Evaluation in Quantitative Structure−Property Relationship Models of Structural Descriptors Derived from Information-Theory Operators

Ovidiu Ivanciuc,*[†,‡] Teodora Ivanciuc,[†] Daniel Cabrol-Bass,*[§,||] and Alexandru T. Balaban*[†,⊥]

Department of Organic Chemistry, Faculty of Chemical Technology, University "Politehnica" of Bucharest, Oficiul 12 CP 243, 78100 Bucharest, Romania, and GRECFO-LARTIC, University of Nice−Sophia Antipolis, Parc Valrose, 06108 Nice Cedex 2, France

Received July 30, 1999

During the search for new structural descriptors we have defined the information-theory operators **U**(**M**), **V**(**M**), **X**(**M**), and **Y**(**M**), that are computed from atomic invariants and measure the information content of the elements of molecular matrices. Structural descriptors computed with these four information-theory operators are used to develop structure−property models for the boiling temperature, molar heat capacity, standard Gibbs energy of formation, vaporization enthalpy, refractive index, and density of alkanes. The information-theory operators were applied to six molecular matrices, namely, the distance **D**, the reciprocal distance **RD**, the distance-path $\mathbf{D_p}$, the reciprocal distance-path $\mathbf{RD_p}$, the path Szeged $\mathbf{Sz_p}$, and the reciprocal path Szeged $\mathbf{RSz_p}$ matrices. In combination with other topological indices, the information-theory indices offer good structure−property models for all six alkane properties investigated in this study.

INTRODUCTION

The properties of a substance (such as the physicochemical behavior, reactivity, or biological activity) are ultimately determined by its molecular structure. Quantitative structure−property relationship (QSPR) and quantitative structure−activity relationship (QSAR) models represent well-established tools for the molecular design of new compounds with desired properties. All QSPR and QSAR models are statistically based and aimed at extracting the maximum information from experimental data on compounds of known structure. These structure−property studies require atomic and molecular descriptors to encode in a numerical form the local (atomic) or global (molecular) structure. The most efficient software used in QSPR or QSAR studies integrates the computation of structural descriptors with the generation of structure−property models. Several programs from this category, such as ADAPT,[1−5] OASIS,[6−8] SciQSAR,[9] CODESSA,[10−16] and Cerius,[2,17] were used with success in developing a large number of QSPR and QSAR models. These programs compute more than 1000 structural descriptors from five classes: constitutional, graph theoretic, and topological indices and geometrical, electrostatic, and quantum-chemical descriptors. Using statistical methods, such as multilinear regression (MLR), principal component analysis (PCA), partial least square (PLS), or artificial neural networks, the best descriptors are selected in the final structure−property model. A survey of the QSPR and QSAR models developed with the above programs shows that molecular graph descriptors and topological indices are used with success to model various properties and demonstrates that they are valuable descriptors of chemical structure. The interest of developing new graph descriptors for organic compounds revived in recent years, when topological indices found new applications in database mining, similarity, and diversity assessment.

Graph theory is widely used in theoretical chemistry, and numerous reviews[18−29] were published on its applications in characterizing the molecular structure with graph descriptors and topological indices. In a first approximation the chemical structure of a molecule can be represented as a molecular graph. Molecular graphs are nondirected chemical graphs that represent molecules, using different conventions. Usually, only non-hydrogen atoms are taken into account in molecular graphs. In the graph representation of molecules, their geometrical features such as bond lengths or bond angles, are not taken into account and the chemical bonding of atoms is regarded as being their most important characteristic. In molecular graphs, vertexes correspond to atoms and edges represent covalent bonds between atoms. A topological index (TI) is a numerical descriptor of the molecular structure based on certain topological features of the molecular graph, offering an effective way of measuring molecular branching, shape, size, and molecular similarity. Topological indices have several obvious advantages when compared with geometrical, electrostatic, and quantum descriptors: they are computed only from the information contained in the molecular graph, they have a unique value for a particular chemical compound, and their calculation requires small computational resources. Considering the advantages of graph descriptors, TIs represent valuable descriptors that complement (and do not substitute) the structural information encoded by other classes of descriptors. Because TIs are global descriptors of the molecular graph, they do not contain explicit information regarding the number of functional groups, pharmacophores, volume, surface area, interatomic distances, charge distribution, orbital energy, or electrostatic potential; for the generation of QSPR and QSAR models such information must be provided by other structural descriptors.

Using information theory applied to graph distances at the atomic (vertex) level, the highly discriminating topological

* Corresponding author.
[†] University "Politehnica" of Bucharest.
[‡] E-mail: o_ivanciuc@chim.upb.ro.
[§] University of Nice−Sophia Antipolis.
[||] E-mail: cabrol@unice.fr.
[⊥] E-mail: atbalaban@hotmail.com.

10.1021/ci9900884 CCC: $19.00
Published on Web 01/12/2000

indices *U*, *V*, *X*, and *Y* were introduced.[30,31] These four information indices were used with good results in QSPR models to estimate the critical volumes, temperatures, and pressures of 49 alkanes, as well as for the coefficients of the Antoine equation that give the saturation pressure versus the absolute temperatures.[32] Although by design the "information-on-distances" indices *U*, *V*, *X*, and *Y* are very selective, it was found that certain acyclic[33] and cyclic[34] graphs that have nonequivalent vertexes with identical distance degree sequences can generate pairs of graphs with degenerate *U*, *V*, *X*, and *Y* indices. Recently, these information indices were extended for any symmetric molecular matrix derived from vertex- and edge-weighted molecular graphs, giving the operators **U**(**M**), **V**(**M**), **X**(**M**), and **Y**(**M**) that are a measure of the information on matrix elements.[35] These four operators were also applied to molecular matrices derived from three-dimensional molecular geometry, giving topological−geometrical indices that were used with success in a quantitative structure−retention relationship model for 50 alkylphenols in gas−liquid chromatography.[36]

In the present study we will investigate the correlational ability of a large collection of TIs computed from the information-theory operators **U**(**M**), **V**(**M**), **X**(**M**), and **Y**(**M**). Structural descriptors computed with these four information-theory operators are used to develop structure−property models for the boiling temperature, molar heat capacity, standard Gibbs energy of formation, vaporization enthalpy, refractive index, and density of alkanes. The information-theory operators are computed from six molecular matrices, namely, the distance **D**, the reciprocal distance **RD**, the distance-path $\mathbf{D_p}$, the reciprocal distance-path $\mathbf{RD_p}$, the path Szeged $\mathbf{Sz_p}$, and the reciprocal path Szeged $\mathbf{RSz_p}$ matrices.

INFORMATION-THEORY OPERATORS

The indices *U*, *V*, *X*, and *Y* for information on distances are computed from the elements of the distance matrix of the molecular graph, and these TIs provided good results both for structure discrimination and in structure−property models.[30−32] Because new graph matrices were defined in recent years,[28] it is possible to extend the definition of these four indices for all dense molecular matrices **M** (a dense matrix is a matrix without zero nondiagonal elements). We have recently introduced four information-theory operators that can be applied to a matrix with integer value elements, such as the distance matrix **D**, or to a matrix with real value elements, such as the reciprocal distance matrix **RD**. The graph vertex operators **VUinf**(**M**,*G*), **VVinf**(**M**,*G*), **VXinf**(**M**,*G*), and **VYinf**(**M**,*G*) apply the information-theory equations to the nonzero elements of the molecular matrix **M** that correspond to a vertex v_i:

$$\mathbf{VUinf}(\mathbf{M})_i = -\sum_{j=1}^{N} \frac{[\mathbf{M}]_{ij}}{\mathbf{VS}(\mathbf{M})_i} \log_2\left[\frac{[\mathbf{M}]_{ij}}{\mathbf{VS}(\mathbf{M})_i}\right] \qquad (1)$$

$$\mathbf{VVinf}(\mathbf{M})_i = \mathbf{VS}(\mathbf{M})_i \log_2 \mathbf{VS}(\mathbf{M})_i - \mathbf{VUinf}(\mathbf{M})_i \qquad (2)$$

$$\mathbf{VXinf}(\mathbf{M})_i = \mathbf{VS}(\mathbf{M})_i \log_2 \mathbf{VS}(\mathbf{M})_i - \mathbf{VYinf}(\mathbf{M})_i \qquad (3)$$

$$\mathbf{VYinf}(\mathbf{M})_i = \sum_{j=1}^{N} [\mathbf{M}]_{ij} \log_2 [\mathbf{M}]_{ij} \qquad (4)$$

where **M** is a molecular graph matrix, $\mathbf{VS}(\mathbf{M})_i$ represents the vertex sum of the vertex v_i, and the summations in eqs 1 and 4 are done for the nonzero elements of the molecular matrix **M**, $[\mathbf{M}]_{ij} \neq 0$. In a molecular graph *G* with *N* vertexes, the vertex sum operator for the vertex v_i, $\mathbf{VS}(\mathbf{M},G)_i$, is defined as the sum of the elements in column *i*, or row *i*, of the molecular matrix **M**:

$$\mathbf{VS}(\mathbf{M}, G)_i = \sum_{j=1}^{N} [\mathbf{M}]_{ij} = \sum_{j=1}^{N} [\mathbf{M}]_{ij} \qquad (5)$$

For a general dense molecular graph matrix **M**, the matrix elements $[\mathbf{M}]_{ij}$ may have values lower than 1, giving negative terms for certain vertex structural descriptors computed with the graph vertex operators **VUinf**(**M**,*G*), **VVinf**(**M**,*G*), **VXinf**(**M**,*G*), and **VYinf**(**M**,*G*). The Randić-like formula used in the case of the indices *U*, *V*, *X*, and *Y* is therefore replaced by the following equation:

$$f(x, y) = \begin{cases} (xy)^{-1/2} & \text{if } xy > 0 \\ -(|xy|)^{-1/2} & \text{if } xy < 0 \end{cases} \qquad (6)$$

The operators **U**(**M**), **V**(**M**), **X**(**M**), and **Y**(**M**), representing information on matrix elements, are computed with the equations

$$\mathbf{U}(\mathbf{M}, G) = \frac{M}{\mu + 1} \sum_{E(G)} f\,(\mathbf{VUinf}(\mathbf{M})_i, \mathbf{VUinf}(\mathbf{M})_j) \qquad (7)$$

$$\mathbf{V}(\mathbf{M}, G) = \frac{M}{\mu + 1} \sum_{E(G)} f\,(\mathbf{VVinf}(\mathbf{M})_i, \mathbf{VVinf}(\mathbf{M})_j) \qquad (8)$$

$$\mathbf{X}(\mathbf{M}, G) = \frac{M}{\mu + 1} \sum_{E(G)} f\,(\mathbf{VXinf}(\mathbf{M})_i, \mathbf{VXinf}(\mathbf{M})_j) \qquad (9)$$

$$\mathbf{Y}(\mathbf{M}, G) = \frac{M}{\mu + 1} \sum_{E(G)} f\,(\mathbf{VYinf}(\mathbf{M})_i, \mathbf{VYinf}(\mathbf{M})_j) \qquad (10)$$

METHOD

(1) Data. The QSPR models were developed for a data set consisting of 134 alkanes between C_6 and C_{10}, for the following six physical properties:[37] t_b, boiling temperature at normal pressure (°C); C_p, molar heat capacity at 300 K (J K^{-1} mol^{-1}); $\Delta_f G°_{300}$ (g), standard Gibbs energy of formation in the gas phase at 300 K (kJ mol^{-1}); $\Delta_{vap}H_{300}$, vaporization enthalpy at 300 K (kJ mol^{-1}); n_D^{25}, refractive index at 25 °C; ρ, density at 25 °C (kg m^{-3}). The value of the refractive index of 2,2,3,3-tetramethylbutane is missing, while the reported density of this compound, 821.70 kg m^{-3},[37] is too high when compared with the density of similar alkanes and it was not considered in the computation of the density QSPR models. As it is known, there are 142 constitutional isomers for these alkanes, but data for all six properties are missing for the following eight of them: *n*-hexane, *n*-nonane, *n*-decane, 2-methylnonane, 3-methylnonane, 4-methylnonane, 5-methylnonane, and 3-ethyl-2,4-dimethylhexane. This database of 134 alkanes together with their six physical properties forms the Alkane1 database proposed as a standard QSPR test for structural descriptors.[38] The utility of structural descriptors may be judged by comparing their performances in computing the alkane physical properties from the Alkane1 database.

(2) Molecular Matrices. The structural descriptors defined on molecular matrices were computed from the following six graph matrices: distance **D**, reciprocal distance **RD**,[39−43] distance-path $\mathbf{D_p}$,[44,45] reciprocal distance-path $\mathbf{RD_p}$,[44,45] path Szeged $\mathbf{Sz_p}$,[46−49] and reciprocal path Szeged $\mathbf{RSz_p}$.[46−49]

(3) Structural Descriptors. The scope of this paper is to investigate the correlational ability of the information-theory operators **U**(**M**), **V**(**M**), **X**(**M**), and **Y**(**M**), computed from six molecular matrices. Owing to the complexity of the molecular structure, it seems to be impossible to expect that a single set of descriptors would contain all the relevant information. This is the main reason QSPR and QSAR models are developed by selecting descriptors from a large pool of structural descriptors. In the present study the QSPR models were generated with the information-theory indices and a selection of the most used molecular graph descriptors. The list of the 60 structural descriptors used in the QSPR study is presented below:

(1) The molecular weight, **MW**: This constitutional descriptor is a measure of molecular size, and experience shows that in many structure−property studies this is an important parameter.

(2) The Kier and Hall connectivity indices[18] ${}^0\chi$, ${}^1\chi$, ${}^2\chi$, ${}^3\chi_p$, ${}^3\chi_c$: These topological indices are, by far, the most used descriptors in QSPR and QSAR models. The ${}^1\chi$ index was first described by Randić and named molecular connectivity.[19]

(3) Wiener indices computed with the Wiener operator **Wi**(**M**):[25−28,50] These descriptors represent an extension of the Wiener index that is defined for any molecular matrix. Molecular graph descriptors computed with the Wiener operator were used with success to compute the boiling temperatures of acyclic compounds containing oxygen or sulfur atoms.[51] The Wiener operator $\mathbf{Wi}(\mathbf{M}) = \mathbf{Wi}(\mathbf{M},G)$ of a molecular graph G with N vertexes is computed from the symmetric $N \times N$ molecular matrix $\mathbf{M} = \mathbf{M}(G)$:

$$\mathbf{Wi}(\mathbf{M},G) = \sum_{j=1}^{N}\sum_{j=1}^{N}[\mathbf{M}(G)]_{ij} \quad (11)$$

The Wiener operator **Wi**(**M**) is an extension of the topological index W introduced by Wiener[52,53] for alkanes, and extended to cycloalkanes by Hosoya.[54] While W is computed from the distance matrix, the Wiener operator **Wi**(**M**) can be applied to any molecular matrix, derived either from the molecular graph or from the three-dimensional structure of a chemical compound. A detailed comparison of these structural descriptors was made by Ivanciuc.[27]

(4) Hyper-Wiener indices computed with the hyper-Wiener operator **HyWi**(**M**):[27,50,51] The hyper-Wiener operator $\mathbf{HyWi}(\mathbf{M}) = \mathbf{HyWi}(\mathbf{M},G)$ of a molecular graph G with N vertexes is computed from the symmetric $N \times N$ molecular matrix $\mathbf{M} = \mathbf{M}(G)$:

$$\mathbf{HyWi}(\mathbf{M},G) = \frac{1}{2}\sum_{i=1}^{N}\sum_{i=1}^{N}([\mathbf{M}]_{ij}^{\,2} + [\mathbf{M}]_{ij}) \quad (12)$$

The hyper-Wiener index WW was defined for alkanes by Randić[55−57] and extended to cycloalkanes by Klein, Lukovits, and Gutman.[58] Diudea proposed an alternative method for computing the hyper-Wiener from the distance-path matrix $\mathbf{D_p}$.[44,45] Equation 12, related to the formula proposed in ref 58, extends the computation of the hyper-Wiener indices to molecular graph matrices or matrices derived from the three-dimensional molecular structure. The definitions and examples for the computation of all these descriptors were recently reviewed.[27]

(5) The spectral operators MinSp(M) and MaxSp(M):[25−27,50] Structural descriptors computed with these two spectral operators were used with good results to develop QSPR models for the boiling temperature, vaporization enthalpy, molar refraction, molar volume, critical pressure, critical temperature, and surface tension of alkanes,[43,59] to estimate the boiling temperature of acyclic compounds containing oxygen or sulfur atoms,[51] and to model the amine boiling temperature.[60] The matrix spectrum operator[50] $\mathbf{Sp}(\mathbf{M},G) = \{x_i, i = 1, 2, ..., N\}$ represents the eigenvalues of the matrix **M** or the roots of the polynomial $\mathbf{Ch}(\mathbf{M},G,x)$, $\mathbf{Ch}(\mathbf{M},G,x) = 0$. The spectral operators $\mathbf{MinSp}(\mathbf{M},G)$ and $\mathbf{MaxSp}(\mathbf{M},G)$ are equal to the minimum and maximum values of $\mathbf{Sp}(\mathbf{M},G)$, respectively:[45]

$$\mathbf{MinSp}(\mathbf{M},G) = \mathbf{min}\{\mathbf{Sp}(\mathbf{M},G)\} \quad (13)$$

$$\mathbf{MaxSp}(\mathbf{M},G) = \mathbf{max}\{\mathbf{Sp}(\mathbf{M},G)\} \quad (14)$$

(6) The Ivanciuc−Balaban operator **IB**(**M**):[28,51] By design, this operator reflects mainly the molecular shape and reduces the contribution of the size term in the numerical value of the descriptor. Molecular graph descriptors computed with the Ivanciuc−Balaban operator were used with success to generate structure−property models for the boiling temperature of acyclic compounds containing oxygen or sulfur atoms[51] and to estimate the boiling temperature, molar heat capacity, standard Gibbs energy of formation, vaporization enthalpy, refractive index, and density of alkanes.[61] The Ivanciuc−Balaban operator of a graph G, $\mathbf{IB}(\mathbf{M}) = \mathbf{IB}(\mathbf{M},G)$ is computed from the vertex sum local invariant defined by eq 5 computed from the symmetric $N \times N$ molecular matrix $\mathbf{M} = \mathbf{M}(G)$:

$$\mathbf{IB}(\mathbf{M},G) = \frac{M}{\mu + 1}\sum_{e_{ij}\in E(G)}(\mathbf{VS}(\mathbf{M})_i \times \mathbf{VS}(\mathbf{M})_j)^{-1/2} \quad (15)$$

where $\mathbf{VS}(\mathbf{M})_i$ and $\mathbf{VS}(\mathbf{M})_j$ denote the vertex sums of the two adjacent vertexes v_i and v_j that are incident with an edge e_{ij} in the molecular graph G, M is the number of edges in the molecular graph, μ is the cyclomatic number (the number of cycles in the graph, $\mu = M - N + 1$, where N is the number of atoms the molecular graph), and the summation goes over all edges from the edge set $E(G)$.

(7) The information-theory operators **U**(**M**), **V**(**M**), **X**(**M**), and **Y**(**M**).[35]

(4) QSPR Model. All studies that develop QSPR models from a large set of computed descriptors use a wide range of algorithms for selecting significant descriptors. Because an exhaustive test of all multilinear regression (MLR) equations require too large computational resources, we have used a heuristic method for descriptor selection. This heuristic algorithm starts from the set of 60 structural descriptors and develops QSPR models by applying the following steps:

(1) All one-parameter correlation equations are computed. All descriptors with a correlation coefficient greater than a threshold $|r_{\min}| > 0.15$ are selected for further use.

(2) Biparametric regression equations are computed with all possible pairs of descriptors selected in step 1 that are not significantly correlated. Two descriptors are considered to be not significantly correlated if their intercorrelation coefficient r_{ij} is lower than a threshold $|r_{ij}| < 0.8$. The most significant 60 pairs of molecular descriptors were used in the next (third) step.

(3) To an MLR model containing n descriptors, a new descriptor is added to generate a model with $n + 1$ descriptors if the new descriptor is not significantly correlated with the previous n descriptors.

(4) The most significant 60 MLR models containing $n + 1$ descriptors are selected.

Steps 3 and 4 are repeated until MLR models with a certain maximum number of descriptors are obtained.

RESULTS AND DISCUSSION

(1) Normal Boiling Temperature. In Tables 1 and 2 we present the coefficients, confidence interval, structural descriptors, and statistical indices for the best ten MLR equations with three and four, respectively, independent variables that model the alkane boiling temperature. The best MLR equation with three independent variables contains a connectivity index, ${}^3\chi_p$, the spectral descriptor **MaxSp**(**RD**), and the index **V**(**RD**). The index ${}^3\chi_p$ represents the weighted contribution of butane-like subgraphs and is a measure of molecular branching, and **MaxSp**(**RD**) is a global measure of branching derived from the reciprocal distance matrix, while **V**(**RD**) represents the weighted total information content in the elements of the reciprocal distance matrix. All ten MLR equations from Table 1 contain the connectivity index ${}^3\chi_p$, and an information index, namely, **V**(**RD**), **V**(**D**), **X**(**RD**), **X**($\mathbf{D_p}$), or **X**($\mathbf{RD_p}$). The third index included by the descriptor selection algorithm either is the molecular weight **MW** or is derived from the Wiener, hyper-Wiener, or **MaxSp** operators. The majority of the indices are derived from reciprocal matrices, such as the reciprocal distance **RD** and the reciprocal distance-path $\mathbf{RD_p}$ matrices.

The addition of the fourth independent variable improves the MLR models for the alkane boiling temperature, as can be seen from Table 2. The best MLR equation with four independent variables contains, apart from the indices from eq 1, the information index **V**($\mathbf{Sz_p}$) computed from the elements of the path Szeged matrix $\mathbf{Sz_p}$; eq 11, with $r = 0.9951$, $s = 2.67$, and $F = 3256$, is only slightly better than the remaining nine equations (12)−(20); this situation appears when a QSPR model is derived from a large set of descriptors and indicates that several MLR models with close statistical indices can be generated with different sets of topological indices. We have to mention that almost all indices are computed from **RD** and $\mathbf{RD_p}$ matrices, but one finds also five information indices derived from $\mathbf{Sz_p}$, indicating the importance of the path Szeged matrix in generating TIs with good correlational ability. The connectivity index ${}^3\chi_p$ is present in all eqs 11−20; seven equations contain two information indices, while the remaining three have one information index. Equally important is the spectral descriptor **MaxSp**(**RD**) that appears in five equations.

Table 1. Coefficients, Confidence Interval, Structural Descriptors $\mathbf{SD}_i$ ($i = 1-3$), and Statistical Indices for the Best 10 MLR Equations with Three Independent Variables That Model the Alkane Boiling Temperature at Normal Pressure, t_b (°C)[a]

eq	a_0	a_1	$\mathbf{SD}_1$	a_2	$\mathbf{SD}_2$	a_3	$\mathbf{SD}_3$	r	s	F
1	−146.782 ± 8.218	8.8334 ± 0.4946	${}^3\chi_p$	45.479 ± 2.546	**MaxSp(RD)**	6.5352 ± 0.3659	**V(RD)**	0.9939	2.97	3495
2	−46.764 ± 2.802	11.514 ± 0.690	${}^3\chi_p$	9.0677 ± 0.5434	**HyWi($\mathbf{RD_p}$)**	4.5813 ± 0.2745	**V(RD)**	0.9930	3.18	3051
3	−41.997 ± 2.542	11.145 ± 0.675	${}^3\chi_p$	7.1334 ± 0.432	**Wi($\mathbf{RD_p}$)**	4.8711 ± 0.2949	**V(RD)**	0.9928	3.21	2990
4	−75.612 ± 4.637	1.1880 ± 0.0729	MW	12.531 ± 0.769	${}^3\chi_p$	2.7534 ± 0.1689	**V(RD)**	0.9926	3.25	2913
5	−34.714 ± 2.184	11.313 ± 0.712	${}^3\chi_p$	6.1756 ± 0.3886	**HyWi(RD)**	4.5683 ± 0.2875	**V(RD)**	0.9923	3.33	2767
6	−87.987 ± 5.578	12.424 ± 0.788	${}^3\chi_p$	21.646 ± 1.372	**MaxSp(RD)**	20.088 ± 1.274	**X(RD)**	0.9921	3.36	2726
7	−22.157 ± 1.421	1.2379 ± 0.0794	MW	15.319 ± 0.983	${}^3\chi_p$	−39.115 ± 2.509	**V(D)**	0.9920	3.40	2662
8	−18.113 ± 1.169	1.1641 ± 0.0751	MW	15.361 ± 0.992	${}^3\chi_p$	−40.546 ± 2.617	**X($\mathbf{D_p}$)**	0.9919	3.42	2629
9	−28.303 ± 1.834	11.275 ± 0.731	${}^3\chi_p$	4.8254 ± 0.3127	**Wi(RD)**	4.4946 ± 0.2912	**V(RD)**	0.9918	3.43	2609
10	−80.408 ± 5.240	11.157 ± 0.727	${}^3\chi_p$	27.648 ± 1.802	**MaxSp(RD)**	9.9073 ± 0.6457	**X($\mathbf{RD_p}$)**	0.9917	3.45	2579

[a] The MLR equations have the general form $t_b = a_0 + a_1\mathbf{SD}_1 + a_2\mathbf{SD}_2 + a_3\mathbf{SD}_3$.

Table 2. Coefficients, Confidence Interval, Structural Descriptors $\mathbf{SD}_i$ (i = 1−4), and Statistical Indices for the Best 10 MLR Equations with Four Independent Variables That Model the Alkane Boiling Temperature at Normal Pressure, t_b (°C)

eq	a_0	a_1	$\mathbf{SD}_1$	a_2	$\mathbf{SD}_2$	a_3	$\mathbf{SD}_3$	a_4	$\mathbf{SD}_4$	r	s	F
11	−97.884 ± 6.227	10.993 ± 0.699	$^3\chi_p$	38.777 ± 2.467	**MaxSp(RD)**	5.3194 ± 0.3384	**V(RD)**	−111.597 ± 7.099	**V(Sz$_p$)**	0.9951	2.67	3256
12	−101.424 ± 6.472	10.964 ± 0.700	$^3\chi_p$	39.005 ± 2.489	**MaxSp(RD)**	5.4272 ± 0.3463	**V(RD)**	−55.943 ± 3.570	**Y(Sz$_p$)**	0.9951	2.68	3236
13	−22.575 ± 1.479	12.541 ± 0.821	$^3\chi_p$	35.354 ± 2.316	**IB(D)**	−86.856 ± 5.690	**V(D)**	15.917 ± 1.043	**X(RD)**	0.9948	2.75	3071
14	16.029 ± 1.070	11.121 ± 0.742	$^3\chi_p$	130.993 ± 8.744	**MinSp(RD)**	54.228 ± 3.620	**MaxSp(RD)**	5.2343 ± 0.3494	**V(RD)**	0.9946	2.80	2957
15	−37.520 ± 2.513	10.239 ± 0.686	$^3\chi_p$	54.677 ± 3.662	**MaxSp(RD)**	83.360 ± 5.584	**MinSp(RD$_p$)**	5.2676 ± 0.3528	**V(RD)**	0.9946	2.81	2937
16	−10.954 ± 0.737	12.974 ± 0.873	$^3\chi_p$	31.871 ± 2.146	**IB(D)**	13.353 ± 0.899	**X(RD)**	−88.201 ± 5.938	**X(D$_p$)**	0.9945	2.82	2907
17	−180.638 ± 12.562	8.1713 ± 0.5683	$^3\chi_p$	54.686 ± 3.803	**MaxSp(RD)**	−1.1665 ± 0.0811	**IB(RD$_p$)**	8.2955 ± 0.5769	**V(RD)**	0.9941	2.91	2725
18	−7.3897 ± 0.5143	13.060 ± 0.909	$^3\chi_p$	6.0343 ± 0.4200	**Wi(RD$_p$)**	3.8737 ± 0.2696	**V(RD)**	−115.302 ± 8.025	**V(Sz$_p$)**	0.9941	2.92	2720
19	−13.231 ± 0.922	13.294 ± 0.926	$^3\chi_p$	7.7348 ± 0.5388	**HyWi(RD$_p$)**	3.6810 ± 0.2564	**V(RD)**	−109.324 ± 7.616	**V(Sz$_p$)**	0.9941	2.92	2715
20	−10.665 ± 0.745	13.034 ± 0.911	$^3\chi_p$	6.0762 ± 0.4246	**Wi(RD$_p$)**	3.9812 ± 0.2782	**V(RD)**	−57.508 ± 4.018	**Y(Sz$_p$)**	0.9941	2.93	2699

[a] The MLR equations have the general form $t_b = a_0 + a_1\mathbf{SD}_1 + a_2\mathbf{SD}_2 + a_3\mathbf{SD}_3 + a_4\mathbf{SD}_4$.

(2) Molar Heat Capacity. The best ten QSPR models with three or four independent variables that model the alkane molar heat capacity are presented in Tables 3 and 4, respectively. The first three-parametric model, eq 21, with r = 0.9883, s = 3.93, and F = 1812, contains two spectral descriptors, **MinSp**(**RD**) and **MaxSp**(**RD**), and an index computed from the Ivanciuc−Balaban operator, **IB**(**RD$_p$**). Only eqs 23, 27, and 28 contain an information index, indicating that for the alkane molar heat capacity the Ivanciuc−Balaban indices are more suitable than the information indices. As previously noted, almost all indices are derived from the reciprocal distance **RD** and the reciprocal distance-path **RD$_p$** matrices. Although eq 21 does not contain a connectivity index, there are five MLR models that contain two such indices, showing their importance in modeling the alkane molar heat capacity.

As can be seen from Table 4, the best MLR model with four independent variables, eq 31, is not derived from any of the models with three parameters reported in Table 3. This equation, with r = 0.9891, s = 3.79, and F = 1461, contains as descriptors **MW**, $^3\chi_p$, **Wi**(**RSz$_p$**), and **IB**(**D**). While the Ivanciuc−Balaban index computed from the distance matrix, **IB**(**D**), appears only in eq 31, the Wiener index derived from the reciprocal path Szeged matrix, **Wi**(**RSz$_p$**), appears in the first six equations from Table 4. All models from Tabe 4 contain at least one connectivity index, with $^3\chi_p$ being present in nine equations; equally important is the spectral descriptor **MaxSp**(**RD**) that appears in seven equations and **Wi**(**RSz$_p$**) that is found in six QSPR models. Although the first three QSPR models from Table 4 do not contain information indices, they appear in the next six equations, suggesting that they are important descriptors in modeling the alkane molar heat capacity. As observed in the previous cases, the majority of the indices are computed from **RD** and **RD$_p$** matrices, but one finds also two descriptors computed from the reciprocal path Szeged matrix, namely, **Wi**(**RSz$_p$**) and **V**(**RSz$_p$**).

(3) Standard Gibbs Energy of Formation. In Tables 5 and 6 we give the best ten QSPR models with three and four, respectively, independent variables that model the alkane standard Gibbs energy of formation. The best MLR eq 41 with three independent variables, with r = 0.9528, s = 4.52, F = 427, contains the Ivanciuc−Balaban index **IB**(**D**) and the information descriptors **V**(**RD**) and **Y**(**RD**). It appears that **IB**(**D**) is quite important in modeling the alkane standard Gibbs energy of formation, because only eq 48 does not contain this index. Equally important are the information indices derived from the **RD**, **RD$_p$**, and **RSz$_p$** matrices; they are present in all QSPR models from Table 5; moreover four equations, namely, equations 41−43 and 45 contain two such descriptors. A spectral descriptor appears only in eqs 45 and 50, two models contain a Wiener index, and only one model contains a connectivity index.

The QSPR equations that model the alkane standard Gibbs energy of formation are significantly improved by the addition of a fourth independent variable, as one can note from the inspection of the results reported in Table 6. By adding to the three descriptors from eq 41 a fourth one, namely, the Wiener index **Wi**(**RSz$_p$**), one obtains the best MLR eq 51, the best MLR equation with four independent variables, with r = 0.9635, s = 4.00, and F = 417. All QSPR models from Table 6 contain the Ivanciuc−Balaban index

Table 3. Coefficients, Confidence Interval, Structural Descriptors $\mathbf{SD}_i$ (i = 1−3), and Statistical Indices for the Best 10 MLR Equations with Three Independent Variables That Model the Alkane Molar Heat Capacity at 300 K, C_p (J K^{-1} mol^{-1})[a]

eq	a_0	a_1	$\mathbf{SD}_1$	a_2	$\mathbf{SD}_2$	a_3	$\mathbf{SD}_3$	r	s	F
21	139.743 ± 10.867	120.900 ± 9.402	**MinSp(RD)**	46.098 ± 3.585	**MaxSp(RD)**	2.8452 ± 0.2213	**IB(RD$_p$)**	0.9883	3.93	1812
22	8.0233 ± 0.6255	−2.9300 ± 0.2284	$^3\chi_p$	33.710 ± 2.628	**MaxSp(RD)**	3.5992 ± 0.2806	**IB(RD$_p$)**	0.9882	3.93	1803
23	102.038 ± 7.984	121.595 ± 9.515	**MinSp(RD)**	56.403 ± 4.413	**MaxSp(RD)**	6.6785 ± 0.5226	**X(RD$_p$)**	0.9881	3.95	1789
24	109.097 ± 8.554	94.738 ± 7.428	**MinSp(RD)**	41.359 ± 3.243	**MaxSp(RD)**	4.5600 ± 0.3575	**IB(RD)**	0.9881	3.96	1782
25	3.4666 ± 0.2718	43.223 ± 3.389	$^1\chi$	10.061 ± 0.789	$^2\chi$	0.15329 ± 0.01202	**MinSp(D$_p$)**	0.9881	3.96	1782
26	−1.4223 ± 0.1117	40.225 ± 3.158	$^1\chi$	8.2934 ± 0.6511	$^2\chi$	5.3908 ± 0.4232	**IB(D)**	0.9880	3.96	1777
27	−40.489 ± 3.180	−2.8825 ± 0.2264	$^3\chi_p$	46.573 ± 3.658	**MaxSp(RD)**	8.4685 ± 0.6651	**X(RD$_p$)**	0.9880	3.96	1776
28	12.091 ± 0.950	46.729 ± 3.671	$^1\chi$	8.0588 ± 0.6332	$^2\chi$	−1.7939 ± 0.1409	**V(RD)**	0.9880	3.96	1775
29	4.9847 ± 0.3918	43.213 ± 3.397	$^1\chi$	10.272 ± 0.807	$^2\chi$	0.71420 ± 0.05614	**MinSp(D)**	0.9880	3.97	1773
30	−1.2280 ± 0.0965	44.298 ± 3.482	$^1\chi$	9.8663 ± 0.7756	$^2\chi$	−0.003367 ± 0.000265	**HyWi(D$_p$)**	0.9880	3.97	1773

[a] The MLR equations have the general form $C_p = a_0 + a_1\mathbf{SD}_1 + a_2\mathbf{SD}_2 + a_3\mathbf{SD}_3$.

Table 4. Coefficients, Confidence Interval, Structural Descriptors $\mathbf{SD}_i$ (i = 1−4), and Statistical Indices for the Best 10 MLR Equations with Four Independent Variables That Model the Alkane Molar Heat Capacity at 300 K, C_p (J K^{-1} mol^{-1})[a]

eq	a_0	a_1	$\mathbf{SD}_1$	a_2	$\mathbf{SD}_2$	a_3	$\mathbf{SD}_3$	a_4	$\mathbf{SD}_4$	r	s	F
31	−22.653 ± 2.151	1.6836 ± 0.1599	MW	−6.0995 ± 0.5792	$^3\chi_p$	−2.5875 ± 0.2457	**Wi(RSz$_p$)**	13.307 ± 1.264	**IB(D)**	0.9891	3.79	1461
32	−29.391 ± 2.791	−5.9861 ± 0.5685	$^3\chi_p$	−2.3728 ± 0.2254	**Wi(RSz$_p$)**	49.751 ± 4.725	**MaxSp(RD)**	4.0817 ± 0.3877	**IB(RD)**	0.9891	3.79	1461
33	−23.122 ± 2.212	−6.2087 ± 0.5939	$^3\chi_p$	−2.0313 ± 0.1943	**Wi(RSz$_p$)**	49.360 ± 4.722	**MaxSp(RD)**	2.8099 ± 0.2688	**IB(RD$_p$)**	0.9890	3.82	1440
34	−69.421 ± 6.673	−5.9530 ± 0.5722	$^3\chi_p$	−2.3861 ± 0.2294	**Wi(RSz$_p$)**	57.883 ± 5.564	**MaxSp(RD)**	12.721 ± 1.223	**X(RD)**	0.9889	3.83	1426
35	−39.812 ± 3.838	1.9627 ± 0.1892	MW	−5.2051 ± 0.5018	$^3\chi_p$	−2.2735 ± 0.2192	**Wi(RSz$_p$)**	33.240 ± 3.205	**X(D$_p$)**	0.9888	3.85	1418
36	−60.776 ± 5.871	−6.1373 ± 0.5928	$^3\chi_p$	−2.0070 ± 0.1939	**Wi(RSz$_p$)**	59.272 ± 5.725	**MaxSp(RD)**	6.6261 ± 0.6401	**X(RD$_p$)**	0.9888	3.85	1412
37	6.4565 ± 0.6238	−3.7440 ± 0.3617	$^3\chi_p$	34.572 ± 3.340	**MaxSp(RD)**	3.5490 ± 0.3429	**IB(RD$_p$)**	0.033931 ± 0.003278	**V(RSz$_p$)**	0.9888	3.85	1412
38	−41.426 ± 4.018	−3.7534 ± 0.3640	$^3\chi_p$	47.291 ± 4.587	**MaxSp(RD)**	8.3465 ± 0.8095	**X(RD$_p$)**	0.036404 ± 0.003531	**V(RSz$_p$)**	0.9887	3.87	1401
39	4.7918 ± 0.4667	42.546 ± 4.144	$^1\chi$	10.179 ± 0.991	$^2\chi$	0.1439 ± 0.0140	**MinSp(D$_p$)**	0.031346 ± 0.003053	**V(RSz$_p$)**	0.9886	3.88	1381
40	−17.665 ± 1.721	−5.0643 ± 0.4934	$^3\chi_p$	−3.0416 ± 0.2963	$^3\chi_c$	41.699 ± 4.063	**MaxSp(RD)**	4.7585 ± 0.4636	**IB(RD)**	0.9886	3.88	1388

[a] The MLR equations have the general form $C_p = a_0 + a_1\mathbf{SD}_1 + a_2\mathbf{SD}_2 + a_3\mathbf{SD}_3 + a_4\mathbf{SD}_4$.

Table 5. Coefficients, Confidence Interval, Structural Descriptors **SD**$_i$ (i = 1−3), and Statistical Indices for the Best 10 MLR Equations with Three Independent Variables That Model the Alkane Standard Gibbs Energy of Formation in the Gas Phase at 300 K, $\Delta_f G°_{300}$(g) (kJ mol^{-1})[a]

eq	a_0	a_1	**SD**$_1$	a_2	**SD**$_2$	a_3	**SD**$_3$	r	s	F
41	−139.505 ± 22.340	42.313 ± 6.776	**IB(D)**	4.8331 ± 0.7740	**V(RD)**	−1.4790 ± 0.2368	**Y(RD)**	0.9528	4.52	427
42	−102.669 ± 16.779	38.510 ± 6.294	**IB(D)**	13.3278 ± 2.1781	**X(RD)**	−3.2052 ± 0.5238	**Y(RD)**	0.9510	4.60	410
43	−145.129 ± 23.778	38.904 ± 6.374	**IB(D)**	5.0194 ± 0.8224	**V(RD)**	−0.70056 ± 0.11478	**Y(RD$_p$)**	0.9508	4.61	408
44	−176.186 ± 28.929	43.568 ± 7.154	**IB(D)**	−1.1827 ± 0.1942	**IB(RD$_p$)**	6.7550 ± 1.1091	**V(RD)**	0.9506	4.62	406
45	−66.224 ± 10.919	2.3868 ± 0.3935	**MaxSp(D)**	39.880 ± 6.575	**IB(D)**	−4.2195 ± 0.6957	**Y(RD)**	0.9502	4.64	403
46	−94.900 ± 15.740	39.034 ± 6.474	**IB(D)**	−2.8102 ± 0.4661	**Y(RD)**	6.7168 ± 1.1140	**X(RD$_p$)**	0.9497	4.66	398
47	−165.935 ± 27.523	41.857 ± 6.943	**IB(D)**	−1.3278 ± 0.2202	**IB(RD)**	6.0855 ± 1.0094	**V(RD)**	0.9497	4.66	398
48	−43.666 ± 7.249	−6.4796 ± 1.0757	$^3\chi_c$	6.1000 ± 1.0127	**Wi(RD)**	−1.1827 ± 0.1963	**X(RSz$_p$)**	0.9496	4.67	398
49	−159.920 ± 26.551	−0.010882 ± 0.001807	**Wi(Sz$_p$)**	39.119 ± 6.495	**IB(D)**	5.0975 ± 0.8463	**V(RD)**	0.9496	4.67	397
50	−157.781 ± 26.220	−0.060519 ± 0.010057	**MaxSp(Sz$_p$)**	38.818 ± 6.451	**IB(D)**	5.0894 ± 0.8458	**V(RD)**	0.9495	4.67	397

[a] The MLR equations have the general form $\Delta_f G°_{300} = a_0 + a_1\mathbf{SD}_1 + a_2\mathbf{SD}_2 + a_3\mathbf{SD}_3$.

Table 6. Coefficients, Confidence Interval, Structural Descriptors **SD**$_i$ (i = 1−4), and Statistical Indices for the Best 10 MLR Equations with Four Independent Variables That Model the Alkane Standard Gibbs Energy of Formation in the Gas Phase at 300 K, $\Delta_f G°_{300}$(g) (kJ mol^{-1})

eq	a_0	a_1	**SD**$_1$	a_2	**SD**$_2$	a_3	**SD**$_3$	a_4	**SD**$_4$	r	s	F
51	−136.232 ± 24.204	4.0032 ± 0.7113	**Wi(RSz$_p$)**	43.759 ± 7.775	**IB(D)**	7.3067 ± 1.2982	**V(RD)**	−4.0401 ± 0.7178	**Y(RD)**	0.9635	4.00	417
52	−139.593 ± 24.943	3.4693 ± 0.6199	**HyWi(RSz$_p$)**	46.786 ± 8.360	**IB(D)**	7.3170 ± 1.3074	**V(RD)**	−3.9993 ± 0.7146	**Y(RD)**	0.9631	4.02	413
53	−156.089 ± 27.899	5.2072 ± 0.9307	**Wi(RSz$_p$)**	35.607 ± 6.364	**IB(D)**	9.4961 ± 1.6973	**V(RD)**	−2.8790 ± 0.5146	**Y(RD$_p$)**	0.9631	4.03	413
54	−159.749 ± 28.906	4.3844 ± 0.7933	**HyWi(RSz$_p$)**	39.566 ± 7.159	**IB(D)**	9.3492 ± 1.6917	**V(RD)**	−2.7739 ± 0.5019	**Y(RD$_p$)**	0.9622	4.07	402
55	−142.485 ± 27.800	6.9594 ± 1.3579	$^2\chi$	38.278 ± 7.468	**IB(D)**	7.3116 ± 1.4266	**V(RD)**	−2.7016 ± 0.5271	**Y(RD$_p$)**	0.9564	4.36	346
56	−76.407 ± 15.035	2.8408 ± 0.5590	**Wi(RSz$_p$)**	38.795 ± 7.634	**IB(D)**	−5.0892 ± 1.0014	**Y(RD)**	9.1294 ± 1.7964	**X(RD$_p$)**	0.9557	4.40	340
57	−79.048 ± 15.633	2.4128 ± 0.4772	**HyWi(RSz$_p$)**	40.896 ± 8.088	**IB(D)**	−5.0164 ± 0.9921	**Y(RD)**	9.0902 ± 1.7977	**X(RD$_p$)**	0.9553	4.42	337
58	−101.947 ± 20.290	0.63860 ± 0.12710	**Wi(RD)**	44.252 ± 8.807	**IB(D$_p$)**	4.7959 ± 0.9545	**V(RD)**	−183.389 ± 36.499	**V(Sz$_p$)**	0.9548	4.44	333
59	−136.337 ± 27.158	2.9446 ± 0.5866	$^3\chi_c$	44.744 ± 8.913	**IB(D)**	5.7705 ± 1.1495	**V(RD)**	−2.6045 ± 0.5188	**Y(RD)**	0.9547	4.45	332
60	−6.7332 ± 1.3415	0.008787 ± 0.001751	**HyWi(D$_p$)**	66.808 ± 13.311	**IB(D)**	−84.635 ± 16.862	**X(D)**	−4.5332 ± 0.9032	**Y(RD)**	0.9547	4.45	332

[a] The MLR equations have the general form $\Delta_f G°_{300} = a_0 + a_1\mathbf{SD}_1 + a_2\mathbf{SD}_2 + a_3\mathbf{SD}_3 + a_4\mathbf{SD}_4$.

IB(**D**), with the exception of eq 58 that contains the index **IB**($\mathbf{D_p}$). The importance of information indices in modeling the alkane standard Gibbs energy of formation is emphasized by the presence of two such indices in eqs 51−60, derived from the **RD** or $\mathbf{RD_p}$ matrices, with the exception of **V**($\mathbf{Sz_p}$) and **X**(**D**). The fourth descriptor is a Wiener, hyper-Wiener, or connectivity index, with **Wi**($\mathbf{RSz_p}$) and **HyWi**($\mathbf{RSz_p}$) having the greatest importance. Although the spectral descriptors **MinSp** and **MaxSp** were identified as significant parameters for modeling other physicochemical properties, these indices were not selected in the QSPR models from Table 6.

(4) Vaporization Enthalpy. The best ten QSPR models with three or four independent variables that model the alkane vaporization enthalpy are presented in Tables 7 and 8, respectively. The first three-parametric model, eq 61, with $r = 0.9891$, $s = 0.64$, and $F = 1959$, contains the connectivity index ${}^3\chi_p$, the spectral descriptor **MaxSp**(**RD**), and the information index **V**(**RD**). All MLR models contain the information index **V**(**RD**); eqs 63−65 and 68 contain also a **Y** index, either **Y**($\mathbf{RD_p}$) or **Y**($\mathbf{RSz_p}$). Another important index is **MaxSp** computed from the **RD**, $\mathbf{RD_p}$, and $\mathbf{RSz_p}$ matrices. A connectivity index appears in three QSPR models, an Ivanciuc−Balaban descriptor is found in three equations, and the molecular weight appears in eqs 66 and 67.

As one can see from Table 8, the best QSPR model with four independent variables, eq 71, is obtained from eq 61 by adding the Ivanciuc−Balaban index computed from the distance-path matrix, **IB**($\mathbf{D_p}$). The statistical indices of this equation, $r = 0.9898$, $s = 0.62$, and $F = 1564$, are only marginally better than those of eq 61, indicating that the addition of the fourth parameter does not significantly improve the modeling of alkane vaporization enthalpy. The trends regarding the most frequent descriptors observed for the QSPR models in Table 7 are noticed also in Table 8. The information index **V**(**RD**) is present in all equations; three models contain a second information index derived from the $\mathbf{Sz_p}$ matrix, namely, **U**($\mathbf{Sz_p}$), **V**($\mathbf{Sz_p}$), and **Y**($\mathbf{Sz_p}$), respectively. A connectivity index appears in seven QSPR models, the spectral index **MaxSp** computed from the **RD** or $\mathbf{RSz_p}$ matrices was selected in six MLR equations, and a **IB** index is found in eqs 71−75, 78, and 80.

(5) Refractive Index. In Tables 9 and 10 we present the best ten MLR equations with three and four, respectively, independent variables that model the alkane refractive index. The first MLR eq 71 with three independent variables has good statistical indices, i.e., $r = 0.9840$, $s = 0.0025$, and $F = 1309$, and contains a connectivity index, ${}^3\chi_p$, the Ivanciuc−Balaban index computed from the distance matrix, **IB**(**D**), and the information index **X**(**D**). All QSPR equations from this table contain the connectivity index ${}^3\chi_p$, indicating the importance of the weighted count of butane-like subgraphs in modeling the alkane refractive index. Although not present in the best two models, **MaxSp**(**RD**) appears in six equations, indicating its significance in this QSAR. The **IB** index appears in five equations, and an information index is selected in the first QSPR models from Table 9. Unlike previous results in which information indices were mainly computed from reciprocal matrices, in this case they are derived from the **D** and $\mathbf{Sz_p}$ matrices, with one exception, **V**(**RD**); the indices **X**(**D**) and **X**($\mathbf{Sz_p}$) are the most important descriptors from this class.

Table 7. Coefficients, Confidence Interval, Structural Descriptors $\mathbf{SD}_i$ ($i = 1-3$), and Statistical Indices for the Best 10 MLR Equations with Three Independent Variables That Model the Alkane Vaporization Enthalpy at 300 K, $\Delta_{vap}H_{300}$ (kJ mol^{-1})[a]

eq	a_0	a_1	$\mathbf{SD}_1$	a_2	$\mathbf{SD}_2$	a_3	$\mathbf{SD}_3$	r	s	F
61	−9.7040 ± 0.7256	−0.35708 ± 0.02670	${}^3\chi_p$	7.2484 ± 0.5420	**MaxSp**(**RD**)	1.5553 ± 0.1163	**V**(**RD**)	0.9891	0.64	1959
62	−9.3009 ± 0.6961	6.6116 ± 0.4948	**MaxSp**(**RD**)	0.37860 ± 0.02834	**MaxSp**($\mathbf{RSz_p}$)	1.6333 ± 0.1222	**V**(**RD**)	0.9891	0.64	1956
63	−12.740 ± 0.957	8.8377 ± 0.6638	**MaxSp**($\mathbf{RD_p}$)	1.6813 ± 0.1263	**V**(**RD**)	0.077031 ± 0.005786	**Y**($\mathbf{RSz_p}$)	0.9890	0.64	1942
64	−8.2712 ± 0.6226	6.3898 ± 0.4810	**MaxSp**(**RD**)	1.4981 ± 0.1128	**V**(**RD**)	0.076361 ± 0.005748	**Y**($\mathbf{RD_p}$)	0.9890	0.64	1934
65	−15.255 ± 1.150	9.0155 ± 0.6798	**MaxSp**($\mathbf{RD_p}$)	1.7043 ± 0.1285	**V**(**RD**)	0.14373 ± 0.01084	**Y**($\mathbf{RD_p}$)	0.9889	0.65	1927
66	−3.3821 ± 0.2556	0.16509 ± 0.01248	MW	2.0646 ± 0.1560	**IB**($\mathbf{D_p}$)	1.3658 ± 0.1032	**V**(**RD**)	0.9889	0.65	1918
67	−3.5348 ± 0.2673	0.12786 ± 0.00967	MW	2.3658 ± 0.1789	**IB**(**D**)	1.4219 ± 0.1075	**V**(**RD**)	0.9889	0.65	1917
68	−7.3777 ± 0.5585	6.4274 ± 0.4866	**MaxSp**(**RD**)	1.5060 ± 0.1140	**V**(**RD**)	0.031672 ± 0.002398	**Y**($\mathbf{RSz_p}$)	0.9889	0.65	1912
69	−8.3251 ± 0.6312	0.17523 ± 0.01328	${}^2\chi$	6.5012 ± 0.4929	**MaxSp**(**RD**)	1.5807 ± 0.1198	**V**(**RD**)	0.9888	0.65	1906
70	−5.3834 ± 0.4090	1.8169 ± 0.1380	${}^0\chi$	2.8035 ± 0.2130	**IB**(**D**)	1.7097 ± 0.1299	**V**(**RD**)	0.9888	0.65	1898

[a] The MLR equations have the general form $\Delta_{vap}H_{300} = a_0 + a_1\mathbf{SD}_1 + a_2\mathbf{SD}_2 + a_3\mathbf{SD}_3$.

Table 8. Coefficients, Confidence Interval, Structural Descriptors $\mathbf{SD}_i$ (i = 1−4), and Statistical Indices for the Best 10 MLR Equations with Four Independent Variables That Model the Alkane Vaporization Enthalpy at 300 K, $\Delta_{vap}H_{300}$ (kJ mol^{-1})

eq	a_0	a_1	$\mathbf{SD}_1$	a_2	$\mathbf{SD}_2$	a_3	$\mathbf{SD}_3$	a_4	$\mathbf{SD}_4$	r	s	F
71	−13.304 ± 1.221	−0.75902 ± 0.06966	$^3\chi_p$	6.8004 ± 0.6241	**MaxSp(RD)**	1.7033 ± 0.1563	**IB(D$_p$)**	1.8267 ± 0.1676	**V(RD)**	0.9898	0.62	1564
72	−7.3423 ± 0.6750	0.10693 ± 0.00983	MW	−0.68510 ± 0.06298	$^3\chi_p$	3.9093 ± 0.3594	**IB(D)**	1.6526 ± 0.1519	**V(RD)**	0.9898	0.62	1559
73	−7.0419 ± 0.6474	0.168510 ± 0.015493	MW	−0.67885 ± 0.06241	$^3\chi_p$	3.3933 ± 0.3120	**IB(D$_p$)**	1.5568 ± 0.1431	**V(RD)**	0.9898	0.62	1559
74	−6.8288 ± 0.6300	0.092611 ± 0.008544	MW	0.65493 ± 0.06042	**MaxSp(RSz$_p$)**	3.2334 ± 0.2983	**IB(D)**	1.7571 ± 0.1621	**V(RD)**	0.9897	0.62	1548
75	−6.4304 ± 0.5947	0.14475 ± 0.01339	MW	0.62565 ± 0.05786	**MaxSp(RSz$_p$)**	2.7691 ± 0.2561	**IB(D$_p$)**	1.6623 ± 0.1537	**V(RD)**	0.9897	0.63	1541
76	−15.006 ± 1.389	−0.59122 ± 0.05474	$^3\chi_p$	7.9750 ± 0.7383	**MaxSp(RD)**	1.6871 ± 0.1562	**V(RD)**	12.101 ± 1.120	**V(Sz$_p$)**	0.9897	0.63	1537
77	−9.2045 ± 0.8526	6.6490 ± 0.6159	**MaxSp(RD)**	0.75438 ± 0.06988	**MinSp(RSz$_p$)**	0.64725 ± 0.05996	**MaxSp(RSz$_p$)**	1.6337 ± 0.1513	**V(RD)**	0.9897	0.63	1536
78	−8.8466 ± 0.8198	1.5143 ± 0.1403	$^0\chi$	−0.67028 ± 0.06211	$^3\chi_p$	4.2655 ± 0.3953	**IB(D)**	1.8931 ± 0.1754	**V(RD)**	0.9897	0.63	1535
79	−14.5608 ± 1.3499	−0.58524 ± 0.05426	$^3\chi_p$	7.9416 ± 0.7362	**MaxSp(RD)**	1.6739 ± 0.1552	**V(RD)**	5.9902 ± 0.5553	**Y(Sz$_p$)**	0.9896	0.63	1533
80	−3.4262 ± 0.3177	−0.74041 ± 0.06866	$^3\chi_p$	4.1683 ± 0.3865	**IB(D)**	1.7890 ± 0.1659	**V(RD)**	0.32635 ± 0.03026	**U(Sz$_p$)**	0.9896	0.63	1532

[a] The MLR equations have the general form $\Delta_{vap}H_{300} = a_0 + a_1\mathbf{SD}_1 + a_2\mathbf{SD}_2 + a_3\mathbf{SD}_3 + a_4\mathbf{SD}_4$.

Table 9. Coefficients, Confidence Interval, Structural Descriptors $\mathbf{SD}_i$ (i = 1−3), and Statistical Indices for the Best 10 MLR Equations with Three Independent Variables That Model the Alkane Refractive Index at 25 °C, n_D^{25} [a]

eq	a_0	a_1	$\mathbf{SD}_1$	a_2	$\mathbf{SD}_2$	a_3	$\mathbf{SD}_3$	r	s	F
81	1.3768 ± 0.1259	0.008563 ± 0.000783	$^3\chi_p$	0.035447 ± 0.003243	**IB(D)**	−0.090690 ± 0.008296	**X(D)**	0.9840	0.0025	1309
82	1.3599 ± 0.1352	0.010478 ± 0.001041	$^3\chi_p$	0.013957 ± 0.001387	**IB(D)**	−0.094928 ± 0.009435	**X(Sz$_p$)**	0.9812	0.0027	1109
83	1.3162 ± 0.1319	0.010368 ± 0.001039	$^3\chi_p$	0.018681 ± 0.001872	**MaxSp(RD)**	−0.034151 ± 0.003422	**X(Sz$_p$)**	0.9809	0.0027	1091
84	1.3159 ± 0.1323	0.010450 ± 0.001051	$^3\chi_p$	0.018481 ± 0.001858	**MaxSp(RD)**	−0.069961 ± 0.007033	**V(Sz$_p$)**	0.9807	0.0027	1084
85	1.3155 ± 0.1329	0.010491 ± 0.001060	$^3\chi_p$	0.018446 ± 0.001863	**MaxSp(RD)**	−0.037071 ± 0.003744	**Y(Sz$_p$)**	0.9806	0.0027	1074
86	1.2849 ± 0.1319	0.009145 ± 0.000939	$^3\chi_p$	0.022541 ± 0.002314	**MaxSp(RD)**	0.000833 ± 0.000086	**V(RD)**	0.9799	0.0027	1040
87	1.3064 ± 0.1341	0.010001 ± 0.001027	$^3\chi_p$	0.023680 ± 0.002431	**MaxSp(RD)**	−0.015213 ± 0.001562	**X(D)**	0.9799	0.0028	1039
88	1.3587 ± 0.1404	0.010938 ± 0.001130	$^3\chi_p$	0.013139 ± 0.001358	**IB(D)**	−0.18926 ± 0.01956	**V(Sz$_p$)**	0.9797	0.0028	1026
89	1.3867 ± 0.1443	0.011018 ± 0.001146	$^3\chi_p$	0.017379 ± 0.001808	**IB(D$_p$)**	−0.14559 ± 0.01515	**X(Sz$_p$)**	0.9794	0.0028	1012
90	1.3123 ± 0.1372	0.010441 ± 0.001092	$^3\chi_p$	0.020036 ± 0.002095	**MaxSp(RD)**	−0.015078 ± 0.001577	**IB(Sz$_p$)**	0.9792	0.0028	1002

[a] The MLR equations have the general form $n_D^{25} = a_0 + a_1\mathbf{SD}_1 + a_2\mathbf{SD}_2 + a_3\mathbf{SD}_3$.

Table 10. Coefficients, Confidence Interval, Structural Descriptors $\mathbf{SD}_i$ (i = 1−4), and Statistical Indices for the Best 10 MLR Equations with Four Independent Variables That Model the Alkane Refractive Index at 25 °C, n_D^{25} [a]

eq	a_0	a_1	$\mathbf{SD}_1$	a_2	$\mathbf{SD}_2$	a_3	$\mathbf{SD}_3$	a_4	$\mathbf{SD}_4$	r	s	F
91	1.4034 ± 0.1573	0.008944 ± 0.001003	$^3\chi_p$	0.045182 ± 0.005065	**IB(D)**	−0.000682 ± 0.000076	**IB($\mathbf{RD_p}$)**	−0.12974 ± 0.01454	**X(D)**	0.9851	0.0024	1049
92	1.4183 ± 0.1590	−0.006829 ± 0.000766	$^1\chi$	0.009901 ± 0.001110	$^3\chi_p$	0.045975 ± 0.005155	**IB(D)**	−0.13410 ± 0.01504	**X(D)**	0.9851	0.0024	1048
93	1.4061 ± 0.1581	0.008761 ± 0.000985	$^3\chi_p$	0.046167 ± 0.005190	**IB(D)**	−0.001060 ± 0.000119	**IB(RD)**	−0.13245 ± 0.01489	**X(D)**	0.9850	0.0024	1043
94	1.4159 ± 0.1597	−0.000260 ± 0.000029	MW	0.008317 ± 0.000938	$^3\chi_p$	0.051042 ± 0.005758	**IB(D)**	−0.13922 ± 0.01571	**X(D)**	0.9849	0.0024	1035
95	1.3982 ± 0.1583	0.008622 ± 0.000976	$^3\chi_p$	0.041032 ± 0.004646	**IB(D)**	−0.11887 ± 0.01346	**X(D)**	−0.000242 ± 0.000027	**X($\mathbf{RSz_p}$)**	0.9848	0.0024	1028
96	1.4166 ± 0.1604	0.008877 ± 0.001005	$^3\chi_p$	0.044615 ± 0.005052	**IB(D)**	−0.13281 ± 0.01504	**X(D)**	−0.003912 ± 0.000443	**X(RD)**	0.9848	0.0024	1028
97	1.4073 ± 0.1597	0.009065 ± 0.001029	$^3\chi_p$	0.042099 ± 0.004777	**IB(D)**	−0.12458 ± 0.01414	**X(D)**	−0.001704 ± 0.000193	**X($\mathbf{RD_p}$)**	0.9847	0.0024	1024
98	1.3932 ± 0.1585	0.008740 ± 0.000994	$^3\chi_p$	0.039800 ± 0.004527	**IB(D)**	−0.000062 ± 0.000007	**IB($\mathbf{RSz_p}$)**	−0.11308 ± 0.01286	**X(D)**	0.9847	0.0024	1019
99	1.3962 ± 0.1589	0.008491 ± 0.000966	$^3\chi_p$	0.044401 ± 0.005052	**IB(D)**	−0.12227 ± 0.01391	**X(D)**	−0.000459 ± 0.000052	**U($\mathbf{RD_p}$)**	0.9846	0.0024	1018
100	1.3947 ± 0.1588	0.008430 ± 0.000960	$^3\chi_p$	0.045034 ± 0.005126	**IB(D)**	−0.12201 ± 0.01389	**X(D)**	−0.000548 ± 0.000062	**U(RD)**	0.9846	0.0024	1017

[a] The MLR equations have the general form $n_D^{25} = a_0 + a_1\mathbf{SD}_1 + a_2\mathbf{SD}_2 + a_3\mathbf{SD}_3 + a_4\mathbf{SD}_4$.

The results from Table 10 show that the addition of a fourth parameter does not significantly improve the modeling of alkane refractive index. A close inspection of these data reveals that eqs 91−100 are obtained from eq 81 by adding a topological index that is not highly intercorrelated with the three descriptors from eq 81. The statistical indices of these equations do not show a significant improvement over the best models from Table 9, indicating that using our collection of structural descriptors it is not possible to improve the QSPR model represented by eq 81.

(6) Density. The best ten MLR equations with three and four, respectively, independent variables that model the alkane density are presented in Tables 11 and 12. The first QSPR model, eq 101, with $r = 0.9902$, $s = 3.73$, F = 2156, contains the same descriptors from eq 81, namely, $^3\chi_p$, **IB**(**D**), and **X**(**D**). The similarity with the descriptors selected for modeling the alkane density is even greater: all QSPR equations from this table contain the connectivity index $^3\chi_p$, **IB**(**D**) appears in four models, the information indices are derived from the **D** and $\mathbf{Sz_p}$ matrices, and **X** appears as the most important information descriptor. Also, **MaxSp**(**RD**) appears in four equations, but not in the first four.

The QSPR models for the alkane density are slightly improved by the addition of a fourth independent variable, as one can note from the inspection of the results reported in Table 12. With the exception of eq 115, all other models from Table 12 are obtained from eq 101 by adding a topological index. The statistical indices of eq 111, $r = 0.9927$, $s = 3.22$, and $F = 2182$, show a small but significant improvement over those of eq 101, and the incorporation of the molecular weight increases the value of F, indicating that the contribution of this descriptor is important for modeling the alkane density.

(7) Frequency of the Structural Descriptors in the QSPR Models. The results obtained in this study indicate that all six alkane properties can be modeled with the set of 60 structural descriptors, but each property requires a particular combination of molecular descriptors. An inspection of the QSPR models reported in Tables 1−12 reveals that some descriptors appear with a greater frequency, while others are rarely present in these equations. To obtain an indication of the importance of each descriptor in modeling the six alkane properties, for each set of QSPR models we have computed the counts of the **MW**, χ, and molecular operators **Wi**, **HyWi**, **MinSp**, **MaxSp**, **IB**, **U**, **V**, **X**, and **Y**. We have to mention that each operator is computed for six molecular matrices, and that we have used five χ indices. The frequencies of the structural descriptors in the QSPR models from Tables 1−12 are reported in Table 13; the structural descriptors are arranged in the decreasing order of their frequencies. Overall, the size descriptor **MW** appears with a low frequency, indicating that the molecular size is incorporated into other structural descriptors. Other descriptors that appear, in general, with a low frequency are **Wi**, **HyWi**, **MinSp**, **U**, and **Y**; however, in specific QSPR models they exhibit a higher importance, such as **Wi** in four-descriptor QSPR models for the molar heat capacity(Table 4) or **Y** in three- and four-descriptor QSPR models for the standard Gibbs energy of formation (Tables 5 and 6). All five connectivity indices were selected in the QSPR models, with $^3\chi_p$ appearing more frequently, indicating that subgraph-counting descriptors cannot be substituted with global

Table 11. Coefficients, Confidence Interval, Structural Descriptors $\mathbf{SD}_i$ (i = 1−3), and Statistical Indices for the Best 10 MLR Equations with Three Independent Variables That Model the Alkane Density at 25 °C, ρ (kg m^{-3})[a]

eq	a_0	a_1	$\mathbf{SD}_1$	a_2	$\mathbf{SD}_2$	a_3	$\mathbf{SD}_3$	r	s	F
101	655.305 ± 46.714	19.217 ± 1.370	$^3\chi_p$	62.894 ± 4.483	**IB**(**D**)	−156.304 ± 11.142	**X**(**D**)	0.9902	3.73	2156
102	597.331 ± 43.988	20.786 ± 1.531	$^3\chi_p$	52.881 ± 3.894	**IB**(**D**)	−115.283 ± 8.490	**V**(**D**)	0.9895	3.85	2020
103	625.994 ± 51.092	22.580 ± 1.843	$^3\chi_p$	25.790 ± 2.105	**IB**(**D**)	−162.602 ± 13.271	**X**($\mathbf{Sz_p}$)	0.9872	4.26	1645
104	675.367 ± 56.196	23.304 ± 1.939	$^3\chi_p$	32.762 ± 2.726	**IB**($\mathbf{D_p}$)	−258.712 ± 21.527	**X**($\mathbf{Sz_p}$)	0.9867	4.34	1582
105	461.883 ± 39.202	−8.3256 ± 0.7066	$^2\chi$	12.568 ± 1.067	$^3\chi_p$	64.237 ± 5.452	**MaxSp**(**RD**)	0.9862	4.42	1521
106	482.755 ± 41.118	16.134 ± 1.374	$^3\chi_p$	−5.1919 ± 0.4422	$^3\chi_c$	51.281 ± 4.368	**MaxSp**(**RD**)	0.9861	4.44	1510
107	623.937 ± 53.587	23.382 ± 2.008	$^3\chi_p$	24.374 ± 2.093	**IB**(**D**)	−323.685 ± 27.800	**V**($\mathbf{Sz_p}$)	0.9858	4.47	1485
108	548.727 ± 47.873	22.778 ± 1.987	$^3\chi_p$	33.576 ± 2.929	**MaxSp**(**RD**)	−52.535 ± 4.583	**X**($\mathbf{Sz_p}$)	0.9854	4.54	1439
109	696.297 ± 60.827	23.442 ± 2.048	$^3\chi_p$	20.391 ± 1.781	**Y**(**D**)	−329.030 ± 28.744	**X**($\mathbf{Sz_p}$)	0.9854	4.55	1436
110	548.222 ± 48.004	22.894 ± 2.005	$^3\chi_p$	33.300 ± 2.916	**MaxSp**(**RD**)	−107.241 ± 9.390	**V**($\mathbf{Sz_p}$)	0.9853	4.56	1429

[a] The MLR equations have the general form $\rho = a_0 + a_1\mathbf{SD}_1 + a_2\mathbf{SD}_2 + a_3\mathbf{SD}_3$.

Table 12. Coefficients, Confidence Interval, Structural Descriptors $\mathbf{SD}_i$ (i = 1−4), and Statistical Indices for the Best 10 MLR Equations with Three Independent Variables That Model the Alkane Density at 25 °C, ρ (kg m^{-3})

eq	a_0	a_1	$\mathbf{SD}_1$	a_2	$\mathbf{SD}_2$	a_3	$\mathbf{SD}_3$	a_4	$\mathbf{SD}_4$	r	s	F
111	781.897 ± 60.762	−0.84327 ± 0.06553	MW	18.420 ± 1.431	$^3\chi_p$	113.446 ± 8.816	**IB**(**D**)	−313.625 ± 24.372	**X**(**D**)	0.9927	3.22	2182
112	694.435 ± 54.523	18.666 ± 1.466	$^3\chi_p$	−0.15015 ± 0.01179	**MaxSp**($\mathbf{Sz_p}$)	81.529 ± 6.401	**IB**(**D**)	−227.787 ± 17.885	**X**(**D**)	0.9926	3.25	2138
113	734.718 ± 57.921	−11.416 ± 0.900	$^0\chi$	16.821 ± 1.326	$^3\chi_p$	104.547 ± 8.242	**IB**(**D**)	−266.271 ± 20.991	**X**(**D**)	0.9925	3.27	2120
114	685.173 ± 54.074	18.985 ± 1.498	$^3\chi_p$	−0.023699 ± 0.001870	**Wi**($\mathbf{Sz_p}$)	79.373 ± 6.264	**IB**(**D**)	−217.958 ± 17.201	**X**(**D**)	0.9925	3.27	2116
115	567.279 ± 44.806	16.186 ± 1.278	$^3\chi_p$	54.617 ± 4.314	**IB**($\mathbf{D_p}$)	5.4818 ± 0.4330	**V**(**RD**)	−412.355 ± 32.570	**V**($\mathbf{Sz_p}$)	0.9925	3.27	2112
116	706.057 ± 55.797	18.336 ± 1.449	$^3\chi_p$	99.233 ± 7.842	**IB**(**D**)	−261.115 ± 20.635	**X**(**D**)	−2.0521 ± 0.1622	**U**($\mathbf{Sz_p}$)	0.9925	3.27	21108
117	718.312 ± 57.021	18.556 ± 1.473	$^3\chi_p$	97.634 ± 7.750	**IB**(**D**)	−2.0173 ± 0.1601	**U**(**D**)	−268.022 ± 21.276	**X**(**D**)	0.9924	3.29	2091
118	719.418 ± 57.153	18.400 ± 1.462	$^3\chi_p$	94.321 ± 7.493	**IB**(**D**)	−262.644 ± 20.865	**X**(**D**)	−1.7086 ± 0.1357	**U**($\mathbf{D_p}$)	0.9924	3.29	2088
119	675.302 ± 53.871	18.816 ± 1.501	$^3\chi_p$	−0.001481 ± 0.000118	**HyWi**($\mathbf{Sz_p}$)	74.716 ± 5.960	**IB**(**D**)	−200.459 ± 15.991	**X**(**D**)	0.9924	3.30	2071
120	719.810 ± 57.440	17.891 ± 1.428	$^3\chi_p$	−6.1616 ± 0.4917	**HyWi**($\mathbf{RD_p}$)	108.974 ± 8.696	**IB**(**D**)	−279.340 ± 22.291	**X**(**D**)	0.9924	3.30	2069

[a] The MLR equations have the general form $\rho = a_0 + a_1\mathbf{SD}_1 + a_2\mathbf{SD}_2 + a_3\mathbf{SD}_3 + a_4\mathbf{SD}_4$.

Table 13. Frequency of the Structural Descriptors in the QSPR Models Reported in Tables 1−12

table	χ	IB	V	X	MaxSp	Y	Wi	MW	HyWi	MinSp	U
1	10	0	7	3	3	0	2	3	2	0	0
2	10	3	12	3	5	2	2	0	1	2	0
3	12	4	1	2	5	0	0	0	1	5	0
4	12	5	3	4	7	0	6	2	0	1	0
5	1	11	6	3	2	5	2	0	0	0	0
6	2	10	8	3	0	9	4	0	4	0	0
7	3	3	10	0	8	4	0	2	0	0	0
8	8	7	11	0	7	1	0	4	0	1	1
9	10	5	3	5	6	1	0	0	0	0	0
10	11	13	0	13	0	0	0	1	0	0	2
11	12	5	3	5	4	1	0	0	0	0	0
12	11	10	2	9	1	0	1	1	2	0	3
Total	102	76	66	50	48	23	17	13	10	9	6

descriptors. Other important descriptors for modeling the six alkane properties are the spectral operator **MaxSp** and the Ivanciuc−Balaban operator **IB**. From the four information-theory operators, our computations revealed a greater importance of the **V** and **X** descriptors; nevertheless, the **U** and **Y** operators can exhibit an important contribution for modeling certain properties, such as the **Y** descriptors for the standard Gibbs energy of formation. Among the four information-theory operators, **U** was found from the outset to be the least promising one.[30]

CONCLUDING REMARKS

Molecular graph descriptors represent valuable structural descriptors that can be used with success in developing QSPR and QSAR models; in such structure−property studies, graph descriptors can be used in conjunction with other classes of structural descriptors, such as constitutional, geometrical, electrostatic, and quantum descriptors. In this study we have investigated the utility of the information-theory operators **U**(**M**), **V**(**M**), **X**(**M**), and **Y**(**M**) in developing QSPR models for six alkane properties: boiling temperature, molar heat capacity, standard Gibbs energy of formation, vaporization enthalpy, refractive index, and density. For the generation of the QSPR models we have used also a selection of the most used molecular graph descriptors: molecular weight, **MW**; connectivity indices $^0\chi$, $^1\chi$, $^2\chi$, $^3\chi_p$, $^3\chi_c$; Wiener indices **Wi**(**M**); hyper-Wiener indices **HyWi**(**M**); spectral indices **MinSp**(**M**) and **MaxSp**(**M**); Ivanciuc−Balaban indices **IB**(**M**). The molecular graph operators were applied to six molecular matrices, namely, the distance **D**, the reciprocal distance **RD**, the distance-path $\mathbf{D_p}$, the reciprocal distance-path $\mathbf{RD_p}$, the path Szeged $\mathbf{Sz_p}$, and the reciprocal path Szeged $\mathbf{RSz_p}$ matrices. The results obtained show that all six alkane properties can be modeled with three or four descriptors selected from the group of descriptors, but each property requires a particular combination of molecular descriptors. From the four information-theory operators, our results indicate a greater importance of the **V**, **X**, and **Y** descriptors in developing QSPR models for the six alkane properties: **V** for boiling temperature, standard Gibbs energy of formation and vaporization enthalpy; **X** for refractive index and density; **Y** for standard Gibbs energy of formation. For the molar heat capacity, the Ivanciuc−Balaban indices **IB** are more important than the information-theory indices, but the **X** indices give also good QSPR models for this property.

ACKNOWLEDGMENT

O.I. thanks the Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche of France for a PAST grant. T.I. and O.I. acknowledge the kind hospitality of the LARTIC group during their stay in Nice. We acknowledge the partial financial support of this research by the Romanian Ministry of National Education under Grant 7001 T34.

REFERENCES AND NOTES

(1) Jurs, P. C. *ADAPT*; Chemistry Department, Penn State University: University Park, PA. E-mail: pcj@psu.edu: www http://zeus.chem.psu.edu/ADAPT.html.

(2) Sutter, J. M.; Dixon, S. L.; Jurs, P. C. Automated Descriptor Selection for Quantitative Structure−Activity Relationships Using Generalized Simulated Annealing. *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 77−84.

(3) Mitchell, B. E.; Jurs, P. C. Prediction of Aqueous Solubility of Organic Compounds from Molecular Structure. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 489−496.

(4) Mitchell, B. E.; Jurs, P. C. Computer Assisted Simulations of ^{13}C Nuclear Magnetic Spectra of Monosaccharides. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 58−64.

(5) Clouser, D. L.; Jurs, P. C. Simulation of the ^{13}C Nuclear Magnetic Resonance Spectra of Ribonucleosides Using Multiple Linear Regression Analysis and Neural Networks. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 168−172.

(6) Mekenyan, O.; Karabunarliev, S.; Bonchev, D. The Microcomputer OASIS System for Predicting the Biological Activity of Chemical Compounds. *Comput. Chem*. **1990**, *14*, 193−200.

(7) Mekenyan, O. G.; Karabunarliev, S. H.; Ivanov, J. M.; Dimitrov, D. N. A New Development of the OASIS Computer System for Modeling Molecular Properties. *Comput. Chem*. **1994**, *18*, 173−187.

(8) Bonchev, D.; Mountain, C. F.; Seitz, W. A.; Balaban, A. T. Modeling the Anticancer Action of Some Retinoid Compounds by Making Use of the OASIS Method. *J. Med. Chem*. **1993**, *36*, 1562−1569.

(9) *SciQSAR*, SciVision, Inc.: Burlington, MA. E-mail: scivision@delphi.com; www http://www.scivision.com.

(10) Murugan, R.; Grendze, M. P.; Toomey, J. E., Jr.; Katritzky, A. R.; Karelson, M.; Lobanov, V. S.; Rachwal, P. Predicting Physical Properties from Molecular Structure. *CHEMTECH* **1994**, *24*, 17−23.

(11) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. QSPR: The Correlation and Quantitative Prediction of Chemical and Physical Properties from Structure. *Chem. Soc. Rev.* **1995**, 279−287.

(12) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* **1996**, *96*, 1027−1043.

(13) Katritzky, A. R.; Wang, Y.; Sild, S.; Tamm, T.; Karelson, M. QSPR Studies on Vapor Pressure, Aqueous Solubility, and the Prediction of Water−Air Partition Coefficients. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 720−725.

(14) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. Normal Boiling Points for Organic Compounds: Correlation and Prediction by a Quantitative Structure−Property Relationship. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 28−41.

(15) Ivanciuc, O.; Ivanciuc, T.; Balaban, A. T. Quantitative Structure−Property Relationship Study of Normal Boiling Points for Halogen-/Oxygen-/Sulfur-Containing Organic Compounds Using the CODESSA Program. *Tetrahedron* **1998**, *54*, 9129−9142.

(16) Ivanciuc, O.; Ivanciuc, T.; Filip, P. A.; Cabrol-Bass, D. Estimation of the Liquid Viscosity of Organic Compounds with a Quantitative Structure−Property Model. *J. Chem. Inf. Comput. Sci*. **1999**, *39*, 515−524.

(17) *Cerius[2] 3.0 QSAR*+; Molecular Simulations Inc.: San Diego, CA, 1997.

(18) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure−Activity Analysis*; Research Studies Press: Letchworth, U.K., 1986.

(19) Randić, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc*. **1975**, *97*, 6609−6615.

(20) Balaban, A. T. Using Real Numbers as Vertex Invariants for Third Generation Topological Indexes. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 23−28.

(21) Balaban, A. T. Local versus Global (i.e. Atomic versus Molecular) Numerical Modeling of Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 398−402.

(22) Balaban, A. T. Local (Atomic) and Global (Molecular) Graph-Theoretical Descriptors. *SAR QSAR Environ. Res*. **1995**, *3*, 81−95.

(23) Ivanciuc, O.; Balaban, A. T. Graph Theory in Chemistry. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, U.K., 1998; pp 1169−1190.
(24) Balaban, A. T.; Ivanciuc, O. Historical Development of Topological Indices. In: *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach Science Publishers: Amsterdam, The Netherlands, 1999; pp. 21−57.
(25) Ivanciuc, O.; Balaban, A. T. The Graph Description of Chemical Structures. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach Science Publishers: Amsterdam, The Netherlands, 1999; pp. 59−167.
(26) Ivanciuc, O.; Ivanciuc, T.; Balaban, A. T. Vertex- and Edge-Weighted Molecular Graphs and Derived Structural Descriptors. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach Science Publishers: Amsterdam, The Netherlands, 1999; pp 169−220.
(27) Ivanciuc, O.; Ivanciuc, T. Matrices and Structural Descriptors Computed from Molecular Graph Distances. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach Science Publishers: Amsterdam, The Netherlands, 1999; pp 221−277.
(28) Ivanciuc, O.; Ivanciuc, T.; Diudea, M. V. Molecular Graph Matrices and Derived Structural Descriptors. *SAR QSAR Environ. Res*. **1997**, *7*, 63−87.
(29) Diudea, M. V.; Gutman, I. Wiener-Type Topological Indices. *Croat. Chem. Acta* **1998**, *71*, 21−51.
(30) Balaban, A. T.; Balaban, T.-S. New Vertex Invariants and Topological Indices of Chemical Graphs Based on Information on Distances. *J. Math. Chem*. **1991**, *8*, 383−397.
(31) Balaban, A. T.; Balaban, T.-S. Correlations Using Topological Indices Based on Real Graph Invariants. *J. Chim. Phys. Phys.-Chim. Biol.* **1992**, *89*, 1735−1745.
(32) Balaban, A. T.; Feroiu, V. Correlations Between Structure and Critical Data or Vapor Pressures of Alkanes by Means of Topological Indices. *Rep. Mol. Theor*. **1990**, *1*, 133−139.
(33) Ivanciuc, O.; Balaban, T.-S.; Balaban, A. T. Chemical Graphs with Degenerate Topological Indices Based on Information on Distances. *J. Math. Chem*. **1993**, *14*, 21−33.
(34) Ivanciuc, O.; Laidboeur, T.; Cabrol-Bass, D. Degeneracy of Topologic Distance Descriptors for Cubic Molecular Graphs: Examples of Small Fullerenes. *J. Chem. Inf. Comput. Sci*. **1997**, *37*, 485−488.
(35) Ivanciuc, O.; Balaban, A. T. Design of Topological Indices. Part 20. Molecular Structure Descriptors Computed with Information on Distances Operators. *Rev. Roum. Chim*., in press.
(36) Ivanciuc, O.; Balaban, A. T. Design of Topological Indices. Part 21. Molecular Graph Operators for the Computation of Geometric Structural Descriptors. *Rev. Roum. Chim*., in press.
(37) Gakh, A. A.; Gakh, E. G.; Sumpter, B. G.; Noid, D. W. Neural Network-Graph Theory Approach to the Prediction of the Physical Properties of Organic Compounds. *J. Chem. Inf. Comput. Sci*. **1994**, *34*, 832−839.
(38) Ivanciuc, O. Design of Topological Indices. Part 18. Modeling the Physical Properties of Alkanes with Molecular Graph Descriptors Derived from the Hosoya Operator. *Rev. Roum. Chim*., in press.
(39) Ivanciuc, O. Design on Topological Indices. 1. Definition of a Vertex Topological Index in the Case of 4-Trees. *Rev. Roum. Chim*. **1989**, *34*, 1361−1368.
(40) Balaban, T. S.; Filip, P. A.; Ivanciuc, O. Computer Generation of Acyclic Graphs Based on Local Vertex Invariants and Topological Indices. Derived Canonical Labeling and Coding of Trees and Alkanes. *J. Math. Chem*. **1992**, *11*, 79−105.
(41) Plavšić, D.; Nikolić, S.; Trinajstić, N.; Mihalić, Z. On the Harary Index for the Characterization of Chemical Graphs. *J. Math. Chem*. **1993**, *12*, 235−250.
(42) Ivanciuc, O.; Balaban, T.-S.; Balaban, A. T. Design of Topological Indices. Part 4. Reciprocal Distance Matrix, Related Local Vertex Invariants and Topological Indices. *J. Math. Chem*. **1993**, *12*, 309−318.
(43) Diudea, M. V.; Ivanciuc, O.; Nikolić, S.; Trinajstić, N. Matrices of Reciprocal Distance, Polynomials and Derived Numbers. *MATCH (Commun. Math. Comput. Chem.)* **1997**, *35*, 41−64.
(44) Diudea, M. V. Walk Numbers $^{e}W_{M}$: Wiener-type Numbers of Higher Rank. *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 535−540.
(45) Diudea, M. V. Wiener and Hyper-Wiener Numbers in a Single Matrix. *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 833−836.
(46) Diudea, M. V.; Minailiuc, O.; Katona, G.; Gutman, I. Szeged Matrices and Related Numbers. *MATCH (Commun. Math. Comput. Chem.)* **1997**, *35*, 129−143.
(47) Diudea, M. V. Cluj Matrix Invariants. *J. Chem. Inf. Comput. Sci*. **1997**, *37*, 300−305.
(48) Diudea, M. V.; Pârv, B.; Topan, M. I. Derived Szeged and Cluj Indices. *J. Serb. Chem. Soc*. **1997**, *62*, 267−276.
(49) Ivanciuc, O. Design of Topological Indices. Part 13. Structural Descriptors Computed from the Szeged Molecular Matrices. *Rev. Roum. Chim*., in press.
(50) Ivanciuc, O. Design of Topological Indices. Part 19. Computation of Vertex and Molecular Graph Structural Descriptors with Operators. *Rev. Roum. Chim*., in press.
(51) Ivanciuc, O.; Ivanciuc, T.; Balaban A. T. Design of Topological Indices. Part 10. Parameters Based on Electronegativity and Covalent Radius for the Computation of Molecular Graph Descriptors for Heteroatom-Containing Molecules. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 395−401.
(52) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc*. **1947**, *69*, 17−20.
(53) Wiener, H. Correlation of Heats of Isomerization and Differences in Heats of Vaporization of Isomers among the Paraffin Hydrocarbons. *J. Am. Chem. Soc*. **1947**, *69*, 2636−2638.
(54) Hosoya, H. A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332−2339.
(55) Randić, M. Novel Molecular Descriptor for Structure−Property Studies. *Chem. Phys. Lett*. **1993**, *211*, 478−483.
(56) Randić, M.; Guo, X.; Oxley, T.; Krishnapriyan, H. Wiener Matrix: Source of Novel Graph Invariants. *J. Chem. Inf. Comput. Sci*. **1993**, *33*, 709−716.
(57) Randić, M.; Guo, X.; Oxley, T.; Krishnapriyan, H.; Naylor, L. Wiener Matrix Invariants. *J. Chem. Inf. Comput. Sci*. **1994**, *34*, 361−367.
(58) Klein, D. J.; Lukovits, I.; Gutman, I. On the Definition of the Hyper-Wiener Index for Cycle-Containing Structures. *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 50−52.
(59) Ivanciuc, O.; Diudea, M. V.; Khadikar, P. V. New Topological Matrices and Their Polynomials. *Ind. J. Chem*. **1998**, *37A*, 574−585.
(60) Ivanciuc, O. Design of Topological Indices. Part 14. Distance-Valency Matrices and Structural Descriptors for Vertex- and Edge-Weighted Molecular Graphs. *Rev. Roum. Chim*., in press.
(61) Ivanciuc, O.; Ivanciuc, T.; Cabrol-Bass, D.; Balaban A. T. Optimum Structural Descriptors Derived from the Ivanciuc−Balaban Operator. *J. Chem. Inf. Comput. Sci.*, in press.

CI9900884