

Can 3D Structural Parameters Be Predicted from 2D (Topological) Molecular Descriptors?

Ernesto Estrada,* Enrique Molina, and Iliana Perdomo-López

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

Received December 28, 2000

The dihedral angle between both phenyl rings determined by photoelectron spectroscopy in a series of seven alkylbiphenyl is described by the local spectral moments of the bond matrix. This series is extended to 78 alkylbiphenyl compounds by estimating the dihedral angle from molecular mechanics force field calculations. The linear correlation obtained between this angle and the local spectral moments shown a correlation coefficient of 0.9838. This result proves that 2D (topological) descriptors can account for 3D structural parameters. A new substituent constant is calculated as the contribution of groups to the studied rotational angle by using the information encoded into the local spectral moments. This substituent constant is not linearly related to the Taft's steric constants E_s as they have a correlation coefficient of only 0.75. These steric constants are able to account only for 71% of the variance in the studied 3D parameter. The implications for QSPR/QSAR studies of the demonstration that 2D (topological) descriptors can describe 3D structural parameters are also analyzed.

INTRODUCTION

Molecular structure can be understood as a hybrid of topologic, geometric, and electronic features.^{1,2} The molecular topology is concerned to the number and type of atoms in the molecule and the way they are connected to each other.^{3,4} Molecular geometry is defined as the disposition of atoms in a 3D space, and it is characterized by three main structural parameters, bond lengths, bond angles, and dihedral angles, which define completely the three-dimensional structure of the molecule.⁵ On the other hand, electronic molecular structure is well-characterized through the molecular wave function from which the electronic distribution (and electronic properties so-derived) in the molecule can be known.⁶

These three molecular features (topological, geometric, and electronic) are intimately related to each other. Thus, some topological patterns are responsible for some geometrical features of the molecule, and they determine most of the electronic parameters of such molecule. As an example, we can mention a simple topological parameter such as the multiplicity of a bond. A single C–C bond is characterized by a geometrical parameter that is more or less constant despite the molecular environment in which the bond is located. That is, the standard bond distance of this single bond is 1.54 Å. However, a change in the topology of this bond, e.g., by transforming its multiplicity to a double or triple bond, will transform its geometric parameter to values of 1.34 and 1.20 Å, respectively.⁷ It is clear that the electronic distribution around these bonds will change as a function of their multiplicity having the classical shapes for the different hybridizations involved in them.⁸

A very illustrative lesson of this interrelation can be taken borrowed from protein structures. It is well-known that the

primary structure of a protein contains most of the necessary information to obtain its 3D structure. In this case, primary structure represents protein molecular topology because it contains the information on what amino acids compose this structure and in what order they are bonded to each other (the sequence). In fact, it has been stated that “the three-dimensional structure of a protein is determined by its genetically encoded sequence”.^{9,10}

In the context of characterization of molecular structure, topological, geometric, and electronic descriptors and approaches can also be distinguished.¹¹ Most of the topological descriptors are based on the use of graph-theoretical concepts applied to chemistry.^{12,13} Among the several approaches to find geometric parameters from theoretical methods, force fields molecular mechanic appears to be the most attractive one,¹⁴ and the use of semiempirical, ab initio, or density functional theory quantum chemical calculations are the main election for computing electronic molecular descriptors.¹⁵ It has been believed for some time that topological indices (TIs) contain some important information on the 3D molecular structure. This possibility arises probably from the great success of these descriptors in structure–activity relationship studies,^{15–17} where it is well-known that 3D molecular features play an important role. There have been some attempts to relate the graph-theoretical characterization of molecules to their “real” three-dimensional structure as determined by X-ray crystallography.^{18,19} However, up to now nobody has answered with facts such as the following: Can 3D (geometric) structural parameters be predicted from 2D (topological) molecular descriptors.

Here we will be concerned with the problem of predicting geometric parameters from topological descriptors. Thus, we study the dihedral angle between both phenyl rings of a set of biphenyl derivatives. The values of this angle experimentally determined by photoelectron spectroscopy or calculated by molecular mechanics MM+ force field will be described

* Corresponding author phone: 34-981-563100 ext 14938; fax: 34-981-594912; e-mail: estrada66@yahoo.com.

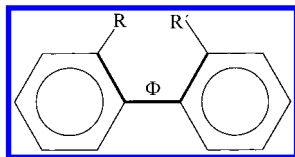


Figure 1. Dihedral angle Φ between both phenyls in alkylbiphenyl derivatives studied here.

by using local spectral moments of the bond adjacency matrix. It will be proved how this geometrical parameter can be expressed as an additive function of group contributions showing its complete determination from a topological parameter.

METHODS

Data Set and Calculations. We have selected the dihedral angle between both phenyl rings in a series of alkylbiphenyl compounds to be described by the local spectral moments of the bond matrix. The first data set that will be studied here is composed by seven alkyl-substituted biphenyls for which the geometry have been determined in gas phase by photoelectron spectroscopy (PES).²⁰ However, this data set is very small to obtain conclusive results on the type we are pretending in the current study. Thus, we have selected a data set of 78 biphenyls having different alkyl substituents at positions 2 and 2' of the phenyl rings, whose geometries were calculated by molecular mechanics. The molecular structures of all these molecules were drawn, and full-geometry optimization was carried out by using the molecular mechanics MM+ force field.²¹ Then we compute the dihedral angle between both phenyl rings, Φ , as shown in Figure 1. In carrying out this geometry optimization the dihedral angle Φ was systematically changed in order to guarantee that the minimal energy conformation was achieved and not be fooled by local minima.

2D Molecular Descriptors. The structural parameters that we will use in the current work are the local spectral moments of the bond matrix.^{22,23} These molecular descriptors have been recently introduced in the TOPS-MODE strategy²⁴ for molecular design.^{22,23} Bond matrix is a square and symmetric matrix whose elements e_{ij} are ones or zeros depending if the corresponding bonds i and j in the molecule are adjacent or not.^{25,26} Two bonds are adjacent if they are incident to the same atom. The total spectral moments of the bond matrix, which the TOPS-MODE is based on, are defined as

$$\mu_k = \text{Tr}(\mathbf{B}^k) = \sum_{i=1}^s (e_{ii})^k \quad (1)$$

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 k th power of the bond matrix.

Local spectral moments are defined as the sum of the diagonal entries of the different powers of the bond matrix corresponding to a given fragment of the molecule. In mathematical terms local spectral moments of the bond matrix are defined as follows^{22,23}

$$\mu_k = (f) = \sum_{i=1}^f (e_{ii})^k \quad (2)$$

where f is the corresponding fragment for which the moments

are defined, and the sum is carried out over all bonds conforming the fragment f . The simplest case is when f corresponds to a single bond, and in this case the k th local moment is defined as the diagonal entry corresponding to this bond in the matrix raised to the k th power.²² Bond spectral moments of the bond matrix corresponding to the central bond of the dihedral angle Φ will be used in the current work as 2D molecular descriptors (see Figure 1).

Fragment Contributions. In previous papers on the TOPS-MODE approach we have shown that the spectral moments of the bond matrix can be expressed as linear combinations of the different fragments of the molecule.^{27–36} Similarly, local moments of a fragment can be expressed in terms of substructures of different sizes containing the corresponding fragment. The simplest case is that of spectral moments defined for bonds in the molecule without considering heteroatoms in the molecule. In this case the local bond moments, $\mu_k(i)$, are expressed in terms of structural fragments through the following expressions:

$$\mu_2(i) = |F_2(i)| \quad (3)$$

$$\mu_3(i) = 2|F_3(i)| \quad (4)$$

$$\mu_4(i) = |F_2(i)| + 4|F_3(i)| + 6|F_4(i)| + |F_5^1(i)| + 2|F_5^2(i)| \quad (5)$$

$$\mu_5(i) = 10|F_3(i)| + 30|F_4(i)| + 2|F_6^1(i)| + 4|F_6^2(i)| + 2|F_6^3(i)| \quad (6)$$

$$\begin{aligned} \mu_6(i) = & |F_2(i)| + 20|F_3(i)| + 120|F_4(i)| + 3|F_5^1(i)| + \\ & 6|F_5^2(i)| + 5|F_6^1(i)| + 10|F_6^2(i)| + 4|F_6^3(i)| + |F_7^1(i)| + \\ & 2|F_7^2(i)| + 6|F_8^1(i)| + 12|F_8^2(i)| + 6|F_8^3(i)| + 4|F_9^1(i)| + \\ & 8|F_9^2(i)| \quad (7) \end{aligned}$$

$$\begin{aligned} \mu_7(i) = & 42|F_3(i)| + 420|F_4(i)| + 18|F_6^1(i)| + 34|F_6^2(i)| + \\ & 14|F_6^3(i)| + 36|F_8^1(i)| + 72|F_8^2(i)| + 30|F_8^3(i)| + \\ & 18|F_9^1(i)| + 40|F_9^2(i)| + 2|F_{10}^1(i)| + 4|F_{10}^2(i)| + \\ & 4|F_{10}^3(i)| + 2|F_{10}^4(i)| + 2|F_{11}^1(i)| + 4|F_{11}^2(i)| + \\ & 12|F_{12}^1(i)| + 24|F_{12}^2(i)| + 12|F_{12}^3(i)| \quad (8) \end{aligned}$$

$$\begin{aligned} \mu_8(i) = & |F_2(i)| + 84|F_3(i)| + 1386|F_4(i)| + 7|F_5^1(i)| + \\ & 14|F_5^2(i)| + 44|F_6^1(i)| + 82|F_6^2(i)| + 30|F_6^3(i)| + \\ & 5|F_7^1(i)| + 11|F_7^2(i)| + 178|F_8^1(i)| + 342|F_8^2(i)| + \\ & 132|F_8^3(i)| + 74|F_9^1(i)| + 168|F_9^2(i)| + 5|F_{10}^1(i)| + \\ & 10|F_{10}^2(i)| + 8|F_{10}^3(i)| + 5|F_{11}^1(i)| + 12|F_{11}^2(i)| + \\ & 6|F_{11}^3(i)| + 96|F_{12}^1(i)| + 204|F_{12}^2(i)| + 90|F_{12}^3(i)| + \\ & |F_{13}^1(i)| + 2|F_{13}^2(i)| + 2|F_{13}^3(i)| + 6|F_{14}^1(i)| + 12|F_{14}^2(i)| + \\ & 12|F_{14}^3(i)| + 6|F_{14}^4(i)| + 6|F_{15}^1(i)| + 12|F_{15}^2(i)| + \\ & 6|F_{15}^3(i)| + 4F_{16}^1(i) + 8F_{16}^2(i) + 4F_{16}^3(i) + 8F_{16}^4(i) + \\ & 4F_{16}^5(i) + 4F_{17}^1(i) + 8F_{17}^2(i) + 36|F_{18}^1(i)| + 72|F_{18}^2(i)| \quad (9) \end{aligned}$$

where the terms $F_p^q(f)$ represent subgraphs of type p , subtype q , containing the bond f (given in Figure 2), and the vertical bars mean the number of such subgraphs.

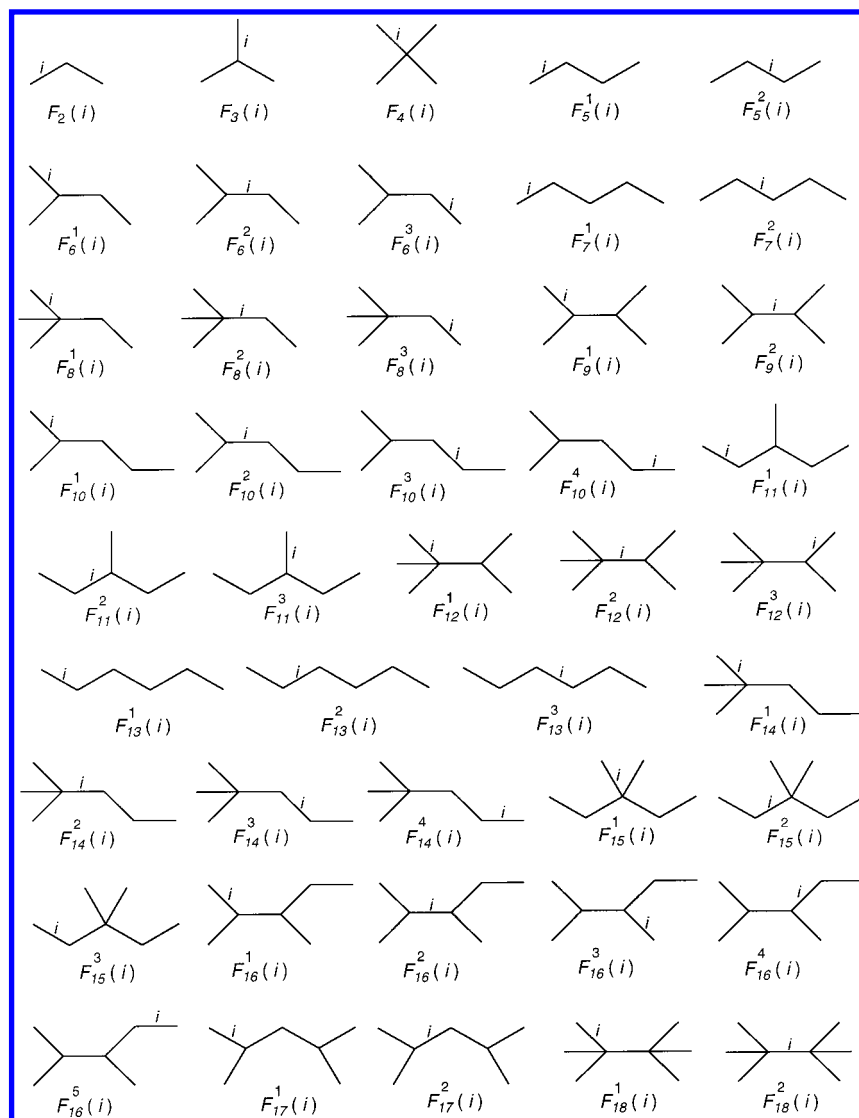


Figure 2. Acyclic structural fragments contributing to the local spectral moments of order not greater than 8.

RESULTS AND DISCUSSION

3D Parameter vs 2D Descriptors. The main objective of the current work is to demonstrate whether a 3D parameter, such as a dihedral angle, can be predicted from a QSPR model using 2D descriptors. The molecular descriptors used in the current are the local spectral moments of the bond matrix. This matrix is a “purely” topological object, as it represents only the connectivity of the different edges of the molecular graph representing the molecule. No one kind of vertex or edge weight was used in these calculations. Consequently, there is no doubt that the molecular descriptors used here are of “pure” topological nature.

In the first part of this study we obtained a linear model relating the dihedral angle Φ experimentally determined by PES²⁰ and the local spectral moments of the bond matrix calculated by our system TOPS-MODE.³⁷ The best linear model is given below together with the statistical parameters of the regression

$$\Phi_{\text{PES}}(\text{deg}) = -39.23(\pm 12.44) + 0.067(\pm 0.0083)\mu_8(i) \quad (10)$$

$$N = 7, R = 0.9636, R_{\text{CV}} = 0.9562, s = 4.40,$$

$$F = 65, \text{RMSECV} = 6.39$$

Table 1. Dihedral Angles between Phenyl Rings Φ in Alkylbiphenyls Determined by Photoelectron Spectroscopy (PES) and the Values of Φ Predicted with Local Spectral Moments and Residuals and Cross-Validation Residuals from the Linear Regression

compound	exp. (PES) ^a	pred.	res.	CV-res
biphenyl	41.6	45.6	-4.0	-5.9
2-CH ₃	70	64.6	5.4	6.4
3-CH ₃	46	47.2	-1.2	-1.7
4-CH ₃	46	45.9	0.1	0.1
2, 2'-CH ₃	78	83.7	-5.7	-13.5
2-C ₂ H ₅	70	66.1	3.9	4.7
2- <i>i</i> -C ₃ H ₇	70	68.5	1.5	1.9

^a Taken from ref 20.

where N is the number of cases, R is the correlation coefficient, F is the Fisher ratio of the regression, s is the standard deviation of the regression, and R_{CV} and RMSECV are the correlation coefficient and the root of the mean squared error of the leave-one-out cross-validation, respectively. Observed and calculated values as well as residuals and cross-validation residuals are given in Table 1.

As can be observed this model is not good enough to obtain general conclusions about the possibility of expressing a 3D property in terms of topological (2D) descriptors.

The model is significant and explains almost 93% of the variance in the experimental gas-phase dihedral angles of biphenyls but it does not have a good predictive ability. The greater error in the cross-validation experiment is observed for 2,2'-dimethylbiphenyl (CV-residual = -13.5). It is probably a consequence of the fact that only one bisubstituted biphenyl was included in this data set. Unfortunately, there are no other experimental data of 2,2'-alkylbiphenyl to be included in this model. Consequently, we will try to find another model with the use of dihedral angles calculated from MM+ force field. The use of molecular mechanic force fields calculations for predicting geometrical parameters of organic molecules is widely used when experimental data are not available. In the case of alkylbiphenyl derivatives some previous experience exists with the use of MM2 force field.³⁸ We have computed the dihedral angle Φ for the seven alkylbiphenyls studied experimentally by PES.²⁰ A good linear correlation exists between the experimental angles and those calculated by MM+ which in part justify the use of this type of calculation for estimating Φ :

$$\Phi_{\text{PES}}(\text{deg}) = 44.17 + 0.789\Phi_{\text{MM+}}(\text{deg}) \quad (11)$$

$$N = 7, R = 0.9856, R_{\text{CV}} = 0.9827, s = 2.78, F = 170$$

We selected a series of 78 alkylbiphenyl derivatives with substituents at positions 2 and 2' and carried out full geometry optimization with MM+ force field. The values of the calculated dihedral angle were then described with the use of the local spectral moments of the central bond of such angle. The best model obtained by using linear regression analysis is given below together with its statistical parameters. Predicted, observed values, residuals, and CV-residuals are given in Table 2.

$$\Phi_{\text{MM+}}(\text{deg}) = 89.639(\pm 14.395) - 2.0503(\pm 0.1542)\mu_6(i) + 0.2070(\pm 0.0098)\mu_8(i) \quad (12)$$

$$N = 78, R = 0.9838, R_{\text{CV}} = 0.9833, s = 2.12, F = 1128, \text{RMSECV} = 2.28$$

The existence of model 12 confirms our hypothesis about the possibility of predicting 3D parameters from 2D molecular descriptors. In this case it has been expected that the 3D parameter depends on the interaction between substituents at positions 2 and 2'. This kind of interactions has been well accounted for by the local spectral moments of the bond matrix as can be seen from the statistical parameters of the regression in model 12. In closing, we have demonstrated that "purely" topological (2D) molecular descriptors can describe a 3D structural parameter, such as the dihedral angle between phenyl rings in alkylbiphenyls.

STRUCTURAL INTERPRETATION

The first step in the structural interpretation of the results obtained here is to transform model 12 into contribution of the different structural fragments present in the alkylbiphenyl studied. This transformation is carried out straightforwardly by replacing the values of the spectral moments included into model 12 by their expressions in terms of fragment contributions. That is, we replace $\mu_6(i)$ and $\mu_8(i)$ in model

Table 2. Dihedral Angles between Phenyl Rings Φ in Alkylbiphenyls Calculated by Molecular Mechanics Force Field (MM+) and the Values of Φ Predicted with Local Spectral Moments and Residuals and Cross-Validation Residuals from the Linear Regression

compound	obsd	pred.	res.	CV-res.
biphenyl	1.46	8.48	-7.02	-9.68
2-CH ₃	28.50	26.25	2.25	2.44
2-CH ₃ , 2'-CH ₃	46.07	44.44	1.63	1.70
2-CH ₃ , 2'-C ₂ H ₅	49.75	47.16	2.59	2.68
2-CH ₃ , 2'-n-C ₃ H ₇	49.86	47.36	2.50	2.58
2-CH ₃ , 2'-iso-C ₃ H ₇	53.45	52.35	1.10	1.12
2-CH ₃ , 2'-n-C ₄ H ₉	47.96	47.36	0.60	0.62
2-CH ₃ , 2'-iso-C ₄ H ₉	48.98	47.57	1.41	1.45
2-CH ₃ , 2'-sec-C ₄ H ₉	51.84	52.56	-0.72	-0.73
2-CH ₃ , 2'-tert-C ₄ H ₉	58.52	61.27	-2.75	-2.84
2-CH ₃ , 2'-n-C ₅ H ₁₁	48.36	47.36	1.00	1.03
2-CH ₃ , 2'-iso-C ₅ H ₁₁	47.55	47.36	0.19	0.19
2-CH ₃ , 2'-CH(C ₂ H ₅)C ₂ H ₅	52.47	52.76	-0.29	-0.30
2-C ₂ H ₅	30.12	28.96	1.16	1.24
2-C ₂ H ₅ , 2'-C ₂ H ₅	53.94	49.87	4.07	4.18
2-C ₂ H ₅ , 2'-n-C ₃ H ₇	52.97	50.07	2.90	2.97
2-C ₂ H ₅ , 2'-iso-C ₃ H ₇	55.69	55.06	0.63	0.64
2-C ₂ H ₅ , 2'-n-C ₄ H ₉	49.93	50.07	-0.14	-0.15
2-C ₂ H ₅ , 2'-iso-C ₄ H ₉	50.71	50.28	0.43	0.44
2-C ₂ H ₅ , 2'-sec-C ₄ H ₉	53.65	55.27	-1.62	-1.64
2-C ₂ H ₅ , 2'-tert-C ₄ H ₉	62.82	63.98	-1.16	-1.21
2-C ₂ H ₅ , 2'-n-C ₅ H ₁₁	49.88	50.07	-0.19	-0.20
2-C ₂ H ₅ , 2'-iso-C ₅ H ₁₁	49.44	50.07	-0.63	-0.65
2-C ₂ H ₅ , 2'-CH(C ₂ H ₅)C ₂ H ₅	53.98	55.47	-1.49	-1.52
2-n-C ₃ H ₇	30.64	29.17	1.47	1.58
2-n-C ₃ H ₇ , 2'-n-C ₃ H ₇	53.24	50.28	2.96	3.03
2-n-C ₃ H ₇ , 2'-iso-C ₃ H ₇	56.93	55.27	1.66	1.69
2-n-C ₃ H ₇ , 2'-n-C ₄ H ₉	49.83	50.28	-0.45	-0.46
2-n-C ₃ H ₇ , 2'-iso-C ₄ H ₉	51.63	50.49	1.14	1.17
2-n-C ₃ H ₇ , 2'-sec-C ₄ H ₉	53.19	55.47	-2.28	-2.32
2-n-C ₃ H ₇ , 2'-tert-C ₄ H ₉	63.18	64.19	-1.01	-1.05
2-n-C ₃ H ₇ , 2'-n-C ₅ H ₁₁	49.88	50.28	-0.40	-0.41
2-n-C ₃ H ₇ , 2'-iso-C ₅ H ₁₁	49.45	50.28	-0.83	-0.85
2-n-C ₃ H ₇ , 2'-CH(C ₂ H ₅)C ₂ H ₅	53.55	55.68	-2.13	-2.16
2-iso-C ₃ H ₇	34.01	34.16	-0.15	-0.16
2-iso-C ₃ H ₇ , 2'-iso-C ₃ H ₇	57.69	60.25	-2.56	-2.62
2-iso-C ₃ H ₇ , 2'-n-C ₄ H ₉	52.88	55.27	-2.39	-2.42
2-iso-C ₃ H ₇ , 2'-iso-C ₄ H ₉	53.72	55.47	-1.75	-1.78
2-iso-C ₃ H ₇ , 2'-sec-C ₄ H ₉	57.33	60.46	-3.13	-3.21
2-iso-C ₃ H ₇ , 2'-tert-C ₄ H ₉	68.47	69.17	-0.70	-0.76
2-iso-C ₃ H ₇ , 2'-n-C ₅ H ₁₁	53.43	55.27	-1.84	-1.87
2-iso-C ₃ H ₇ , 2'-iso-C ₅ H ₁₁	53.01	55.27	-2.26	-2.29
2-iso-C ₃ H ₇ , 2'-CH(C ₂ H ₅)C ₂ H ₅	59.60	60.67	-1.07	-1.09
2-n-C ₄ H ₉	30.81	29.17	1.64	1.76
2-n-C ₄ H ₉ , 2'-n-C ₄ H ₉	50.12	50.28	-0.16	-0.16
2-n-C ₄ H ₉ , 2'-iso-C ₄ H ₉	50.93	50.49	0.44	0.45
2-n-C ₄ H ₉ , 2'-sec-C ₄ H ₉	52.89	55.47	-2.58	-2.62
2-n-C ₄ H ₉ , 2'-tert-C ₄ H ₉	68.07	64.19	3.88	4.06
2-n-C ₄ H ₉ , 2'-n-C ₅ H ₁₁	49.75	50.28	-0.53	-0.54
2-n-C ₄ H ₉ , 2'-iso-C ₅ H ₁₁	49.42	50.28	-0.86	-0.88
2-n-C ₄ H ₉ , 2'-CH(C ₂ H ₅)C ₂ H ₅	54.33	55.68	-1.35	-1.37
2-iso-C ₄ H ₉	32.92	29.38	3.54	3.80
2-iso-C ₄ H ₉ , 2'-iso-C ₄ H ₉	52.18	50.69	1.49	1.52
2-iso-C ₄ H ₉ , 2'-sec-C ₄ H ₉	53.87	55.68	-1.81	-1.84
2-iso-C ₄ H ₉ , 2'-tert-C ₄ H ₉	66.33	64.39	1.94	2.03
2-iso-C ₄ H ₉ , 2'-n-C ₅ H ₁₁	50.84	50.49	0.35	0.36
2-iso-C ₄ H ₉ , 2'-iso-C ₅ H ₁₁	50.36	50.49	-0.13	-0.13
2-iso-C ₄ H ₉ , 2'-CH(C ₂ H ₅)C ₂ H ₅	54.84	55.89	-1.05	-1.06
2-sec-C ₄ H ₉	35.01	34.36	0.65	0.69
2-sec-C ₄ H ₉ , 2'-sec-C ₄ H ₉	57.42	60.67	-3.25	-3.33
2-sec-C ₄ H ₉ , 2'-tert-C ₄ H ₉	70.38	69.38	1.00	1.09
2-sec-C ₄ H ₉ , 2'-n-C ₅ H ₁₁	53.35	55.47	-2.12	-2.16
2-sec-C ₄ H ₉ , 2'-iso-C ₅ H ₁₁	53.23	55.47	-2.24	-2.28
2-sec-C ₄ H ₉ , 2'-CH(C ₂ H ₅)C ₂ H ₅	59.10	60.87	-1.77	-1.82
2-tert-C ₄ H ₉	43.00	43.08	-0.08	-0.09
2-tert-C ₄ H ₉ , 2'-tert-C ₄ H ₉	80.59	78.09	2.50	3.11
2-tert-C ₄ H ₉ , 2'-n-C ₅ H ₁₁	68.19	64.19	4.00	4.19
2-tert-C ₄ H ₉ , 2'-iso-C ₅ H ₁₁	69.55	64.19	5.36	5.61
2-tert-C ₄ H ₉ , 2'-CH(C ₂ H ₅)C ₂ H ₅	73.40	69.59	3.81	4.17
2-n-C ₅ H ₁₁	30.78	29.17	1.61	1.73
2-n-C ₅ H ₁₁ , 2'-n-C ₅ H ₁₁	50.48	50.28	0.20	0.21
2-n-C ₅ H ₁₁ , 2'-iso-C ₅ H ₁₁	49.42	50.28	-0.86	-0.88
2-n-C ₅ H ₁₁ , 2'-CH(C ₂ H ₅)C ₂ H ₅	53.09	55.68	-2.59	-2.63
2-iso-C ₅ H ₁₁	30.59	29.17	1.42	1.52
2-iso-C ₅ H ₁₁ , 2'-iso-C ₅ H ₁₁	49.98	50.28	-0.30	-0.31
2-iso-C ₅ H ₁₁ , 2'-CH(C ₂ H ₅)C ₂ H ₅	53.41	55.68	-2.27	-2.31
2-CH(C ₂ H ₅)C ₂ H ₅	34.07	34.57	-0.50	-0.54
2-CH(C ₂ H ₅)C ₂ H ₅ , 2'-CH(C ₂ H ₅)C ₂ H ₅	60.97	61.08	-0.11	-0.11

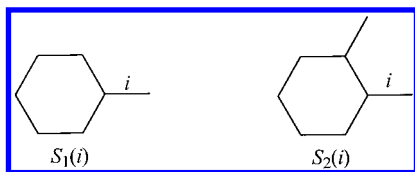


Figure 3. Cyclic fragments present in alkylbiphenyl derivatives and contributing to $\mu_8(i)$.

12 by expressions 7 and 9, and then we sum the similar terms and obtain the following expression:

$$\begin{aligned} \Phi(deg) = & -1.843|F_2(i)| - 23.619|F_3(i)| - \\ & 4.702|F_5^1(i)| - 9.404|F_5^2(i)| - 1.144|F_6^1(i)| - \\ & 3.530|F_6^2(i)| - 1.992|F_6^3(i)| - 1.015|F_7^1(i)| - \\ & 1.824|F_7^2(i)| + 7.115|F_9^1(i)| + 18.370|F_9^2(i)| + \\ & 1.035|F_{10}^1(i)| + 2.070|F_{10}^2(i)| + 1.656|F_{10}^3(i)| + \\ & 0.828|F_{10}^4(i)| + 1.035|F_{11}^1(i)| + 2.484|F_{11}^2(i)| + \\ & 1.242|F_{11}^3(i)| + 0.207|F_{13}^1(i)| + 0.414|F_{13}^2(i)| + \\ & 0.414|F_{13}^3(i)| + 0.828|F_{16}^1(i)| + 1.656|F_{16}^2(i)| + \\ & 0.828|F_{16}^3(i)| + 1.656|F_{16}^4(i)| + 0.828|F_{16}^5(i)| + \\ & 0.828|F_{17}^1(i)| + 1.656|F_{17}^2(i)| + 0.828|S_1(i)| + \\ & 0.414|S_2(i)| \quad (13) \end{aligned}$$

In this expression we have included two terms that were not included in expressions 7 and 9. These terms, $S_1(i)$ and $S_2(i)$, are fragments containing a cycle as illustrated in Figure 3. This is the main reason for which they not appear in expressions 7 and 9 which are only concerned to acyclic fragments. The coefficients in eq 13 represent the contribution of the corresponding fragment.

The expression 13 contains all necessary structural information for calculating the contribution of different groups and substituents in the alkylbiphenyl studied. For instance, a biphenyl framework is decomposed as the sum of 18 fragments giving the following expression:

$$\begin{aligned} C(biphenyl) = & 4F_2(i) + 2F_3(i) + 4F_5^1(i) + 4F_5^2(i) + \\ & 4F_6^1(i) + 4F_6^2(i) + 4F_7^1(i) + 8F_7^2(i) + F_9^2(i) + 4F_{10}^1(i) + \\ & 4F_{10}^2(i) + 8F_{11}^2(i) + 2F_{11}^3(i) + 4F_{13}^1(i) + 8F_{13}^2(i) + \\ & 4F_{13}^3(i) + 4F_{16}^2(i) + 2S_1(i) + 89.64 \quad (14) \end{aligned}$$

To calculate the contribution of biphenyl to the dihedral angle Φ we have to substitute the values of the contributions of each fragment given in eq 13 in the expression 14. For instance, the contribution of one $F_2(i)$ fragment in eq 13 is -1.843 , in biphenyl there are four of these fragments as can be seen from the coefficient of this term in eq 14. Thus, the contribution of this fragment in biphenyl is $-1.843 \cdot 4 = 7.372$. Similarly, we calculate the contributions of the other fragments in eq 14, which result in the total contribution of biphenyl to Φ . This value is $C(biphenyl) = 8.48$, which corresponds with the value of Φ predicted by model 12 (see Table 2).

In a similar way we can obtain the contribution of the different alkyl substituents studied in the current work. The decomposition of these groups in terms of fragments are:

$$\begin{aligned} C(CH_3) = & F_5^1(i) + F_6^1(i) + F_6^3(i) + 2F_7^2(i) + F_9^1(i) + \\ & F_{10}^2(i) + F_{10}^3(i) + F_{11}^1(i) + 2F_{11}^2(i) + F_{11}^3(i) + 2F_{13}^3(i) + \\ & F_{16}^1(i) + F_{16}^2(i) + F_{16}^3(i) + 2F_{16}^4(i) + F_{17}^2(i) + S_2(i) \quad (15) \end{aligned}$$

$$\begin{aligned} C(C_2H_5) = & F_5^1(i) + F_6^1(i) + F_6^3(i) + F_7^1(i) + 2F_7^2(i) + \\ & F_9^1(i) + F_{10}^1(i) + F_{10}^2(i) + 2F_{10}^3(i) + 2F_{11}^1(i) + 2F_{11}^2(i) + \\ & F_{11}^3(i) + 2F_{13}^2(i) + 2F_{13}^3(i) + 2F_{16}^1(i) + F_{16}^2(i) + F_{16}^3(i) + \\ & 2F_{16}^4(i) + F_{17}^2(i) + S_2(i) \quad (16) \end{aligned}$$

$$\begin{aligned} C(n - C_3H_7) = & F_5^1(i) + F_6^1(i) + F_6^3(i) + F_7^1(i) + 2F_7^2(i) + \\ & F_9^1(i) + F_{10}^1(i) + F_{10}^2(i) + 2F_{10}^3(i) + 2F_{11}^1(i) + 2F_{11}^2(i) + \\ & F_{11}^3(i) + F_{13}^1(i) + 2F_{13}^2(i) + 2F_{13}^3(i) + 2F_{16}^1(i) + F_{16}^2(i) + \\ & F_{16}^3(i) + 2F_{16}^4(i) + F_{17}^2(i) + S_2(i) \quad (17) \end{aligned}$$

$$\begin{aligned} C(i - C_3H_7) = & F_5^1(i) + F_6^1(i) + F_6^3(i) + 2F_7^1(i) + 2F_7^2(i) + \\ & F_9^1(i) + 2F_{10}^1(i) + F_{10}^2(i) + 2F_{10}^3(i) + F_{10}^4(i) + 3F_{11}^1(i) + \\ & 2F_{11}^2(i) + F_{11}^3(i) + 4F_{13}^2(i) + 2F_{13}^3(i) + 3F_{16}^1(i) + \\ & F_{16}^2(i) + F_{16}^3(i) + 2F_{16}^4(i) + F_{16}^5(i) + F_{17}^1(i) + \\ & F_{17}^2(i) + S_2(i) \quad (18) \end{aligned}$$

$$\begin{aligned} C(n - C_4H_9) = & F_5^1(i) + F_6^1(i) + F_6^3(i) + F_7^1(i) + 2F_7^2(i) + \\ & F_9^1(i) + F_{10}^1(i) + F_{10}^2(i) + 2F_{10}^3(i) + 2F_{11}^1(i) + 2F_{11}^2(i) + \\ & F_{11}^3(i) + F_{13}^1(i) + 2F_{13}^2(i) + 2F_{13}^3(i) + 2F_{16}^1(i) + F_{16}^2(i) + \\ & F_{16}^3(i) + 2F_{16}^4(i) + F_{17}^2(i) + S_2(i) \quad (19) \end{aligned}$$

$$\begin{aligned} C(i - C_4H_9) = & F_5^1(i) + F_6^1(i) + F_6^3(i) + 2F_7^1(i) + 2F_7^2(i) + \\ & F_9^1(i) + 2F_{10}^1(i) + F_{10}^2(i) + 2F_{10}^3(i) + F_{10}^4(i) + 3F_{11}^1(i) + \\ & 2F_{11}^2(i) + F_{11}^3(i) + F_{13}^1(i) + 4F_{13}^2(i) + 2F_{13}^3(i) + \\ & 3F_{16}^1(i) + F_{16}^2(i) + F_{16}^3(i) + 2F_{16}^4(i) + F_{16}^5(i) + F_{17}^1(i) + \\ & F_{17}^2(i) + S_2(i) \quad (20) \end{aligned}$$

$$\begin{aligned} C(s - C_4H_9) = & F_5^1(i) + F_6^1(i) + F_6^3(i) + F_7^1(i) + 2F_7^2(i) + \\ & F_9^1(i) + F_{10}^1(i) + F_{10}^2(i) + 2F_{10}^3(i) + 2F_{11}^1(i) + 2F_{11}^2(i) + \\ & F_{11}^3(i) + 2F_{13}^1(i) + 2F_{13}^2(i) + 2F_{13}^3(i) + 3F_{16}^1(i) + \\ & F_{16}^2(i) + F_{16}^3(i) + 2F_{16}^4(i) + F_{17}^2(i) + S_2(i) \quad (21) \end{aligned}$$

$$\begin{aligned} C(t - C_4H_9) = & F_5^1(i) + F_6^1(i) + F_6^3(i) + 3F_7^1(i) + 2F_7^2(i) + \\ & F_9^1(i) + 3F_{10}^1(i) + F_{10}^2(i) + 2F_{10}^3(i) + 3F_{10}^4(i) + 4F_{11}^1(i) + \\ & 2F_{11}^2(i) + F_{11}^3(i) + 6F_{13}^2(i) + 2F_{13}^3(i) + F_{14}^4(i) + \\ & 4F_{16}^1(i) + F_{16}^2(i) + F_{16}^3(i) + 2F_{16}^4(i) + 3F_{16}^5(i) + 3F_{17}^1(i) + \\ & F_{17}^2(i) + S_2(i) \quad (22) \end{aligned}$$

$$\begin{aligned} C(n - C_5H_{11}) = & F_5^1(i) + F_6^1(i) + F_6^3(i) + F_7^1(i) + \\ & 2F_7^2(i) + F_9^1(i) + F_{10}^1(i) + F_{10}^2(i) + 2F_{10}^3(i) + 2F_{11}^1(i) + \\ & 2F_{11}^2(i) + F_{11}^3(i) + F_{13}^1(i) + 2F_{13}^2(i) + 2F_{13}^3(i) + \\ & 2F_{16}^1(i) + F_{16}^2(i) + F_{16}^3(i) + 2F_{16}^4(i) + F_{17}^2(i) + S_2(i) \quad (23) \end{aligned}$$

$$C(i - C_5H_{11}) = F_5^1(i) + F_6^1(i) + F_6^3(i) + 2F_7^1(i) + 2F_7^2(i) + F_9^1(i) + 2F_{10}^1(i) + F_{10}^2(i) + 2F_{10}^3(i) + F_{10}^4(i) + 3F_{11}^1(i) + 2F_{11}^2(i) + F_{13}^1(i) + 4F_{13}^2(i) + 2F_{13}^3(i) + 2F_{16}^1(i) + F_{16}^2(i) + F_{16}^3(i) + 2F_{16}^4(i) + F_{16}^5(i) + F_{17}^2(i) + S_2(i) \quad (24)$$

$$C(CH(C_2H_5)_2) = F_5^1(i) + F_6^1(i) + F_6^3(i) + 2F_7^1(i) + 2F_7^2(i) + F_9^1(i) + 2F_{10}^1(i) + F_{10}^2(i) + 2F_{10}^3(i) + F_{10}^4(i) + 3F_{11}^1(i) + 2F_{11}^2(i) + F_{11}^3(i) + 2F_{13}^1(i) + 4F_{13}^2(i) + 2F_{13}^3(i) + 3F_{16}^1(i) + F_{16}^2(i) + F_{16}^3(i) + 2F_{16}^4(i) + F_{16}^5(i) + F_{17}^1(i) + F_{17}^2(i) + S_2(i) \quad (25)$$

By using these expressions and the coefficients of eq 13 we can calculate the contributions of these substituents at positions 2 or 2' of the biphenyl to the dihedral angle Φ . These contributions are given in Table 3, and they represent a sort of substituent constants based on the dihedral angle between phenyl rings in alkylbiphenyls. The term $I(2,2')$ given in Table 3 represents the interaction between any two substituents in positions 2 and 2'. It arises from the contribution of the term $F_{13}^3(i)$ that is present when substituents appear in such positions.

These substituent constants can be used to estimate the dihedral angle of alkylbiphenyl derivatives. For instance, the angle Φ in 2,2'-dimethylbiphenyl is calculated as the sum of the contribution of biphenyl plus two times the contribution of methyl plus the interaction term $I(2,2')$: $(8.48 \div 2 * 17.17 + 0.414 = 44.44)$. By these means we can express the dihedral angle Φ as an additive function in terms of the contributions C_i and the interaction term $I(2,2')$:

$$\Phi_{MM+}(\text{deg}) = 8.48 + \sum_i C_i + I(2,2') \quad (26)$$

The substituent constants found in the current work are a sort of "steric" parameter. In fact, they reflect the repulsion between substituents at positions 2 and 2' of biphenyl due to steric hinder that obligates phenyl rings to be not coplanar. However, these new substituent constants are not linearly related to other steric constants, such as the Taft's steric constants $-E_s$.³⁹ For instance, the linear correlation coefficient between C_i and $-E_s$ is only 0.752, which indicates that both constants account for different structural features of the substituents. In fact, only a modest correlation can be obtained to describe the angle Φ with the sum of Taft's steric constants $-E_s$ of alkyl substituents at positions 2 and 2':

$$\Phi_{MM+}(\text{deg}) = 21.119(\pm 2.317) - 8.4889(\pm 0.622) \sum_i E_{S_i} \quad (27)$$

$$N = 78, R = 0.8427, s = 6.32, F = 186.2$$

As can be seen the topological descriptors used here to account for the 3D parameter studied describe in a more appropriate way than the Taft's steric constants $-E_s$ the interactions of alkyl groups in the biphenyl framework. In fact, the model based on graph-theoretical (topological)

Table 3. Group Contributions Calculated from Local Spectral Moments for the Alkyl Substituents in Biphenyls Studied Here and the Taft's Steric Constants $-E_s$

R	C	$-E_s$	R	C	$-E_s$
CH ₃	17.77	1.24	<i>s</i> -C ₄ H ₉	25.88	2.37
C ₂ H ₅	20.48	1.31	<i>t</i> -C ₄ H ₉	34.60	2.78
<i>n</i> -C ₃ H ₇	20.69	1.6	<i>n</i> -C ₅ H ₁₁	20.69	1.64
<i>i</i> -C ₃ H ₇	25.68	1.71	<i>i</i> -C ₅ H ₁₁	20.69	1.59
<i>n</i> -C ₄ H ₉	20.69	1.63	CH(C ₂ H ₅) ₂	26.09	3.22
<i>i</i> -C ₄ H ₉	20.90	2.17	<i>I</i> (2,2')	0.414	

descriptors explains 25.8% of the variance in Φ more than the Taft's steric constants.

CONCLUSIONS

We have proved that 2D molecular descriptors, such as those derived from a graph-theoretical representation of molecules, can be directly used to describe 3D structural parameters of molecules. This apparently ingenuous conclusion can have important consequences for the interpretation of 2D (topological) molecular descriptors in QSPR/QSAR studies. For instance, suppose that we are now modeling an experimental property of a subset of biphenyls studied here. This (physicochemical or biological) property is expressed in terms of 2D (topological) molecular descriptors, such as those used here to describe the dihedral angle Φ . In fact, we can obtain a model of the form, $P = b + \sum_k a_k \mu_k(i)$, where $\mu_k(i)$ are the local spectral moments of the bond matrix. After the statistical validation of this QSPR/QSAR model the following question immediately arises: Does this model take into account the 3D structural features of biphenyls that can have an important influence on the physicochemical or biological property P under study. It is clear that the descriptors we are including in this model are 2D (topological), but we have previously shown that they are able to describe at least one of the most important 3D features of these molecules. So, the current QSPR/QSAR model contains "implicitly" a description of these 3D features of biphenyls, and to call it a 2D-QSPR/QSAR model does not reflect its real nature.

Another important point that the current work shows is related to the possibilities of graph-based molecular descriptors. We previously shown that TOPS-MODE descriptors can be transformed into substituent constants of different nature, such as pharmacological activity contributions, toxicity constants, electronic constants, and now as steric constants. These features shown the versatility of graph-theoretically derived descriptors for solving many of the problems actually under study in QSPR and QSAR research.

ACKNOWLEDGMENT

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

REFERENCES AND NOTES

- (1) Randic, M. The nature of chemical structure. *J. Math. Chem.* **1990**, 4, 157-184.

- (2) Randic, M. Chemical structure-What is "she"? *J. Chem. Educ.* **1992**, 69, 713–718.
- (3) Turro, N. J. Geometric and topological thinking in organic chemistry. *Angew. Chem., Int. Ed. Engl.* **1986**, 25, 882–901.
- (4) Gutman, I.; Polansky, O. E. *Mathematical Concepts in Organic Chemistry*; Springer-Verlag: Berlin, 1987.
- (5) Burkert, U.; Allinger, N. L. *Molecular Mechanics*; American Chemical Society: Washington, 1982.
- (6) Helgaker, T.; Jorgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; Wiley: Chichester, 2000.
- (7) Balaban, A. T.; Bonchev, D.; Seitz, W. A. Topological/chemical distances and graph centres in molecular graphs with multiple bonds. *J. Mol. Struct. (Theochem)* **1993**, 280, 253–260.
- (8) Jorgensen, W. L.; Salem, L. *The Organic Chemist's Book of Orbitals*; Academic Press: New York, 1973.
- (9) Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **1973**, 181, 223–230.
- (10) Fersht, A. *Structure and Mechanism in Protein Science*; W. H. Freeman and Co.: New York, 1998; p 4.
- (11) Katritzky, A. R.; Gordeeva, E. V. Traditional topological indices vs electronic, geometrical, and combined molecular descriptors in QSAR/QSPR research. *J. Chem. Inf. Comput. Sci.* **1993**, 33, 835–857.
- (12) Randic, M. Topological indices. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Ed.; John Wiley & Sons: New York, 1998; Vol. 5, pp 3018–3032.
- (13) Balaban, A. T.; Ivanciuc, O. Historical developments of topological indices. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach: The Netherlands, 1999; pp 21–57.
- (14) Rappé, A. K.; Casewit, C. J. *Molecular Mechanics Across Chemistry*; University Sci. Books: Herndon, VA, 1997.
- (15) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Analysis*; Research Studies Press: Letchworth, 1986.
- (16) 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.
- (17) Estrada, E.; Uriarte, E. Recent advances on the role of topological indices in drug discovery research. *Curr. Med. Chem.* **2000**, in press.
- (18) Petitjean, M. Applications of the radius-diameter diagram to the classification of topological and geometric shapes of chemical compounds. *J. Chem. Inf. Comput. Sci.* **1992**, 32, 331–337.
- (19) Bath, P. A.; Poirrette, A. R.; Willett, P. The extent of the relationship between the graph-theoretical and the geometrical shape coefficient of chemical compounds. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 714–716.
- (20) Maier, J. P.; Turner, D. W. Steric inhibition of resonance studied by molecular photoelectron spectroscopy. 1. Biphenyls. *Faraday Discuss. Chem. Soc.* **1972**, 149–67.
- (21) *Hyperchem 3.0 for Windows*; Hypercube, Inc., 1993.
- (22) Estrada, E.; Molina, E. Novel local (fragment-based) topological molecular descriptors for QSPR/QSAR and molecular design. *J. Mol. Graph. Modell.* **2001**, in press.
- (23) Estrada, E.; Molina, E. QSPR/QSAR by graph theoretical descriptors beyond the frontiers. In *QSAR/QSPR Studies by Molecular Descriptors*; Diudea, M., Ed.; Nova Science: New York, 2000; pp 83–107.
- (24) Originally introduced as TOSS-MODE it has been recently renamed to TOPS-MODE: Estrada, E.; Uriarte, E. Quantitative structure-toxicity relationships using TOPS-MODE. 1. Nitrobenzene toxicity to *Tetrahymena pyriformis*. *SAR QSAR Environ. Res.* **2000**, in press.
- (25) Estrada, E. Edge adjacency relationships and a novel topological index related to molecular volume. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 31–33.
- (26) Estrada, E. Novel strategies in the search of topological indices. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach: Amsterdam, The Netherlands, 1999; pp 403–453.
- (27) Estrada, E. Spectral moments of the edge adjacency matrix. 1. Definition and applications to the prediction of physical properties of alkanes. *J. Chem. Inf. Comput. Sci.* **1996**, 36, 844–849.
- (28) Estrada, E. Spectral moments of the edge adjacency matrix. 2. Molecules containing heteroatoms and QSAR applications. *J. Chem. Inf. Comput. Sci.* **1997**, 37, 320–327.
- (29) Estrada, E. Spectral moments of the edge adjacency matrix. 3. Molecules containing cycles. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 23–27.
- (30) Estrada, E. Modelling the diamagnetic susceptibility of organic compounds by a sub-structural graph-theoretical approach. *J. Chem. Soc., Faraday Trans.* **1998**, 94, 1407–1410.
- (31) 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.
- (32) Estrada, E.; Gutierrez, Y. Modeling chromatographic parameters by a novel graph theoretical sub-structural approach. *J. Chromatogr. A* **1999**, 858, 187–199.
- (33) 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.
- (34) Estrada, E.; Uriarte, E.; Montero, A.; Teixeira, 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.
- (35) Estrada, E.; Peña, A. *In silico* studies for the rational discovery of anticonvulsant compounds. *Bioorg. Med. Chem.* **2000**, 8, 2755–2770.
- (36) 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.
- (37) Gutierrez, Y.; Estrada, E.; Muñoz, I. TOSS-MODE for Windows, version 3.0; Universidad de Santiago de Compostela: Spain, 1999.
- (38) Tsuzuki, S.; Tanabe, K.; Nagawa, Y.; Nakanishi, H.; Osawa, E. Calculations of structures of biphenyl and alkylbiphenyls by molecular mechanics. *J. Mol. Struct. (Theochem)* **1988**, 178, 277–285.
- (39) Taft, R. W. separation of polar, steric, and resonance effects in reactivity. In *Steric Effects in Organic Chemistry*; Newman, M. S., Ed.; Wiley Publishers: Dordrecht, 1956; pp 556–675.

CI000170V