic Descriptors in QSPR/QSAR? Modeling Dipole Moments of Aromatic Compounds with TOPS-MODE Descriptors

Ernesto Estrada* and Humberto González

Faculty of Pharmacy, Department of Organic Chemistry, University of Santiago de Compostela,
Santiago de Compostela 15706, Spain, and Center of Chemical Bioactive, Central University of Las Villas,
Santa Clara 54830, Villa Clara, Cuba

The numerous possibilities of using graph theoretic descriptors in QSPR/QSAR are analyzed, and some misunderstandings on the role of this theoretical approach in chemistry are clarified. Principal component analysis is used to obtain a property space for several physicochemical properties of aromatic compounds. It is proved that most of the QSPR applications of the graph-theoretic structure descriptors are concentrated to the description of properties in a very limited region of this property space. Here, we show that graph-theoretic approaches are also applicable to the modeling of physicochemical properties that are far away from this region traditionally studied. The molecular dipole moments of benzene derivatives, *mono-*, *ortho-*, *meta-*, and *para-*susbtituted, are modeled by using the Topological Sub-Structural Molecular Design (TOPS-MODE) approach. The TOPS-MODE approach used permits to calculate group dipole moments that are given for several substituents. The differences between these group dipoles and those obtained by simple difference between experimental values are analyzed. Some difficulties arising from this traditional way of deriving substituent constants are identified and analyzed.

INTRODUCTION

What are the limits of application of Graph Theory and Topology in chemistry? Those who identified the chemical graph theory with the Hückel molecular orbital (HMO) approach believed that it represents the only field of application for topological ideas (in a graph-theoretical way) in theoretical chemistry. However, the development of graph theoretical descriptors, also known as topological indices,[1] have permitted the efficient description of many physical, physicochemical and biological properties of organic and organo-metallic compounds.[2−12] The use of this kind of molecular descriptors represents an important structure-explicit alternative in quantitative structure−property (QSPR) and quantitative structure−activity (QSAR) studies. Despite the simplicity of most of these descriptors, they have proved to contain a great quantity of structural information, which is mainly, but not exclusively, related to the connectivity relationships between atoms in the molecules. Consequently, they can be considered as appropriate numerical descriptors of the "chemical structure".[13−17]

Most of the criticism directed against graph theoretic approaches in chemistry is based on the fact that they do not contain "explicit" information about three-dimensional features of molecules. Based on this conclusion, the opinion prevailed that the application of topological descriptors is limited to series of structurally related compounds, for which the three-dimensional features are constant, and the role of topological features are dominant.[18] In favor of this thesis is the fact that a great part of the applications of topological indices in QSPR and QSAR were reported for this kind of data sets. On the other hand, most of the physicochemical properties studied by graph theoretic approaches are mainly size- and shape-dependent, such as boiling points, molar volumes, retention indices in chromatography, and so forth. When the indices are applied to QSAR studies they are believed to be useful only in cases in which the biological activity is nonspecific, or when limited series of structurally analogous compounds are used. However, recent advances in the development and applications of graph theoretic descriptors in physical chemistry and pharmaceutical sciences, including the development of three-dimensional descriptors, indicate that this approach is an important alternative to the development of QSPR/ QSAR models, maintaining both statistical and physical meaning.[19−37]

ON THE ROLE OF GRAPHS DESCRIPTORS IN CHEMISTRY

There is a misunderstanding of the role of graph theoretic descriptors in chemistry. The differences between graph theory and topology have been commented on by Randić several times.[38] However, the identification of graph theoretic descriptors as topological ones needs some conceptual corrections. There is no doubt that there are a number of "pure" topological indices, i.e., those that have information only about the relationships of connectivity between atoms or bonds in the molecules. Such are, for instance, the Randić branching index,[39] Hosoya Z index,[40] Balaban J index,[41] the Wiener index,[42] bond connectivity index,[43] and so forth. If we limit the applications of graph theory in QSPR/QSAR to the development and uses of this kind of descriptors, we also limit the field of applications of these descriptors, attributing them to homologous series of compounds or to the study of bulky molecular properties. However, the extensions of graph theoretic descriptors to molecules containing heteroatoms carried out by the use of several weights in graphs can hardly be considered as "pure" topological

* Corresponding author e-mail: estrada66@yahoo.com. Present address: SEAC, Unilever R&D, Colworth House, Sharnbrook, Bedford MK44 1LQ, UK.

descriptors.[44−50] These indices not only contain information on molecular topological features but also contain important information on the nature of atoms and bonds, as well as on the electronic features of molecules as a whole.

In this sense, the identification of graphs in chemistry to some specific method or characteristic of molecular structure is similar to the identification of the derivative with the concept of velocity. Consequently, the question of the limits of applicability of graph theory in chemistry is similar to the question of applicability of differential calculus in physical sciences. These two questions are, of course, lacking of any sense. Both, graphs and derivatives, are mathematical objects having different uses and applications in many fields of the intellectual activity of men.

Graph theoretic approaches in structure−property studies have the advantages that they are structure-explicit,[51] and as a consequence they permit a direct structural interpretation of the models obtained. This advantage should be maintained in graph theoretic approaches using or not heteroatom differentiation, avoiding the use of unnecessarily complicated invariants, and fulfilling the desirable attributes for this kind of descriptors.[52]

PROPERTY SPACE AND QSPR STUDIES

All physical properties of organic molecules, all the known properties, and many others that can be discovered in the future can be used to form a property space. In this hypothetical multidimensional space, the properties are represented by single points, those properties that have some resemblance are near from each other in the space and those which are not interrelated are far away from each other. It is expected that if a molecular descriptor is able to describe one physical property that is in a specific region of the property space, it will be able to describe also other properties, which are in the same region of the space. On the other hand, it is also expected that this descriptor will be not able to describe properties that are far away from the region where the first properties are. The first of these claims should be accepted, however, with caution. We cannot expect that the ability of molecular descriptors to describe physical properties is transitive, and we cannot, of course, move around the property space by using one single descriptor.

Examples of limited property spaces have been published in the literature by using factor analysis to reduce the dimensions of the variable space. Cramer III[53,54] developed a property space for several physical properties of a structurally diverse series of organic molecules. On the other hand, Needham et al.[55] described property spaces for 8 and 7 important physical properties of alkanes. They found that the first two factors were able to explain more than 94% of the variance in the physical properties and most of them, except the melting point, were placed in the same region of the property space. These properties, boiling points, molar volume, molar refraction, heat of vaporization, critical pressure, critical temperature, and surface tension have significant collinealities as expressed by their high correlation coefficients. They are well described by topological indices, such as the connectivity index, in concordance with our thesis before explained. However, the correlation coefficients for the models using the Randić index[39] vary from 0.96 for boiling point to 0.83 to surface tension, showing the nontransitive character of the ability of molecular descriptors to

Table 1. Results of the Factor Analysis for the Physicochemical Properties of Substituted Benzene Compounds: Eigenvalues, Percentage of Variance and Cumulative Percentage of Variance Reproduced.

factor	eigenvalue	% variance	% cumul. var.
1	2.833	47.22	47.22
2	1.354	22.57	69.79
3	0.965	16.09	85.88

Table 2. Principal Component Analysis Loadings for the Physicochemical Properties[a]

	PCA factor loadings		
property	f1	f2	f3
DM	−0.055	**0.959**	0.002
BP	0.644	0.675	−0.084
MP	−0.096	0.024	**−0.992**
MW	**0.922**	0.039	0.077
MV	**0.794**	−0.124	−0.039
MR	**0.918**	−0.0003	0.098

[a] Loading greater than 0.7 are boldfaced.

describe these physical properties. On the other hand, the melting point, which is out this region of the property space, could not be described by any of the descriptors used in the study.

In the present work, we propose the analysis of a property space composed by the following physical properties: dipole moment (DP), boiling point (BP), melting point (MP), molar volume (MV) and molar refraction (MR). We analyze a data set composed by substituted benzene derivatives. To develop the property space we use principal component analysis (PCA) as described in earlier works.[53−56] Our objective is to show the possibilities of graph theoretic approaches to correctly describe physical properties that are outside of the region of the property space in which these descriptors are used regularly.

The property space is built by using these six properties for a series of 61 mono- and di-substituted benzenes with a great structural variability.[57,58] In Table 1 we show the eigenvalues, the percentage of the variance explained by them individually and the cumulative percentage of the variance of the physical properties reproduced by successive addition of eigenvalues. As can be seen in this table there are three main eigenvalues that explain more than 85% of the variance of the studied properties.

The factor loadings of the PCA are shown in Table 2. The property space projected in the three principal components is illustrated in Figure 1. Taking into account that the factor loadings represent the correlation of the variable with the factors, we can assign physical meanings to these factors or coordinates of the property space. The loadings that are shown in Table 2 and Figure 1 were obtained after a Quartimax rotation of the factor space.[59] The first coordinate, accounting for 47% of the variance, is clearly identified with bulky properties of this series of molecules and most of the properties that are dependent on this structural feature load strongly with this coordinate. These properties are the molar volume, molar refraction and molecular weight. They occupy the same region of the property space as seen in Figure 1. Factor 2, the second coordinate, is clearly identified as a dipole moment coordinate and the third factor is almost exclusively a melting point coordinate.
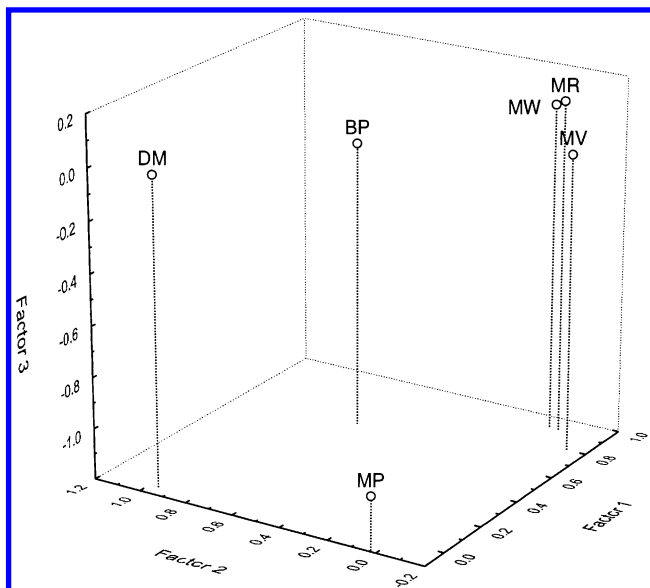
GRAPH THEORETIC DESCRIPTORS IN QSPR/QSAR

J. Chem. Inf. Comput. Sci., Vol. 43, No. 1, 2003 **77**

Figure 1. Three-dimensional representation of the property space obtained for six physicochemical properties of substituted benzene compounds (DM = dipole moment; MP = melting point; BP = boiling point; MW = molecular weight; MR = molar refraction; MV = molar volume).

Boiling point, which is commonly identified as a property mainly dependent from the bulky molecular features, has almost the same contribution to the first two coordinates in the property space, i.e., it is intimately related to the bulky and dipolar features of these molecules.

Now, we analyze the role of graph theoretic approaches in the description of physical properties in the light of this property space. The majority of the "pure" topological indices show good correlation with the properties related to the bulk coordinate in property spaces. For instance, almost all topological indices are tested by correlating them to the boiling points of homologous series of compounds, especially alkanes. However, no one of these graph theoretic descriptors shows good correlation with the dipole moment of series of compounds.

The main purpose of this paper is to show that the graph theoretic approaches can be used to describe not only size and shape-dependent properties, but they can also be applied to describe properties that are out of this region of the property space, such as the dipole moment.

THE TOPS-MODE APPROACH

The general principles of the topological sub-structural approach to molecular design (TOPS-MODE) developed by one of the present authors have been explained elsewhere.[60-69] Here we outline the general scheme of this approach that is based on the calculation of the spectral moments of the bond (edge) adjacency matrix of molecular graphs.[43] This matrix was previously mentioned in a chemical context in some treatises[70,71] and in the work of Heilbronner concerning ionization potentials of saturated hydrocarbons[72] (that is in a quantum chemical context!). For the connection between this matrix and line graphs as well as for a more detailed historical background see Gutman et al.[73] These spectral moments have been expressed in terms of the different structural fragments of organic molecules via algebraic expressions for acyclic and cyclic hydrocarbons as well as

for molecules containing heteroatoms. The method has been successfully applied to QSPR, QSAR and drug design studies.[60-69] This method is related in some extent to the so-called "cluster expansion" approach which has been previously studied by several authors.[74-77] Spectral moments of other structural matrices have also been studied in the chemical literature over a long period of time, in diverse chemical contexts.[78-85]

TOPS-MODE is the acronym of **TOP**ological **S**ubstructural **MO**lecular **De**sign.[60-62] It is based on the computation of the spectral moments of the bond matrix, whose mathematical basis were described in previous reports by one of the current authors.[60-62] We have previously detailed the main steps to be conducted for the application of this approach to structure−property-activity relationships. They can be briefly resumed as follows:

i) draw the hydrogen-depleted molecular graphs for each molecule of the data set,

ii) use appropriated bond weights in order to differentiate the molecular bonds, e.g., bond distance, bond dipoles, bond polarizabilities, etc.,

iii) compute the spectral moments of the bond matrix with the appropriated weights for each molecule in the data set generating a table in which the rows correspond to the compounds and columns correspond to the spectral moments of the bond matrix,

iv) find a quantitative structure−property relationship (QSPR) by using any appropriate linear or nonlinear multivariate statistical technique, such as multilinear regression analysis (MRA), neural networks, etc.

$$P = a_0\mu_0 + a_1\mu_1 + a_2\mu_2 + a_3\mu_3 + a_4\mu_4 + a_5\mu_5 + a_6\mu_6 + b$$

(1)

where P is the measurement of the property, μ_k is the kth spectral moment, and the a_k's are the coefficients obtained by the linear regression analysis,

v) test the predictive capacity of the QSPR model by using cross-validation techniques,

vi) compute the contribution of the different fragments of interest in order to determine their quantitative contribution to the property of the molecules under study.

The computation of fragment contributions to the property under study is probably the most important advance of the TOPS-MODE approach to the study of QSPR/QSAR problems compared to the traditional QSAR and QSPR methods. This procedure can be useful for the identification of possible pharmacophores/toxicophores, that eventually may be further studied by using other theoretical and experimental techniques. It has been described in more details in our previous reports.

CALCULATION OF TOTAL BOND CONTRIBUTIONS

In general there are two different ways in which we can transform a QSAR/QSPR model into an additive model by using the TOPS-MODE approach. The first one is the "classical" way in which we express the QSAR/QSPR model as a linear combination of the contribution of all different structural fragments contributing to the studied property. The second one is to express the total contribution of the different bonds in specific molecules. In this second approach, a more compact additive scheme is obtained. The first approach

permits to calculate the studied property from fragment contributions for any molecule (real or virtual) in a general way. The second approach, however, permits to obtain the contribution of bonds in specific molecules permitting the comparison among them in a more effective way. This means that both approaches are complementary to each other. For comparison an analogy with the use of "standard" and real molecular geometries is appropriated. For instance, if we consider the standard bond distance C−C of 1.54 Å and a standard bond angle of 119.5° we can estimate the distance between any pair of atoms in an alkane molecule, even if it is hypothetical. However, real distances are specific for each molecule and we will have a value of C−C bond distance in ethane, which is different to that of propane or 2-methylpropane. Thus, we cannot use these real distances for estimations in a general way as before but they permit us to compare in a more exact way the geometries of two or more specific molecules. The second approach, based on "local" spectral moments,[86,87] is explained below.

To calculate total bond contributions in the current approach we make use of the decomposition of total spectral moments (TOPS-MODE descriptors) into "local" moments, more specifically into bond spectral moments. The total spectral moments of the bond matrix are defined as[60−62]

$$\mu_k^T = \mu_k = \mathbf{Tr}(\mathbf{B}^k) = \sum_{i=1}^{s}(e_{ii})^k \qquad (2)$$

where Tr means the trace of the matrix, that is the sum of the diagonal entries of the matrix and the elements $(e_{ii})^k$ are the diagonal entries of the kth power of the bond matrix.

A local spectral moment is defined as the sum of diagonal entries of the different powers of the bond matrix corresponding to the respective molecular fragment.[86,87] In mathematical terms local spectral moments of the bond matrix are defined as follows

$$\mu_k(f) = \sum_{i=1}^{f}(e_{ii})^k \qquad (3)$$

where f is the corresponding fragment for which the moments are defined and the summation is carried out over all bonds forming the fragment f. The simplest case is when f corresponds to a single bond, and in this case the kth local moment is defined as the diagonal entry corresponding to this bond in the matrix raised to the kth power.

Consequently, total spectral moments of order k can be expressed as the sum of the bond spectral moments of the same order

$$\mu_k^T = \sum_{B}\mu_k(B) \qquad (4)$$

where B stands for the corresponding bond.

This means that we can substitute expression (4) into the QSAR/QSPR (1) in such a way that we obtain the total contribution of the different bonds in a specific molecule:

$$P = b_0 + \sum_{k} a_k \mu_k = b_0 + \sum_{k}\sum_{B} a_k\mu_k(B) \qquad (5)$$

An example is given here for the calculation of these contributions to molar refraction of an alkane. Consider the following QSPR model to describe molar refraction of alkanes, MR[60]

$$MR(cm^3) = 5.703 + 5.506\mu_0^T - 0.329\mu_2^T + 0.193\mu_3^T - $$
$$0.033\mu_4^T$$
$$N = 69 \quad r = 0.9999 \quad s = 0.05 \quad F = 168569 \qquad (6)$$

where μ_i^T represents the total spectral moments of the bond matrix.

Now we consider the molecule of 2,2-dimethylbutane with the following bond numbering:
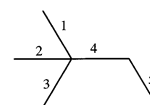
The total spectral moments can be expressed as the sum of the bond spectral moments for this molecule as follows:

$$\mu_0^T = \mu_0(1) + \mu_0(2) + \mu_0(3) + \mu_0(4) + \mu_0(5)$$

$$\mu_2^T = \mu_2(1) + \mu_2(2) + \mu_2(3) + \mu_2(4) + \mu_2(5)$$

$$\mu_3^T = \mu_3(1) + \mu_3(2) + \mu_3(3) + \mu_3(4) + \mu_3(5)$$

$$\mu_4^T = \mu_4(1) + \mu_4(2) + \mu_4(3) + \mu_4(4) + \mu_4(5)$$

The bond spectral moments of this molecule are as follows:

$$\mu_0(1){=}\mu_0(2){=}\mu_0(3){=}\mu_0(4){=}\mu_0(5){=}1$$

$$\mu_2(1){=}\mu_2(2){=}\mu_2(3){=}3 \quad \mu_2(4){=}4 \quad \mu_2(5){=}1$$

$$\mu_3(1){=}\mu_3(2){=}\mu_3(3){=}6 \quad \mu_3(4){=}6 \quad \mu_3(5){=}0$$

$$\mu_4(1){=}\mu_4(2){=}\mu_4(3){=}22 \quad \mu_4(4){=}28 \quad \mu_4(5){=}4$$

The intercept is taken as the contribution of the whole molecule and it is divided by the number of bonds to obtain its contribution to each specific bond.

Now we substitute these expressions into model (6) obtaining bond molar refractions in the following form:

$$MR(1) = MR(2) = MR(3) = (5.703/5) + 5.506\mu_0(1) - $$
$$0.329\mu_2(1) + 0.193\mu_3(1) - 0.033\mu_4(1) = 6.092 \text{ cm}^3$$

$$MR(4) = (5.703/5) + 5.506\mu_0(4) - 0.329\mu_2(4) + $$
$$0.193\mu_3(4) - 0.033\mu_4(4) = 5.5646 \text{ cm}^3$$

$$MR(5) = (5.703/5) + 5.506\mu_0(5) - 0.329\mu_2(5) + $$
$$0.193\mu_3(5) - 0.033\mu_4(5) = 6.1856 \text{ cm}^3$$

It is clear that the sum of these bond contributions gives the value of the molar refraction of the molecule: 30.026 cm³.

MODELING DIPOLE MOMENTS OF AROMATIC COMPOUNDS

The applicability of the current approach to study physicochemical properties of organic compounds is based on

GRAPH THEORETIC DESCRIPTORS IN QSPR/QSAR

J. Chem. Inf. Comput. Sci., Vol. 43, No. 1, 2003 **79**

Table 3. Standard Bond Dipoles Used in the Present Work

bond	dipole[a]	bond	dipole[a]	bond	dipole[a]
C−F	1.51	C≡N	3.602	N→O	3.20
C−Cl	1.56	C−O	0.86	N=O	2.00
C−Br	1.48	C=O	2.40	$C_{sp2}-C_{sp3}$	0.68
C−I	1.29	C−S	2.95	$C_{sp}-C_{sp3}$	1.48
C−N	0.40	C=S	2.80	$C_{sp2}-C_{sp}$	1.15
C=N	0.90	N−O	0.30		

[a] Taken from ref 88.

the flexibility that it offers to use different weights in the diagonal entries of the bond matrix. In the modeling of the dipole moment of aromatic compounds we use the standard bond dipoles as weights in this matrix. This quantity can be obtained from theoretical or experimental approaches. Here we use the values reported by Smyth.[88] In Table 3 we give some of these values for standard dipole moments.

The first data set that will be studied here is composed by 24 monosubstituted benzene derivatives for which the dipole moments have been reported in benzene solution at 25 °C.[58] For the sake of simplicity in the following we will use μ_k instead of μ_k^T.

The best linear regression model obtained to describe the dipole moment of these compounds is given below

$$\mu(D) = 1.681 - 1.110\mu_0 + 0.553 \cdot \mu_1 + 0.445 \cdot \mu_2 - 0.177 \cdot \mu_3 + 0.021 \cdot \mu_4$$

$$n = 24 \quad R = 0.943 \quad s = 0.403 \quad RMSCV = 0.412$$
$$F = 29.1 \quad (7)$$

where $\mu(D)$ is the dipole moment in Debye, n is the number of compounds, R is the regression coefficient, s is the standard deviation of the regression, RMSCV is the standard deviation of the cross-validation, obtained from the leave-one-out procedure, and F is the Fisher ratio. In Table 4 we give the experimental and calculated values of the dipole moment in benzene at 25 °C.

The di-substituted derivatives of benzene represent a more challenging set of compounds for testing the possibilities of the present graph-theoretic approach for the description of molecular dipole moments. In this kind of compounds the vectorial nature of the molecular dipole moments is more remarkable that in the monosubstituted compounds. In this case the molecular dipole moment

depends on the position of the different substituents in the aromatic ring and on the orientation of the dipoles of such groups. This is well illustrated for 1,4-disubstituted benzenes in which both substituents are identical. In this case the resulting dipole moment is zero on grounds of molecular symmetry. These structural features are not explicitly accounted for the present approach that is based on graph-theoretic (topological) ideas. However, as a demonstration of the potentials that these approaches have to the description of molecular structure, we consider the dipole moments of o- (*ortho*), m- (*meta*) and p- (*para*) substituted benzene derivatives as three independent data sets to be studied here.[58]

The best linear regression model describing the dipole moments of *ortho*-substituted aromatic compounds is given below:

$$\mu(D) = 1.055 - 1.106 \cdot \mu_0 - 6.7 \cdot 10^{-4} \cdot \mu_0 \cdot \mu_5 + 6.047 \cdot 10^{-2} \cdot \mu_0 \cdot \mu_2$$

$$n = 24 \quad R = 0.973 \quad s = 0.286 \quad RMSCV = 0.29$$
$$F = 119.7 \quad (8)$$

A similar model was obtained for *meta*-substituted compounds, which is illustrated below:

$$\mu(D) = -3.446 - 1.428 \cdot \mu_1 + 0.189 \cdot \mu_0 \cdot \mu_2 - 0.196 \cdot (\mu_1)^2 - 0.318 \cdot (\mu_0)^2 \, 0.0048 \cdot \mu_0 \cdot \mu_4$$

$$n = 29 \quad R = 0.966 \quad s = 0.319 \quad RMSCV = 0.351$$
$$F = 64.3 \quad (9)$$

In both models describing dipole moments of o- and m-disubstituted benzene derivatives, the spectral moments of the bond matrix are able to explain more than 93% of the variance in the experimental values. Experimental and calculated values of the dipole moments of these molecules are given in Table 5.

As a consequence of the nature of the present approach, which does not consider any geometrical or symmetry-based structural feature, the percentage of the variance in the experimental dipole moments explained for *para*-disubstituted benzene derivatives is smaller than the percentages explained for *ortho*- and *meta*-compounds. Experimental dipole moments of *para*-substituted benzene derivatives and

Table 4. Experimental and Calculated Dipole Moments of Mono-Substituted Benzene Derivatives in Benzene Solution

compound	μ (D) exp.[a]	μ (D) calcd.[b]	residual	compound	μ (D) exp.[a]	μ (D) calcd.[b]	residual
-Br	1.5	1.63	−0.13	-N(CH₃)₂	1.59	0.93	0.66
-F	1.42	1.65	−0.23	-COOCH₃	2.53	2.58	−0.05
-NO₂	3.93	4.10	−0.17	-C₃H₇	0.37	0.83	−0.46
-OH	1.45	1.31	0.14	*i*- C₃H₇	0.38	0.84	−0.46
-CF₃	2.56	2.48	0.08	-OC₂H₅	1.00	0.80	0.20
-CN	4.05	3.58	0.47	-COOC₂H₅	1.85	2.09	−0.24
-CH₃	0.34	1.25	−0.91	-CH₂COOCH₃	1.8	1.99	−0.19
-OCH₃	1.3	1.29	0.01	-COCl	3.16	3.11	0.05
-COCH₃	2.8	2.51	0.29	-CH₂F	1.77	1.76	0.01
-CH₂NH₂	1.38	1.03	0.35	-C_CH	0.79	0.94	−0.15
-NHCH₃	1.65	1.09	0.56	-CH₂CN	3.5	3.79	−0.29
-CHO	2.8	2.61	0.19	-OCH=CH₂	1.1	0.80	0.30

[a] Taken from ref 58. [b] Calculated from eq 7.

Table 5. Experimental and Calculated Dipole Moments of Ortho- and Meta-Substituted Benzene Derivatives.

	ortho substituted benzene derivatives			*meta* substituted benzene derivatives		
compound	μ (D) exp.[a]	μ (D) calcd.[b]	residual	μ (D) exp.[a]	μ (D) calcd.[c]	residual
bromochlorobenzene	2.19	2.03	0.16	1.52	1.43	0.08
bromofluorobenzene	1.84	2.00	−0.16	1.42	1.42	0.00
bromoiodobenzene	1.815	1.88	−0.07	1.15	1.40	−0.25
bromonitrosobenzene	4.143	4.12	0.02	3.32	3.44	−0.12
dibromobenzene	1.55	1.99	−0.44	1.377	1.42	−0.04
chlorofluorobenzene	2.365	2.05	0.32	1.52	1.44	0.08
chloroiodobenzene	1.95	1.93	0.02	1.4	1.42	0.02
chloronitrobenzene	4.277	3.89	0.38	3.207	2.99	0.22
dichlorobenzene	2.56	2.08	0.48	1.49	1.45	0.04
fluoroiodobenzene	2.02	1.90	0.12	1.38	1.41	−0.03
difluorobenzene	2.4	2.02	0.40	1.58	1.43	0.15
iodonitrosobenzene	3.82	4.00	−0.18	3.42	3.42	0.00
diiodobenzene	1.58	1.78	−0.20	1.28	1.37	−0.09
dinitrobenzene	6.065	6.34	−0.27	3.775	3.17	0.61
bromophenol	1.37	1.69	−0.32	d		
bromothiophenol	1.96	2.07	−0.11	1.51	1.91	−0.40
chlorophenol	1.333	1.73	−0.40	2.155	1.58	0.58
chlorothiophenol	1.98	2.12	−0.14	1.56	1.89	0.33
fluorophenol	1.32	1.70	−0.38	d		
iodophenol	1.54	1.58	−0.04	d		
nitrophenol	3.163	3.46	−0.30	3.925	3.82	0.11
bromoaniline	1.79	1.58	0.21	2.665	2.04	0.63
chloroaniline	1.78	1.79	−0.01	2.81	2.09	0.72
phenylendiamine	1.46	1.18	0.28	1.79	2.20	−0.41
aminophenol	d			1.85	1.90	−0.05

[a] Experimental values in benzene solution at 25 °C taken from ref 58. [b] Calculated from eq 8. [c] Calculated from eq 9. [d] Experimental value in benzene solution at 25 °C not reported.

those calculated by the model given below are shown in Table 6:

$$\mu(D) = -34.87 + 4.935 \cdot \mu_0 - 0.416 \cdot \mu_0 \cdot \mu_1 + 0.038 \cdot \mu_4$$
$$n = 30 \quad R = 0.956 \quad s = 0.444 \quad RMSCV = 0.471$$
$$F = 93.2 \quad (10)$$

As can be seen in Table 6, the predicted dipole moments of symmetrically substituted compounds are different from zero. This is a consequence of the nonconsideration of the symmetry of these molecules. However, when a symmetry indicator variable *I* is introduced, the statistical quality of the model is improved and the dipole moments predicted for most of the symmetric di-substituted benzene derivatives are equal to zero (see the last two columns in Table 6). The symmetry indicator variable takes value of one if the symmetry exists or equals to zero otherwise. The equation describing the dipole moments of p-substituted benzenes considering the symmetry is given below:

$$\mu(D) = -37.454 + 4.816 \cdot \mu_0 - 0.345 \cdot \mu_0 \mu_1 +$$
$$0.066 \cdot \mu_4 - 0.022 \cdot \mu_0 \mu_3 + 0.020 \cdot \mu_0 \mu_2 - 0.695 \cdot I$$
$$n = 30 \quad R = 0.976 \quad s = 0.356 \quad RMSCV = 0.383$$
$$F = 76.5 \quad (11)$$

As can be seen, the statistical parameters of this equation are similar to those obtained for o- and m-disubstituted benzene derivatives.

A data set of 26 mono- and di-substituted (*ortho*, *meta* and *para*) benzene derivatives was used as an external prediction set for testing the predictability of models (7)− (11). In Table 7 we give the experimental dipole moments of these compounds as well as those estimated from TOPS-MODE descriptors. For the sake of comparison we provide the dipole moments of these compounds estimated from full-

Table 6. Experimental and Calculated Dipole Moments of Para-Substituted Benzene Derivatives

compound	μ (D) exp.[a]	μ (D) calcd.[b]	residual	μ (D) calcd.[c]	residual
p-bromochlorobenzene	0.1	0.40	−0.30	0.45	−0.35
p-bromofluorobenzene	0.00	0.47	−0.47	0.52	−0.52
p-bromoiodobenzene	0.49	0.86	−0.37	0.88	−0.39
p-bromonitrosobenzene	1.94	1.39	0.54	1.90	0.04
p-dibromobenzene	0.00	0.52	−0.52	0.00	0.00
p-chlorofluorobenzene	0.97	0.35	0.62	0.40	0.57
p-chloroiodobenzene	0.46	0.73	−0.27	0.76	−0.30
p-chloronitrosobenzene	1.84	1.24	0.60	1.74	0.10
p-chloronitrobenzene	2.62	2.67	−0.04	2.57	0.05
p-dichlorobenzene	0.00	0.27	−0.27	0.00	0.00
p-fluoroiodobenzene	0.76	0.81	−0.05	0.83	−0.07
p-fluoronitrobenzene	2.645	2.78	−0.14	2.70	−0.06
p-iodonitrosobenzene	2.18	1.81	0.37	2.32	−0.14
p-iodonitrobenzene	2.873	3.36	−0.48	3.30	−0.43
p-diiodobenzene	0.00	1.20	−1.01	0.50	−0.50
p-nitronitrosobenzene	0.85	1.25	−0.40	0.82	0.03
p-bromophenol	2.238	1.80	0.44	1.76	0.57
p-bromothiophenol	1.14	0.73	0.41	1.05	0.09
p-chlorophenol	2.243	1.67	0.57	1.64	0.60
p-chlorobenzenesulfinic acid	2.2	2.15	0.05	2.19	0.01
p-chlorothiophenol	1.16	0.60	0.56	0.94	0.22
p-fluorophenol	2.135	1.75	0.39	1.72	0.42
p-iodophenol	2.21	2.14	0.07	2.07	0.14
p-iodothiophenol	1.04	1.07	−0.03	1.36	−0.32
p-nitrophenol	5.013	4.65	0.36	4.62	0.39
p-bromoaniline	2.923	2.99	−0.07	2.93	−0.01
p-chloroaniline	2.93	2.87	0.06	2.81	0.12
p-fluoroaniline	2.625	2.95	−0.32	2.88	−0.26
p-iodoaniline	2.84	3.33	−0.49	3.24	−0.40
p-nitroaniline	6.21	6.23	−0.02	6.22	−0.01

[a] Experimental value in benzene solution at 25 °C taken from ref 58. [b] Calculated from eq 10. [c] Calculated from eq 11.

geometry-optimization quantum chemical calculations. In this comparison we have used several semiempirical molecular orbital methods, such as Complete Neglect of Differential

Table 7. Experimental Dipole Moments of Benzene Derivatives in an External Prediction Set and Their Values Calculated by the Present Approach and Six Semiempirical Quantum Chemical Methods

compound	exp.	calcd.	CNDO	INDO	ZINDO/1	MNDO	AM1	PM3
-Cl	1.51	1.68	2.89	6.15	3.84	1.86	1.31	0.95
-NH$_2$	1.52	1.20	1.28	1.42	1.42	1.48	1.58	1.29
-CH$_2$OH	1.67	1.29	1.82	1.61	2.22	1.20	1.46	1.43
-C$_2$H$_5$	0.35	0.93	0.28	0.14	0.52	0.05	0.24	0.21
-I	1.42	1.52				1.60	1.43	0.79
-NH C$_2$H$_5$	1.70	0.80	1.76	1.26	1.59	1.10	1.61	1.36
-CH$_2$Cl	1.84	1.80	2.53	5.64	3.58	2.41	1.69	1.52
o-OH, CH$_3$	1.42	1.33	1.88	1.70	2.45	1.09	1.45	1.40
o-NO$_2$, CH$_3$	3.63	3.39	5.10	4.98	5.34	5.34	5.15	5.00
o-NH$_2$, CH$_3$	1.525	1.23	1.83	1.61	1.86	1.40	1.55	1.33
o-CN, CH$_3$	3.805	3.40	3.26	3.31	3.81	3.30	3.24	3.52
o- CH$_3$, CH$_3$	0.53	1.27	0.23	0.03	0.72	0.06	0.40	0.44
o−O CH$_3$, CH$_3$	1.0	2.51	1.38	1.37	1.26	1.14	1.04	0.90
m-OH, CH$_3$	1.78	1.62	1.98	1.79	2.52	1.14	1.53	1.41
m-NO$_2$, CH$_3$	4.23	4.19	5.22	5.05	5.80	5.41	5.41	5.41
m-NH$_2$, CH$_3$	1.532	1.97	1.24	1.42	1.31	1.46	1.49	1.26
m-CN, CH$_3$	4.21	4.20	3.37	3.32	4.25	3.33	3.49	3.77
m- CH$_3$, CH$_3$	0.415	1.71	0.23	0.10	0.53	0.05	0.26	0.26
m−O CH$_3$, CH$_3$	1.18	1.67	1.80	1.49	2.30	1.04	1.49	1.31
m−O CH$_3$, OCH$_3$	1.59	1.49	1.64	1.38	2.35	1.98	2.29	2.04
m-NO$_2$, OCH$_3$	3.89	3.93	3.55	3.61	3.84	4.62	4.32	4.49
m-NO$_2$, NH$_2$	4.89	4.92	5.91	5.92	6.43	6.40	6.30	6.34
p-OH, CH$_3$	1.65	3.38	1.68	1.60	2.09	1.14	1.31	1.18
p-NO$_2$, CH$_3$	4.47	5.22	5.64	5.36	6.36	5.53	5.73	5.73
p-NO$_2$, OCH$_3$	4.82	3.09	4.94	4.88	5.70	6.16	5.99	6.03
p-CN, CH$_3$	4.40	3.14	3.65	3.51	4.72	3.38	3.73	3.99
RMSE		0.76	0.72	1.34	1.04	0.75	0.65	0.65

Table 8. Group Contributions to Molecular Dipole Moment of Different Substituent in Benzene Derivatives Calculated According to the Traditional Approach (Difference between Property of Substituted and Nonsubstituted Derivative) and the Current Method Based on Total Bond Contributions Obtained from TOPS-MODE

group	traditional	current	group	traditional	current
OH	1.45	0.1	N(CH$_3$)$_2$	1.59	−0.15
CH$_3$	0.34	0.002	COOCH$_3$	2.53	1.48
F	1.42	0.53	OC$_2$H$_5$	1.00	−0.14
Br	1.50	0.51	COOC$_2$H$_5$	1.85	1.09
NO$_2$	3.93	2.94	CH$_2$COOCH$_3$	1.80	0.98
CHO	2.80	1.32	CH$_2$F	1.77	0.64
CN	4.05	2.49	CH≡CH	0.79	−0.09
OCH$_3$	1.30	0.21	CH$_2$CN	3.50	2.40
CF$_3$	2.56	1.45	OCH=CH$_2$	1.10	−0.14
COCH$_3$	2.80	1.30	COCl	3.16	1.87
NHCH$_3$	1.65	−0.08	CH$_2$NH$_2$	1.38	−0.08

Figure 2. Total bond contribution to the molecular dipole moment of four benzene derivatives with different substituent topologies.

Overlapping (CNDO),[89] Intermediate Neglect of Differential Overlapping (INDO),[90] Medium Neglect of Differential Overlapping (MNDO),[91] Austin Model 1 (AM1)[92] and Parametrization Method 3 (PM3).[93] More advanced ab initio computational methods are not used in the current work for the sake of time and because our main objective is to compare the current approach to other semiempirical ones.
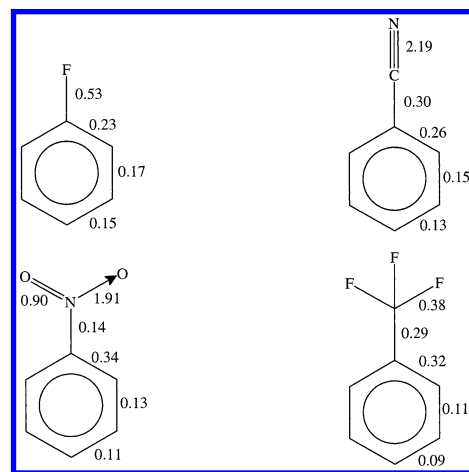
The root-mean-square error (RMSE) has been used as a criterion for the comparison of the predictability of the models. As can be seen in Table 7 the RMSE value of the predictions made by the models developed here is in the same range that those obtained with other theoretical approaches. In this work we do not pretend to propose a new method for calculating the values of molecular dipole moments as several well-reliable methods exist. Our only one intention here is to prove that the use of appropriate weighting schemes in combination with powerful graph-theoretic tools permits to model physicochemical (as well as biological and toxicological) properties outside the traditional region of the property space explored with "topological" descriptors. In this sense, the results showed in Table 7 are quite illustrative that this objective has been fulfilled. An advantage of the use of graph-theoretic approaches in QSPR, in particular of TOPS-MODE approach, is that we can obtain quantitative contributions of the different parts (substructures, groups, fragments, etc.) to the studied property.

PHYSICOCHEMICAL INTERPRETATION OF RESULTS

As early as in 1923 J. J. Thomson suggested that dipole moments of molecules should be the resultant of the vector summation of characteristic atomic, bond, and group contributions.[94] That is, the dipole moment is an additive property. A group moment is understood as the moment produced in a molecule by the replacement of a hydrogen atom by the group X in question. As the dipole moment of benzene molecule is zero the group dipole moment of X is the vector sum of the moments of the bonds of the group X and that of the hydrogen atom in the position *para* with respect to it. The last is generally taken arbitrarily as zero and the moment of the monosubstituted molecule is taken as the group moment of the substituent concerned.[95] This practice is also common in the generation of other group contributions in chemistry. For instance, when calculating the Hansch π constants for different groups in aromatic compounds the difference between the log$P$ of the substituted benzene derivative and that of the benzene molecule is considered:

$$\pi(X) = \log P(AX) - \log P(AH)$$

In the book *Electric Dipole Moments* from J. W. Smith[95] some of the group dipole moments in aromatic compounds are reported, some of them are reproduced in Table 8 first column.

As we have previously stated the main contribution of the TOPS-MODE approach to QSAR/QSPR studies is that it gives the possibility of finding structural contributions to the property studied in an explicit way. Here we give only the total bond contributions to the dipole moment. In Figure 2 we give these values for four benzene derivatives with different topologies of the substituent (fluorobenzene, cy-

anobenzene, nitrobenzene and trifluorobenzene). The sum of these bond contributions gives the value of the molecular dipole moment. That is, we have found bond contributions that can be summed not in a vectorial way giving the value of the molecular dipole moment. As can be seen in this figure the values of the molecular dipole moment are not concentrated on the bonds constituting the substituent but they are spread over all the molecule. This can be interpreted as a transmission of the dipolar effect of the substituent group for all bonds in the molecule. In fact, a picture much more realistic than the previous one assuming zero values for benzene ring bonds.

In the current approach we can obtain the values of the group dipole moment as the sum of the bond dipoles of the corresponding group. It is clear that we are leaving outside this consideration the proper influence of the substituent over the rest of the molecule, but we consider that it is still a more realistic approach that considering the group dipole as that of the whole molecule (as previously done in the literature). By this means the dipole of the F group is 0.53 D, of CN is 2.39 D (0.3 + 2.19), of NO_2 is 2.95 D (0.14 + 1.91 + 0.9) and that of CF_3 is 1.43 D (0.38*3 + 0.29). In Table 8 we give in the second column the values calculated according to this method. A correlation coefficient of 0.92 exists between both kind of group dipole moments, which indicates that they are related to each other but not necessarily closely related. As we explain in the next paragraph the use of these two approaches account for different effects of the substituent over the rest of the molecule.

The question about the appropriateness of the current group contributions compared to that traditionally obtained by difference between the substituted compounds minus the nonsubstituted one is illustrated below. Consider a property that is simply the sum of atomic charges for non-hydrogen atoms in the molecule:

$$P = \sum_i Q_i$$

If we consider a monosubstituted benzene derivative (AX), we can obtain the group property P(X) as the difference P(AX) − P(AH). In fact, we calculate this property for toluene, aniline, phenol and fluorobenzene using full-geometry optimization AM1 calculations.[92] The values of P(AX) are as follows: P(A-CH$_3$) = −0.646; P(A-NH$_2$) = −0.637; P(A-OH) = −0.683 and P(A−F) = −0.705, while P(A−H) = −0.780. These values are obtained by summing the atomic charges on carbon atoms of the benzene ring plus atomic charges on atoms of the corresponding substituent (hydrogen included). According to the traditional approach of using the difference between substituted and not-substituted derivative for calculating group contribution we obtain the following values for these groups: P(CH$_3$) = 0.134; P(NH$_2$) = 0.143; P(OH) = 0.097 and P(F) = 0.075. However, the sum of the charges of the atoms in the corresponding groups give the following values of group contributions: P(CH$_3$) = 0.073; P (NH$_2$) = 0.032; P(OH) = −0.036 and P(F) = −0.107. As can be seen significant differences exist between both series of values, a correlation between them exist but it is only of 0.942, and the most significant thing is the differences in the order and sign of the values obtained by both approaches.

We have not selected this "hypothetical" property incidentally. We have selected the sum of the atomic charges as a property because of its relation with the *n*-octanol/water partition coefficient. According to Cammarata and Rogers[96] the logP values and π constants can be calculated from a quantum chemical approach, using the sum of atomic charges and the sum of electrophilic superdelocalizabilities as the only two parameters. This means that when the substituent constants, such as Hansch π constants, are calculated we are not considering the transmission effects of this property from the substituent to the rest of the molecule. This kind of effect, however, can easily be taken into account by using a graph-theoretic approach as that used in the current work or by using quantum chemical approaches as that of Cammarata and Rogers for logP.[96]
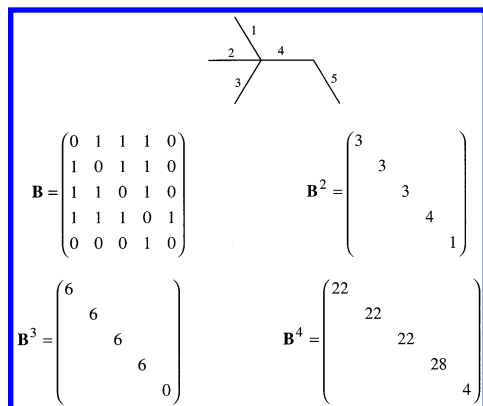
CONCLUSIONS

The main objective of the present study has not been to propose a novel approach to calculate the dipole moments of organic compounds, i.e., benzene derivatives, for which very good quantum chemical and molecular mechanics approaches exist. What we want to show here are the potentials that graph theoretic approaches have in the study of physicochemical problems. When QSPR models are developed, two main questions should be taken into account in order to evaluate the quality of the models found. First, the statistical quality of the QSPR equation, and very specially its predictive ability, should be evaluated. Second, but not least important, the model developed should be physically meaningful, permitting an interpretation of the physicochemical processes in terms of structural features of molecules. In the present study we have shown that by using graph theoretic descriptors it is possible to develop QSPR models that have both statistical quality and physical significance.

We have also demonstrated, that the applications of graph theoretic molecular descriptors are not limited to the description of such properties that are in the region of the property space traditionally related to molecular bulkiness. It is clear that the use of graph theoretic descriptors to model other physicochemical properties depends on an adequate selection of the graph invariants, for which some rules put forward.[52] However, the application of the very powerful tools of algebra to the manipulation of molecular graphs permits the generation of novel and interesting molecular descriptors. The next step is, of course, the development of QSPR models describing novel physicochemical properties. These models should be obtained by following the paradigm of Heisenberg[96] about the phenomenological approaches in physical sciences, that is "inventing formulas which seem to reproduce the experiments".

APPENDIX

The spectral moments of the bond matrix are calculated in the following way. Considering the molecule of 2,2-dimethylbutane as in the example given in the main text we have the following labeled molecular graph and bond adjacency matrix. The second, third and fourth powers of

GRAPH THEORETIC DESCRIPTORS IN QSPR/QSAR

J. Chem. Inf. Comput. Sci., Vol. 43, No. 1, 2003 **83**

this matrix are also given, in these cases we only show the main diagonal of the matrices:

The zeroth spectral moment is simply the order of the matrix, i.e., the number of bonds in the molecule. The spectral moments given in the example provided in the main text of the article are calculated as follows:

$$\mu_2^T = \mu_2 = \mathbf{Tr}(\mathbf{B}^2) = \sum_i^n (\mathbf{B}^2)_{ii} = 3 + 3 + 3 + 4 + 1 = 14$$

$$\mu_3^T = \mu_3 = \mathbf{Tr}(\mathbf{B}^3) = \sum_i^n (\mathbf{B}^3)_{ii} = 6 + 6 + 6 + 6 + 0 = 24$$

$$\mu_4^T = \mu_4 = \mathbf{Tr}(\mathbf{B}^4) = \sum_i^n (\mathbf{B}^4)_{ii} = 22 + 22 + 22 + 28 +$$

$$4 = 98$$

The terms in the summations for calculating the total spectral moments are the so-called bond spectral moments. We have written these terms in the consecutive order of the bond labels given in the graph. For instance, the bond moments of order 2, 3 and 4 for the bond labeled as 1 are 3, 6 and 22, respectively.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) Randić, M. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Ed.; John Wiley & Sons: New York, 1998; Vol. 5, p 3018.
(2) Estrada, E.; Uriarte, E. Recent advances on the role of topological indices in drug discovery research. *Curr. Med. Chem.* **2001**, *8*, 1573−1588.
(3) *QSPR/QSAR Studies by Molecular Descriptors*; Diudea, M. V., Ed.; Nova Science Publishers: Huntington, NY, 2001.
(4) Kier, L. B.; Hall, L. H. *Molecular Structure Description. The Electrotopological State*; Academic Press: New York, 1999.
(5) Devillers, J. New trends in (Q)SAR modeling with topological indices. *Curr. Op. Drug Discov. Dev.* **2000**, *3*, 275, 275−279.
(6) *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; 1999.
(7) Balaban, A. T. Topological and stereochemical molecular descriptors for databases useful in QSAR similarity/dissimilarity and drug design. *SAR QSAR Environ. Res.* **1998**, *8*, 1−21.
(8) *From Chemical Topology to Three-Dimensional Geometry*; Balaban, A. T., Ed.; Plenum Press: New York, 1997.
(9) Mihalić, Z.; Trinajstić, N. A graph-theoretical approach to structure−property relationships. *J. Chem. Educ.* **1992**, *69*, 701−712.
(10) Stankevich, M. I.; Stankevich, I. V.; Zefirov, N. S. Topological indexes in organic chemistry. *Usp. Khim.* **1988**, *57*, 337−366.
(11) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure−Activity Analysis*; Research Studies Press: Letchworth, 1986.
(12) Sabljić, A.; Trinajstić, N. Quantitative structure−activity relationships: The role of topological indices. *Acta Pharm. Jugosl.* **1981**, *31*, 189−214.
(13) Randić, M. The Nature of Chemical Structure. *J. Math. Chem.* **1990**, *4*, 157−184.
(14) Randić M. Chemical Structure-what is "she"? *J. Chem. Educ.* **1992**, *69*, 713−718.
(15) Randić M. On characterization of chemical srructure *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 672−687.
(16) Basak, S. C.; Niemi, G. J.; Veith, G. D. Optimal characterization of structure for prediction of properties. *J. Math. Chem.* **1990**, *4*, 185−205.
(17) Basak, S. C.; Niemi, G. J.; Veith, G. D. Predicting properties of molecules using graph invariants. *J. Math. Chem.* **1991**, *7*, 243−272.
(18) Albert, A. *Selective Toxicity. The Physicochemical basis of Therapy*; Chapman & Hall: London, 1984.
(19) Pogliani, L. Molecular connectivity model for determination of physicochemical properties of α-amino acids. *J. Phys. Chem.* **1993**, *97*, 6731−6736.
(20) Pogliani, L. Molecular connectivity descriptors of the physicochemical properties of the α-amino acids. *J. Phys. Chem.* **1994**, *98*, 1494−1499.
(21) Pogliani, L. Molecular Modeling by Linear Combinations of Connectivity Indexes. *J. Phys. Chem.* **1995**, *99*, 925−937.
(22) Pogliani, L. Modeling with Special Descriptors Derived from a Medium-Sized Set of Connectivity Indices. *J. Phys. Chem.* **1996**, *100*, 18065−18077.
(23) Pogliani, L. Modeling Purines and Pyrimidines with the Linear Combination of Connectivity Indices-Molecular Connectivity "LCCI-MC" Method. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1082−1091.
(24) Gálvez, J.; García-Domenech, R.; de Julian-Ortiz, J. V.; Soler, R. Topological Approach to Drug Design. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 272−284.
(25) Gálvez, J.; García-Domenech, R.; De Julian-Ortiz, V.; Soler, R. Topological Approach to Analgesia. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1198−1203.
(26) Gálvez J.; García-Domenech R.; de Gregorio A. C.; de Julian-Ortiz J. V.; Popa L. Pharmacological distribution diagrams: a tool for de novo drug design. *J. Mol. Graph. Modell.* **1996**, *14*, 272−276.
(27) García-Domenech, R.; De Gregorio A. C.; De Julian-Ortiz, J. V.; Gálvez, J.; Popa, L. Molecular connectivity to find β-blockers with low toxicity. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 567−572.
(28) Randić, M. Molecular topographic descriptors. *Stud. Phys. Theor. Chem.* **1988**, *54*, 101−108.
(29) Randić, M.; Jerman-Blazic, B.; Trinajstić, N. Development of 3-Dimensional molecular Descriptor. *Comput. Chem.* **1990**, *14*, 237−246
(30) Bogdanov, B.; Nikolić, S.; Trinajstić, N. On the three-dimensional Wiener number. *J. Math. Chem.* **1989**, *3*, 299−309.
(31) Zhu, H. Y.; Klein, D. J. Graph-Geometric Invariants for Molecular Structures. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1067−1075.
(32) Klein, D. J.; Zhu, H. Y. Distances and volumina for graphs. *J. Math. Chem.* **1998**, *23*, 179−195.
(33) Estrada, E.; Montero, L. A. Bond order weighted graphs in molecules as structure−property indexes. *Mol. Eng.* **1993**, *2*, 363−373.
(34) Estrada, E. Three-Dimensional Molecular Descriptors Based on Electron Charge Density Weighted Graphs. *J. Chem. Inf. Comput. Sci.* **1995**,
(35) Estrada, E.; Ramirez, A. Edge Adjacency Relationships and Molecular Topographic Descriptors. Definition and QSAR Applications. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 837−843.
(36) Estrada, E.; Molina, E. 3D connectivity indices in QSPR/QSAR studies*J. Chem. Inf. Comput. Sci.* **2001**, *41*, 791−797.
(37) Estrada, E. Physicochemical interpretation of molecular connectivity indices. *J. Phys. Chem. A* **2002**, *106*, 9085−9091.
(38) Randić, M.; Sabljić, A.; Nikolić, S.; Trinajstić, N. A rational selection of graph-theoretical indexes in the QSAR. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1988**, *15*, 267−285.
(39) Randić, M. Characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609−6615.
(40) Hosoya, H. Topological index. Newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332−2339.
(41) Balaban, A. T. Chemical graphs. Part 48. Topological index *J* for heteroatom-containing molecules taking into account periodicities of element properties. *MATCH* **1986**, *21*, 115−122.
(42) Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **1947**, *69*, 17−20.
(43) Estrada, E. Edge Adjacency Relationships and a Novel Topological Index Related to Molecular Volume. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 31−33.

(44) Randić, M. On computation of optimal parameters for multivariate analysis of structure−property relationship. *J. Comput. Chem.* **1991**, *12*, 970−980.

(45) Randić, M. Novel graph theoretical approach to heteroatoms in quantitative structure−activity relationships. *Chemom. Intell. Lab. Syst.* **1991**, *10*, 213−227.

(46) Hall, L. H.; Mohney, B.; Kier, L. B. The electrotopological state: an atom index for QSAR. *Quant. Struct.-Act. Relat.* **1991**, *10*, 43−51.

(47) Hall, L. H.; Mohney, B.; Kier, L. B. The electrotopological state: structure information at the atomic level for molecular graphs. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 76−82.

(48) Estrada, E. Edge adjacency relationships in molecular graphs containing heteroatoms: a new topological index related to molar volume. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 701−707.

(49) Gálvez, J.; García, R.; Salabert, M. T.; Soler, R. Charge Indexes. New Topological Descriptors. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 520−525.

(50) Ivanciuc, O. QSAR Comparative Study of Wiener Descriptors for Weighted Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1412−1422.

(51) Trinajstić, N.; Randić, M.; Klein, D. J. On the quantitative structure−activity relationship in drug research. *Acta Pharm. Jugosl.* **1986**, *362*, 267−279.

(52) Randić, M. Generalized molecular descriptors. *J. Math. Chem.* **1991**, *7*, 155−68.

(53) Cramer, R D., III. BC(DEF) parameters. 2. An empirical structure-based scheme for the prediction of some physical properties. *J. Am. Chem. Soc.* **1980**, *102*, 1849−1859.

(54) Cramer, R. D., III. BC(DEF) parameters. 1. The intrinsic dimensionality of intermolecular interactions in the liquid state. *J. Am. Chem. Soc.* **1980**, *102*, 1837−1849.

(55) Needham, D. E.; Wei, I. C.; Seybold, P. G.. Molecular modeling of the physical properties of alkanes. *J. Am. Chem. Soc.* **1988**, *110*, 4186−4194.

(56) Malinowski, E. R.; Howery, D. G. *Factor Analysis in Chemistry*; Wiley-Interscience: New York, 1980.

(57) The physicochemical properties were taken from the following: Weast, R. C. *CRC Handbook of Chemistry and Physics*, 67th ed.; CRC: Boca Raton, FL, 1986−1987. The dipole moments were take from ref 57.

(58) McClellan, A. L. *Tables of Experimental Dipole Moments*; W. H. Freeman & Co.: New York, 1963.

(59) *STATISTICA for Windows*; Release 4.3. StatSoft, Inc., 1993.

(60) Estrada, E. Spectral Moments of the Edge Adjacency Matrix in Molecular Graphs. 1. Definition and Applications to the Prediction of Physical Properties of Alkanes. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 844−849.

(61) Estrada, E. Spectral Moments of the Edge-Adjacency Matrix of Molecular Graphs. 2. Molecules Containing Heteroatoms and QSAR Applications. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 320−328.

(62) Estrada, E. Spectral Moments of the Edge Adjacency Matrix in Molecular Graphs. 3. Molecules Containing Cycles. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 23−27.

(63) Estrada, E.; Peña, A.; García-Domenech, R. Designing sedative/hypnotic compounds from a novel substructural graph-theoretical approach. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 583−595.

(64) Estrada, E.; Gutiérrez, Y.; González, H. Modeling Diamagnetic and Magnetooptic Properties of Organic Compounds with the TOSS-MODE Approach. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1386-1399.

(65) Estrada, E.; Peña, A. In silico studies for the rational discovery of anticonvulsant compounds. *Bioorg. Med. Chem.* **2000**, *8*, 2755−2770.

(66) Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L.; De Clercq, E. A Novel Approach for the Virtual Screening and Rational Design of Anticancer Compounds. *J. Med. Chem.* **2000**, *43*, 1975−1985.

(67) Estrada, E. On the topological sub-structural molecular design (TOSS-MODE) in QSPR/QSAR and drug design research. *SAR QSAR Environ. Res.* **2000**, *11*, 55−73.

(68) Estrada, E.; Uriarte, E. Quantitative structure-toxicity relationships using TOPS-MODE. 1. Nitrobenzene toxicity to *Tetrahymena pyriformis*. *SAR QSAR Environ. Res.* **2001**, *12*, 309−324.

(69) Estrada, E.; Molina, E.; Uriarte, E. Quantitative structure-toxicity relationships using Tops-Mode. 2. Neurotoxicity of a noncongeneric series of solvents. *SAR QSAR Environ. Res.* **2001**, *12*, 445−459.

(70) *Chemical Applications of Graph Theory*; Balaban, A. T., Ed.; Academic Press: New York; 1976; p 180.

(71) Trinajstić, N. *Chemical Graph Theory*; CRC Press: Boca Raton, 1992.

(72) Bieri, G.; Dill, J. D.; Heilbronner, E.; Schmelzer, A. Application of the equivalent bond orbital model to the C2s-ionization energies of saturated hydrocarbons. *Helv. Chim. Acta* **1977**, *60*, 2234−2347.

(73) Gutman, I.; Popovic, L.; Estrada, E.; Bertz, S. H. The line graph model. Predicting physicochemical properties of alkanes. *ACH−Models Chem.* **1998**, *135*, 147−155.

(74) Smolenskii, E. A. Utilization of the theory of graphs for the calculation of structure-additive properties of hydrocarbons. *Zh. Fiz. Khim.* **1964**, *38*, 1288−1291.

(75) Gordon, M.; Kennedy, J. W. Graphlike state of matter. 2. LCGI [linear combinations of graph-theoretical invariants] schemes for the thermodynamics of alkanes and the theory of inductive inference. *J. Chem. Soc., Faraday Trans. 2* **1973**, *69*, 484−504.

(76) Essam, J. W.; Kennedy, J. W.; Gordon, M.; Whittle, P. The graph-like state of matter. Part 8. LCGI schemes and the statistical analysis of experimental data. *J. Chem. Soc., Faraday Trans. 2* **1977**, *73*, 1289−1307.

(77) Klein, D. J.; Schmalz, T. G.; Bytautas, L. Chemical sub-structural cluster expansions for molecular properties. *SAR QSAR Environ. Res.* **1999**, *10*, 131−136.

(78) Jiang, Y.; Tang, A.; Hoffmann, R. Evaluation of moments and their application in Hueckel molecular orbital theory. *Theor. Chim. Acta* **1984**, *66*, 183−192.

(79) Burdett, J. K.; Lee, S. Moments method and elemental structures. *J. Am. Chem. Soc.* **1985**, *107*, 3063−3082.

(80) Burdett, J. K.; Lee, S. Moments and the energies of solids. *J. Am. Chem. Soc.* **1985**, 107, 3050−3063.

(81) Lee, S. Second-moment scaling and covalent crystal structures. *Acc. Chem. Res.* **1991**, *24*, 249−254.

(82) Markovic, S.; Gutman, I. Dependence of spectral moments of benzenoid hydrocarbons on molecular structure. *J. Mol. Struct. Theochem* **1991**, *81*, 81−87.

(83) Gutman, I. Remark on the moment expansion of total p-electron energy. *Theor. Chim. Acta* **1992**, *83*, 313−318.

(84) Gutman, I.; Rosenfield, V. R. Spectral moments of polymer graphs. *Theor. Chim. Acta* **1996**, *93*, 191−197.

(85) Karwowski, J.; Bielinska-Waz, D.; Jurkowski, J. Eigenvalues of model Hamiltonian matrixes from spectral density distribution moments: the Heisenberg spin Hamiltonian. *Int. J. Quantum Chem.* **1996**, *60*, 185−193.

(86) Estrada, E.; Molina, E. Novel local (fragment-based) topological molecular descriptors for QSPR/QSAR and molecular design. *J. Mol. Graph. Modell.* **2001**, *20*, 54−64.

(87) Estrada, E.; Molina, E. QSPR/QSAR by graph theoretical descriptors beyond the frontiers. In *QSPR/QSAR Studies by Molecular Descriptors*; Diudea, M., Ed.; Nova Sci.: New York, 2001; pp 83−107.

(88) Smyth, C. P. Bond moment and electronegativity. *J. Phys. Chem.* **1937**, *41*, 209−219.

(89) Pople, J. A.; Segal, G. A. Approximate self-consistent molecular orbital theory. III. CNDO results for AB2 and AB3 systems. *J. Chem. Phys.* **1966**, *44*, 3289−3296.

(90) Pople, J. A.; Beveridge, D. L.; Dobosh, P. A. Approximate self-consistent molecular-orbital theory. V. Intermediate neglect of differential overlap. *J. Chem. Phys.* **1967**, *47*, 2026−2033.

(91) (a) Dewar, M. J. S.; Thiel, W. Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J. Am. Chem. Soc.* **1977**, *99*, 4899−4907. (b) Dewar, M. J. S.; Thiel, W. Ground states of molecules. 39. MNDO results for molecules containing hydrogen, carbon, nitrogen, and oxygen. *J. Am. Chem. Soc.* **1977**, *99*, 4907−4917.

(92) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(93) (a) Stewart, James J. P. Optimization of parameters for semiempirical methods. I. Method. *J. Comput. Chem.* **1989**, *10*, 209−320. (b) Stewart, J. J. P. Optimization of parameters for semiempirical methods. II. Applications. *J. Comput. Chem.* **1989**, *10*, 221−64.

(94) Thomson, J. J. Studies in the electron theory of chemistry. The changes in chemical properties produced by the substitution of one element or radical by another, with implications to benzene ring. *Phil Mag.* **1923**, *46*, 497−514.

(95) Smith, J. W. *Electric Dipole Moments*; Butterworth Sci. Pub.: London, 1955.

(96) (a) Rogers, K. S.; Cammarata, A. Superdelocalizability and charge density. A correlation with partition coefficients. *J. Med. Chem.* **1969**, *12*, 692−693. (b) Rogers, K. S.; Cammarata, A. A molecular orbital description of the partitioning of aromatic compounds between polar and nonpolar phases. *Biochim. Biophys. Acta* **1969**, *193*, 22−29. (c) Cammarata, A.; Rogers, K. S. Electronic representation of the lipophilic parameter pi. *J. Med. Chem.* **1971**, *14*, 269−274.

(97) Heisenberg, W. In *From a Life of Physics*; Salam, A., Ed.; World Scientific: Singapore, 1989; p 31.