ELSEVIER

Journal of Molecular Graphics and Modelling 20 (2001) 19–35

Journal of Molecular Graphics and Modelling

www.elsevier.com/locate/JMGM

The connectivity index 25 years after

Milan Randić [a,b]

[a] *Department of Mathematics and Computer Science, Drake University, Des Moines, IA 50014, USA*
[b] *National Institute of Chemistry, Ljubljana, Slovenia*

Received 21 November 2000; received in revised form 6 April 2001; accepted 2 May 2001

Abstract

We review the developments following introduction of the connectivity indices as molecular descriptors in multiple linear regression analysis (MLRA) for structure–property–activity studies. We end the review with discussion of results obtained with applications of the variable connectivity index. A comparison is made between some results obtained with the traditional topological indices and the variable connectivity index. © 2001 Published by Elsevier Science Inc.

Keywords: Connectivity index; Multiple linear regression analysis; Structure–property–activity studies

1. Introduction

According to E.B. Wilson [1]:

> ... every once in a while some new theory or a new experimental method or apparatus makes it possible to enter a new domain. Sometimes it is obvious to all that this opportunity has arisen, but in other cases recognition of the opportunity requires more imagination.

Twenty-five years ago a significant turn in structure–property–activity studies occurred with a paper "Molecular Connectivity. I. Relationship to Nonspecific Local Anesthesia" by Kier et al. [2]. Until that time the discipline of structure–property–activity was dominated by the traditional quantitative structure activity relationship (QSAR) methodology as developed by Hansch and Leo [3]. The fundamental distinction between the two approaches is in use of descriptors to characterize molecules. The novel approach of Kier et al. is based on use of mathematical characterization of compounds in contrast to physico-chemical characterizations used by traditional QSAR. The new approach was made possible by construction of the connectivity index [4] χ (or ${}^1\chi$), a bond additive mathematical invariant of molecules, designed to parallel relative magnitudes of the boiling points in smaller alkanes. Another important contributions by the same authors that followed and expanded the power of the connectivity index was the paper "Molecular Connectivity. V. Connectivity Series Concept Applied to Density" by Kier et al. [5]. Here the higher order connectivity indices ${}^m\chi$ were introduced for the first time. This made it possible to apply the new mathematical characterization of compounds to study structure–property–activity relationships by multiple linear regression analysis (MLRA). Finally, another important development followed: Kier and Hall [6] modified the connectivity indices to discriminate between carbon atoms and other heteroatoms, which has led to the valence connectivity index ${}^m\chi^{\mathrm{v}}$.

Within a short time Kier and Hall compiled a large number of structure–property and structure–activity regressions, which were collected in the book [7] "Molecular Connectivity in Chemistry and Drug Research". There is no doubt that this book played the major role in expansion of the new methodology based on connectivity indices to QSAR. The new approach found immediately followers (for bibliography of early applications of the connectivity indices to quantitative structure–property relationship (QSPR) and QSAR see [8]), just as at the same time it continued to be misunderstood by others [9]. Indeed, it takes some imagination to recognize a novel approach, based on mathematical characterization of chemical structure, as complementary, rather than competitive, to the prevailing approaches of the past based on use of selected properties as descriptors. Besides the already mentioned higher order connectivity indices and the valence connectivity indices soon novel directions in mathematical characterization of chemical, biochemical, and of recently biological systems followed that may have been stimulated by the connectivity index, its generalizations, and associated numerous developments in chemical graph theory [10]. In Table 1 we have briefly listed a few major impacts that the connectivity index may have made or induced in the broad domain of structure–property–activity studies.

1093-3263/01/$ – see front matter © 2001 Published by Elsevier Science Inc.
PII: S1093-3263(01)00098-5

Table 1
A list of novel developments following introduction of the connectivity index that may have been stimulated or owe to some degree their appearance indirectly to the connectivity index[a]

Year	Topic	Authors
1975	The connectivity index	Randić
1976	The higher order connectivity indices	Kier, Murray, Randić, Hall
1976	The valence connectivity indices	Kier and Hall
	Development of new topological indices (TI)	Everybody
1987	Search for pharmacophore	Randić
1989	Inverse problem	Baskin, Gordeeva, Devdariani, Zefirov, Palyulin, Stankevitch
1990	Variable molecular descriptors	Randić
1990	Electrotopological state	Kier and Hall
1991	X'/X topological indices	Randić
1992	The concept of a basis descriptors	Randić
1995	Line graph topological indices	Estrada and Gutman
1997	Double invariants	Randić, Plavšic, Razinger,
1998	TI in combinatorial libraries	Lahana et al.
2000	DNA characterization	Randić
2000	Characterization of proteomics maps	Randić

[a] Most of the references not mentioned in the text can be found in review articles: [15–17,41,60]. Work on DNA has been published in recent issues of J. Chem. Inf. Comput. Sci. to which also the work on characterization of proteomics maps has been submitted.

2. Molecular descriptors

Multivariate linear regression analysis continues to be used in QSAR and QSPR studies. Often the most critical step in such studies is the selection of molecular descriptors, which may have physico-chemical, graph theoretical (topological), and quantum mechanical origin.

2.1. Physico-chemical descriptors

In traditional QSAR studies [11] besides selected physico-chemical descriptors (such as Hammett σ) also a few physico-chemical properties are used to describe correlation (such as $\log P$, molar refraction (MR)). Strictly speaking such studies do not offer a structure–property relationship. They relate physico-chemical descriptors and properties as descriptors to considered biological property (activity). Hence, at best they represent a mixture of property–property and structure–property relationship. For a recent survey on characterization of chemical structures using molecular properties see [12]. Use of property–property correlation should not be discouraged. Such correlation may show if the properties considered depend on the same structural features or not. Moreover, property–property correlation, if properties are closely related may point to possible data inconsistencies, as has been the case with solubilities of alcohols [13].

2.2. Graph theoretical descriptors

Graph theoretical descriptors are usually based on mathematical properties of molecular skeletons (typically not involving hydrogen atoms) [14–16]. In Fig. 1 we illustrate molecular graphs of 18 octane isomers, which depict C–C bonds and their connectivity. There are no limitations on construction of novel structural invariants, though in order to curb proliferation of invariants some guidance was offered [17]. For example, it is desirable that novel descriptors involve some novel structural feature that other descriptors fail to capture, because only then can they improved

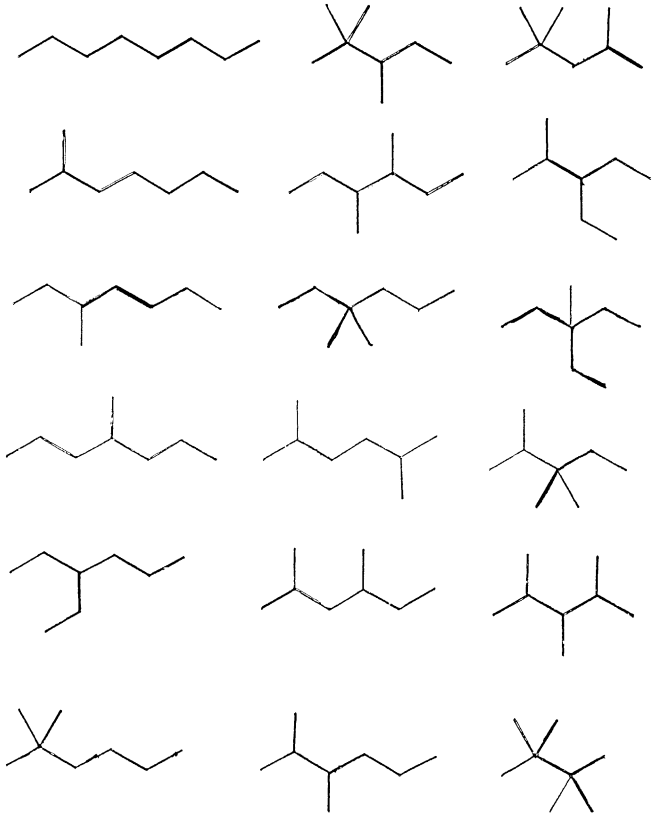

Fig. 1. Molecular skeletons of 18 isomers of octane.

Table 2
Numerical values for a selection of topological indices for octane isomers illustrating the diversity of such indices

		Wiener (W)	Hosoya (Z)	Randić (χ)	Hyper–Wiener	Path eigenvalue	p_2/w_2
1	*n*-Octane	84	34	3.91421	462		0.45833
2	2-Methylheptane	79	29	3.77005	382	10.2866	0.50208
3	3-Methylheptane	76	31	3.80806	336	10.2359	0.50000
4	4-Methylheptane	75	30	3.80806	321	10.2211	0.49167
5	3-Ethylhexane	72	32	3.84606	275	10.1696	0.49167
6	2,2-Dimethylhexane	71	23	3.56066	278	10.0816	0.55625
7	2,3-Dimethylhexane	70	27	3.68073	259	10.0900	0.54167
8	2,4-Dimethylhexane	71	26	3.66390	269	10.1151	0.53958
9	2,5-Dimethylhexane	74	25	3.62859	309	10.1712	0.54583
10	3,3-Dimethylhexane	67	25	3.62132	228	9.9994	0.55417
11	3,4-Dimethylhexane	68	29	3.71874	234	10.0484	0.54167
12	2-Methyl-3-ethylhexane	67	28	3.71874	219	10.0329	0.53810
13	3-Methyl-3-ethylhexane	64	28	3.68198	193	9.9314	0.55982
14	2,2,3-Trimethylpentane	63	22	3.48138	187	9.8771	0.60268
15	2,2,4-Trimethylpentane	66	19	3.41650	217	9.9543	0.59345
16	2,3,3-Trimethylpentane	62	23	3.50403	177	9.8512	0.60774
17	2,3,4-Trimethylpentane	65	24	3.55341	203	9.9526	0.58810
18	2,2,3,3-Tetramethylbutane	58	17	3.25000	145		0.66964

on regressions that have shown limited success previously. The oldest non-trivial structural descriptors are: the Wiener number [18] W, and the topological index Z introduced by Hosoya [19]. The Wiener number counts the lengths of all distances between each pair of atoms in a molecule, and the Hosoya's Z topological index counts all sets of non-adjacent bonds in a structure. In Table 2 we listed for octane isomers these graph theoretical indices, together with the values of the connectivity index χ and a few other indices in order to illustrate the diversity of mathematical invariants available for characterization of molecules. As we see from Table 2 molecular descriptors occasionally show degeneracy, that is, different structures display the same magnitude for a descriptor. This clearly shows that there is a loss of information when molecule is represented by a set of descriptors. Hence, the inverse problem, that of determining a structure when several of its descriptors are given, may not to be easy and need not be possible in some cases. The inverse problem of construction of a graph from a collection of invariants, despite its enormous importance, unfortunately, has received only a limited attention in the literature [20–22].

The Wiener number and Hosoya's Z topological index may serve to illustrate another important problem that has plagued chemical graph theory: the problem of interpretation of topological indices. This problem has been mostly overlooked in the past and still does not receive enough attention. This problem, however, should not be confused with the fact that often construction of topological indices is based on relatively elegant mathematical definitions. What is missing is an *insight* into the meaning of so constructed invariants in terms of *simple structural concepts*. In order to stress the distinction between the definition and interpretation consider the Wiener index. We can calculate W at least in three relatively simple and elegant ways: (1) as originally outlined by Wiener [18] by summing bond contributions given by the product of the number of carbon atoms on each side of a bond; (2) by summing the entries above the main diagonal in the distance matrix as pointed out by Hosoya [19]; (3) as a scalar product of vector P listing the number of paths of different length P (p_1, p_2, p_3, ...) and vector L indicating their corresponding lengths L (1, 2, 3, ...) [23]. Each of these recipe can be used for making an equivalent mathematical definition of W, but the question remains: What does the Wiener index represent in structural terms?

2.3. Quantum chemical descriptors

Quantum chemical descriptors are derived from models based on quantum chemical calculations. Commonly used descriptors include computed atomic charges and HOMO–LUMO energies based on MO calculations. However, it is often overlooked when quantum chemical descriptors are employed, that quantum chemical descriptors, just as various topological indices, are non-observables. Thus, there is no justification for preference of quantum chemical descriptors over graph theoretical descriptors (or vice versa). Quantum chemical descriptors are neither more nor less fundamental than graph theoretical descriptors. Descriptors should, however, have an interpretation within the model employed-quantum chemical descriptors within quantum chemistry and graph theoretical descriptors within chemical graph theory. Different theoretical models even within quantum chemistry may use concepts that have limited interpretation valid for one particular model only. For instance, Kekulé valence structures have no apparent meaning within MO models just as HOMO–LUMO separation has no clear interpretation within VB model. In parallel, therefore, to insist that topological indices ought to have physico-chemical interpretation is tantamount to insisting that physico-chemical concepts must have graph theoretical interpretation, or

quantum chemical interpretation, and that Kekulé valence structures ought to have simple meaning in MO theory. Such concepts should necessarily have a clear meaning within the model in which they play a role, but outside such models they may have some interpretation but need not.

3. The connectivity index

The connectivity χ index that has been proposed 25 years ago [3], in contrast to the Wiener index W and the Hosoya index Z both of which represent an ad hoc constructions, has been *designed*. In fact, this is the first so-called topological index that has been *designed* having specific goals in mind, rather than being "invented" and then tested for its use, which typifies most if not all other topological indices. The idea behind the connectivity index is to use available information on some molecular property for their construction. The relative values of the boiling points of smaller alkanes were in fact used in construction of the connectivity index. The numerical values for bond contributions to the connectivity index for smaller alkane isomers were so selected that the computed index parallels the relative numerical values of the boiling points. The connectivity index has another important advantage over the Wiener index and the Hosoya's Z topological index, in that it is accompanied by a relatively simple interpretation in terms of basic structural element-bonds. The connectivity index χ is a bond additive quantity in which terminal C–C bonds are given a greater contribution than the inner C–C bonds using different bond weights. The greatest weight of $1/\sqrt{2}$ has been assigned to the primary bond between CH_3 and CH_2 groups, the weight $1/\sqrt{3}$ has been assigned to the secondary C–C bond between CH_3 and CH groups, etc. Formally the connectivity index can be constructed, as pointed out by Balaban [24], from the row sums R_i and R_j of the adjacency matrix using the algorithm $1/\sqrt{R_i R_j}$ for the contribution of the bond (i, j).

The first "fruits" of the emergence of the connectivity index is appearance of several indices derived by applying the algorithmic approach to other then the adjacency matrix (Table 3). Thus Balaban [24,25] arrived at the topological index J, which is obtained from the graph distance matrix. This is one of early indices that has shown lesser degeneracy, which is one of the desirable properties of molecular descriptors. Analogous indices could be constructed using other graph matrices, e.g. the Wiener matrix [26,27], the Hosoya matrix [28], the restricted random Walk matrix [29], the Szeged matrix [30], the Cluj matrix [31], the Path matrix [32], etc. In fact any well-defined graph matrix can in this way generate a topological index analogous to the connectivity index χ, by using the row sums from the matrix considered combined by the algorithm $1/\sqrt{R_i R_j}$. This algorithm has been recently referred to as Balaban–Ivanciuc algorithm [33].

4. Sequential topological indices

Most graph theoretical invariants (topological indices [14–16]) are, just as the Wiener index and the Hosoya Z index, ad hoc constructions. In applications such indices are combined or selected so to produce regression equations associated with the smallest standard error. It is therefore not uncommon to find different combinations of descriptors for different properties even when properties are closely related. It would seem better if one could use the same set of descriptors in a study of different properties as this would allow more meaningful comparisons. Structurally closely related indices, to be referred to as "sequential indices", have an advantage in facilitating comparative study of different molecular properties [34,35]. The first sequential indices, the paths of different length, were proposed as potential molecular descriptors over 50 years ago by Platt [36,37]. They were overlooked till the later revival of the chemical graph theory [38–41].

Table 3
List of various matrices associated with molecules and molecular graphs

Matrix	Year	Authors
Bond property	1940	Balandin
Atom connectivity	1963	Spialter
Distance	1969	Harary
Detour	1969	Harary
Expanded distance	1990	Tratch, Stankevitch and Zefirov
Electrotopological	1990	Kier and Hall
Reciprocal distance	1992	Balaban, Filip and Ivanciuc
Wiener	1993	Randić
Hosoya	1994	Randić
Distance/distance (D/D)	1997	Randić, Kleiner, DeAlba
Path	1997	Randić, Plavšic, Razinger
Szeged	1997	Diudea, Minailiuc, Katona and Gutman
Cluj	1997	Diudea
Excluded neighborhood	2000	Randić and M. Basak
Nearest neighborhood	2000	Randić

Table 4
List of families of structural indices that are closely structurally related

Year	Symbol	Name	Authors
1947	p_1, p_2, p_3	Path numbers	Platt
1976	${}^m\chi$	Higher order connectivity indices	Kier, Murray, Randić and Hall
1976	${}^m\chi^v$	Higher order valence connectivity	Kier and Hall
1980	W_k	Walks	Randić
1985	${}^1\kappa$, ${}^2\kappa$, ${}^3\kappa$	Kappa shape indices	Kier
1992	p_m^x	Extended path numbers	Randić
1993	${}^kW$	Higher order Wiener numbers	Randić et al.
1994	${}^mZ$	Higher order Hosoya indices	Herman and Zinn; Randić
1995	${}^kM$	Molecular profiles	Randić
1995	${}^kS$	Shape profiles	Randić and Razinger
1999	${}^k\phi$	Folding profiles	Randić and Krilov
2000	p_k/w_k	Path/walk quotients	Randić

Table 5
Illustration of correlations of molecules having heteroatoms in which the valence connectivity index better describes property than the simple connectivity index

Property	Compounds	Descriptors	r	s
Heat of atomization	Alcohols	n, ${}^1\chi$	0.9999	0.990
		n, ${}^1\chi$, ${}^1\chi^v$	0.9999	0.640
Molecular refraction R_m	Alcohols	${}^1\chi$	0.9886	1.02
		${}^1\chi^v$	0.9926	0.92
	Amines	${}^1\chi$		2.30
		${}^1\chi^v$	0.997	1.92
	Primary amines	${}^1\chi$	0.9936	4.00
		${}^1\chi^v$	0.9991	1.56

In Table 4 we collected topological indices which form sequences of descriptors and can serve as a basis for characterization of different properties for the same set of compounds. This class includes, besides the path numbers, $p_1, p_2, p_3, \ldots$, the higher order connectivity indices [5] ${}^m\chi$, and the higher order weighted paths [40]. Additional descriptors of this kind followed, including the Kier's kappa indices [42,43] ${}^m\kappa$, the path/walks quotients [44], and so-called molecular profiles [45–48]. A "molecular profile" is obtained from suitably normalized matrices derived by individually raising matrix elements of a matrix associated with a graph to higher powers.

5. Modifications for the presence of heteroatoms

A number of topological indices have been generalized to describe the presence of heteroatoms. Kier and Hall [6] were

Table 6
Illustration of correlations in which both the valence connectivity index and the simple connectivity index are combined to offer better correlation (r is the coefficient of regression and s the standard error)

Property	Compounds	Descriptors	r	s
		${}^1\chi^v$	0.991	0.787
	Amines	${}^1\chi$		2.30
		${}^1\chi^v$	0.997	1.92
Molecular refraction R_m	Alcohols	${}^1\chi$, ${}^1\chi^v$		0.720
Boiling points	Ethers	${}^1\chi$	0.9851	5.69
		${}^1\chi$, ${}^1\chi^v$	0.9882	5.39
Partition coefficient	Aliphatic alcohols	${}^1\chi^v$	0.9612	0.195
		${}^1\chi$, ${}^1\chi^v$	0.9655	0.186
	Aliphatic ketones	${}^1\chi^v$	0.9929	0.098
		${}^1\chi$, ${}^1\chi^v$	0.9967	0.068
	Carboxylic acids	${}^1\chi^v$	0.9708	0.346
		${}^1\chi$, ${}^1\chi^v$	0.9979	0.099

Table 7
Illustration of correlations of molecules having heteroatoms in which the simple connectivity index better describes property than the valence connectivity index

Property	Compounds	Descriptors	r	s
Water solubility	Aliphatic ethers	$^1\chi$	0.9833	0.354
	Alcohols and ethers	$^1\chi$, $^1\chi^v$	0.9853	0.337
	Hydrocarbons	$^1\chi$	0.952	1.81
		$^1\chi$, $^2\chi$	0.9997	0.121
	Aliphatic ethers	$^1\chi$	0.9680	0.091
		$^1\chi$, $^2\chi$	0.9762	0.083
Anesthetic activity	Mixed	$^1\chi$	0.982	0.409
Barnacle larvae narcosis	Alcohols	$^1\chi$	0.987	0.163
Inhibitors of thymidine phosphorylase		$^1\chi$	0.920	0.213
Inhibitors of adenosine deaminase		$^1\chi$	0.990	0.086
Butyrycholinesterase inhibitors		$^1\chi$	0.996	0.055
Vapor toxicities for tomato	Alcohols	$^1\chi$	0.963	0.116
Vapor toxicities for red spider	Alcohols	$^1\chi$	0.977	0.090

first to recognize a need for a modification of the connectivity index in order to offer better regressions when compounds having heteroatoms are considered. This has led to the valence connectivity indices $^m\chi^v$, which produced improved correlation for a number of compounds and properties. In Table 5 we collect a few cases that show a better performance of the valence connectivity index in such situations.

In Table 6 we show correlations employing both, the valence connectivity index $^1\chi^v$ and the ordinary connectivity index $^1\chi$. As we see use of both connectivity indices often produce satisfactory results that are visibly better then when these descriptor are used individually. That two descriptors should produce better regression than one is to be expected, but it has not been clear why the simple connectivity index $^1\chi$, which does not differentiate heteroatoms, should be one of such descriptors that improves the performance of the valence connectivity index. Even more curious are the results collected in Table 7 in which regressions using the simple connectivity index $^1\chi$ give better result than the regressions using the valence connectivity index $^1\chi^v$. How can this be understood? The question that could have been raised and considered a long time ago remained, however, unanswered until very recently.

What has been overlooked in the past when using the connectivity indices is that *different properties* of the same compounds may require *different descriptors*. Thus there are no inherently unique descriptors for heteroatoms that will satisfy different regression for different properties equally well. The valence descriptors that have been constructed for heteroatoms by Kier and Hall may be optimal for one property, but at the same time may be quite unsuitable as descriptors for other properties of the same molecules. It appears thus to be desirable that descriptors have some flexibility to accommodate for variability when different properties of the same compounds are considered. Because the valence connectivities (and other modified topological indices for heteroatom) are of "fixed" kind they cannot serve as the best descriptors for numerous molecular properties, but at best just for a few.

6. Variable connectivity index $^1\chi^f$

An answer to the problem of characterization of heteroatoms that allows some flexibility is the variable connectivity index [49,50]. The variable index $^1\chi^f$ was constructed by augmenting the valence (or the row sum in the adjacency matrix) of different atoms by introducing variables x, y, z, ... The numerical values for the variables need to be selected so to minimize the standard error for a regression. In this way in different applications the numerical values of the connectivity index for different structures can adapt. Moreover, one can even use different weights for the same kind of atom in different molecular environments. For example, one can differentiate oxygen atoms in alcohols, ethers, esters, ketones and fatty acids [51], or carbon atoms in cyclic parts of a molecule and in acyclic parts of the same molecules [52].

The variable connectivity index, including also the higher order variable connectivity indices, was introduced about 10 years ago, but apparently until very recently remained overlooked. We will illustrate construction of the variable connectivity index, $^1\chi^f$ on 2,3-dimethylpentane. In Table 8 we show the augmented adjacency matrix for

Table 8
The augmented adjacency matrix and corresponding row sums

	1	2	3	4	5	6	7	Row sum
1	x	1	0	0	0	0	0	$1+x$
2	1	x	1	0	0	1	0	$3+x$
3	0	1	x	1	0	0	1	$3+x$
4	0	0	1	x	1	0	0	$2+x$
5	0	0	0	1	x	0	0	$1+x$
6	0	1	0	0	0	x	0	$1+x$
7	0	0	1	0	0	0	x	$1+x$

2,3-dimethylpentane which has on the main diagonal instead of zeros variable x. The connectivity index $^1\chi^f$ is constructed by summing over all bonds the reciprocal square root function of the product of the row sums for adjacent atoms:

$$^1\chi^f = \frac{3}{\sqrt{(1+x)(3+x)}} + \frac{1}{3+x} + \frac{1}{\sqrt{(3+x)(2+x)}} + \frac{1}{\sqrt{(2+x)(1+x)}}$$

The connectivity index is now a function of a single variable $f(x)$. In Table 9 we have listed the expressions for the variable connectivity index for all isomers of heptane.

In the case of 2-methyl-3-pentanol, the only difference in the augmented matrix is variable y that characterizes the oxygen atom, hence the connectivity index becomes:

$$^1\chi^f = \frac{2}{\sqrt{(1+x)(3+x)}} + \frac{1}{3+x} + \frac{1}{\sqrt{(3+x)(2+x)}} + \frac{1}{\sqrt{(2+x)(1+x)}} + \frac{1}{\sqrt{(3+x)(1+y)}}$$

which is a function of two variables, $f(x, y)$.

If several heteroatoms are present in a molecule we will have several variables. For example, in the case of amino acids [53], we have besides carbon atoms (variable x), oxygen (variable y), also nitrogen (variable z) and sulfur (variable w). Similarly, in the case of chloroflourocarbons we would have separate variables for carbon atom, fluorine, chlorine and bromine. The power of variable molecular descriptors lies in their flexibility to adapt their relative magnitudes to data. One starts analysis by assuming definite values for the variables $x, y, z, \ldots$ from which numerical values of $^1\chi^f$ can be computed for all compounds considered. MLRA will then yield initial r_0, s_0 and F_0 values. In the next step one alters the initial values for the variables and recalculates molecular descriptors. They yield a new set of statistical parameters r_1, s_1 and F_1. If $s_1 < s_0$ one continues changing variables till a minimum for s_i is found.

7. On limitations of topological indices

Limitations of a set of indices used in MLRA is reflected in the magnitudes of the residuals of the regression considered. If residuals (or the standard error) are small, we may say that the descriptors used adequately span the structure space for the property. As is known a set of descriptors that adequately span the structure space for one property may not describe other properties equally well. Hence, for the same set of compounds we may have a multitude of structure–property spaces.

The notion of basis descriptors, that is, descriptors that would describe several different properties reasonably well, runs counter to the notion of the optimal descriptors, that is, the best descriptors for a particular property. Nevertheless, the concept of basis descriptors deserves more attention. Optimal descriptors often are not robust, that is, they may be sensitive on exclusion of a single structure from the set considered (because it is an outlier, or experimental data are not reliable), or an inclusion or exclusion of a single descriptor. Such modifications may dramatically change the composition of the optimal set of descriptors already established. In Table 9 we illustrate the lack of robustness on an example taken from the literature [54]. Each line in Table 9 gives the best model for the correlation of the boiling points for isomers of nonane C_9H_{20} using from one to seven connectivity indices as molecular descriptors in a stepwise regression. As we see in the initial stages of the stepwise regression each time a novel descriptor was introduces the previously

Table 9
The expressions for the variable connectivity index for heptane isomers

1	*n*-Heptane	$\frac{2}{v(1+x)(2+x)} + \frac{4}{2+x}$
2	2-Methylhexane	$\frac{2}{v(1+x)(3+x)} + \frac{2}{2+x}\frac{1}{v(2+x)(3+x)} + \frac{1}{v(1+x)(2+x)}$
3	3-Methylhexane	$\frac{2}{v(1+x)(2+x)} + \frac{1}{2+x} + \frac{2}{v(2+x)(3+x)} + \frac{1}{v(1+x)(3+x)}$
4	3-Ethylpentane	$\frac{3}{v(1+x)(2+x)} + \frac{3}{v(2+x)(3+x)}$
5	2,2-Dimethylpentane	$\frac{2}{v(1+x)(4+x)} + \frac{2}{v(2+x)(4+x)} + \frac{2}{v(1+x)(2+x)}$
6	2,3-Dimethylpentane	$\frac{3}{v(1+x)(3+x)} + \frac{1}{v(1+x)(2+x)} + \frac{1}{v(2+x)(3+x)} + \frac{1}{3+x}$
7	2,4-Dimethylpentane	$\frac{4}{v(1+x)(3+x)} + \frac{2}{v(2+x)(3+x)}$
8	3,3-Dimethylpentane	$\frac{3}{v(1+x)(3+x)} + \frac{1}{3+x} + \frac{1}{v(2+x)(3+x)} + \frac{1}{v(1+x)(2+x)}$
9	2,2,3-Trimethylbutane	$\frac{3}{v(1+x)(4+x)} + \frac{2}{v(1+x)(3+x)} + \frac{1}{v(3+x)(4+x)}$

employed descriptors have been replaced by new descriptors. Under such conditions it is difficult to interpret the results, because the interpretation will depend on descriptors used, and these do not show constancy.

7.1. On stepwise regression method for obtaining a model

Most of regressions that we discuss here are related to the stepwise regression model, in contrast to examination of all possible subsets, which is another method widely used. According to one of the referees of this manuscript: "the set of descriptors obtained at each step in the regression analysis is not very useful; attempting to interpret indices picked in the first few steps does not seem very sound, when the regression quality is not adequate. Meaningful interpretation occurs only when there is a strong relationship between property and descriptors". The only part of this opinion to which we agree is that "attempting to interpret indices ... when the regression quality is not adequate" is not sound. But as the orthogonalization of descriptors for MLRA has shown the contributions of individual descriptors in stepwise regression when a descriptors occurs for the first time are the same as for orthogonalized set, hence, despite that the first descriptors in stepwise regression do not give satisfactory regression they are part of the final regression equation and contribute to interpretation of a correlation. However, to obtain a meaningful interpretation of a regression it is not sufficient only that "there is a strong relationship between property and descriptors", and that the descriptors themselves individually have suitable structural or physico-chemical interpretation. One needs to clarify how to interpret linear combination of descriptors, regardless whether the descriptor used are mutually related or orthogonal. That this problem is difficult is clear from numerous attempts to interpret the principal components in PCA. The principal components are mutually orthogonal, but are expressed as linear combinations of descriptors that are not orthogonal. Even if the components of individual principal components would be orthogonal the interpretation of linear combination of descriptors (components) remains to plague PCA. Recently Randić and Zupan [55] considered the problem of interpretation of linear combination of descriptors, which we will briefly outline in a later section in this manuscript.

7.2. On lack of interpretability of topological indices

Some criticism has been raised concerning physico-chemical meaning of topological indices. Although it may be desirable to have an interpretation of various topological indices in term of physico-chemical concepts, to request that topological indices must have physico-chemical interpretation is unreasonable. Such demand overlooks the fact that chemical graph theory, its models, and its concepts, represents a discipline distinct from physical chemistry, just as physical chemistry represents discipline different from quantum chemistry, etc.

Molecular descriptors should have an interpretation *within* the discipline and the model used, and need not have meaning in related disciplines and *different* models. However, many topological indices do not have clear interpretation even within the structural chemistry. For example, what is the structural interpretation of the Wiener number? Is this an index that measure molecular compactness? What is molecular compactness? Is this an index that characterizes molecular volume? What is molecular volume? Recent work of Bytautas and Klein [56] shows that as the size of some polymeric structures tends to infinity the Wiener number increases with power 5/2. Hence, it could represent a property that grows with power 5/2, but what is this property?

Randić and Zupan [57] recently initiated a systematic approach for interpretation of topological indices in terms of bond contributions as the most elementary structural fragment. For indices that are bond additive, as is the case with the connectivity index, it is trivial to partition the molecular descriptor into bond contributions. This is illustrated in Fig. 2 on 2,3-dimethylhexane (top). Also in the case of Wiener index partitioning of W into bond contribution is not difficult if one follows the original algorithm for calculation of W as proposed by Wiener. Each bond contributions in W is given by the product of the number of atoms on each side of the bond considered. In the top part of Fig. 2 we illustrate bond contributions to W for 2,3-dimethylhexane. In the middle part of Fig. 2 we show partitioning of the Hosoya Z number into bond contributions outlined in [57]. Comparison of bond contributions associated with different indices is instructive. As we see from Fig. 2 the connectivity index and the Hosoya index both give higher weights to contributions arising from peripheral bonds, and smaller weight to internal C–C bonds, while the opposite is true of the Wiener

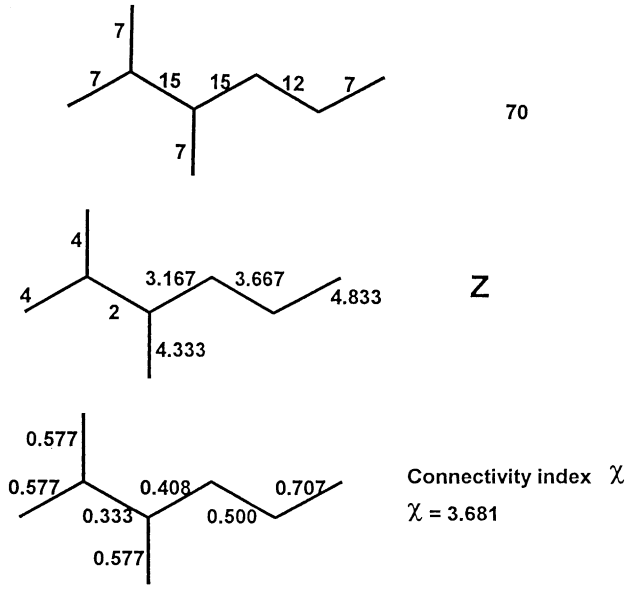

Fig. 2. Partitioning of the Wiener index, the Hosoya index and the connectivity index for 2,3-dimethylhexane.

index (and a few other indices, like Balaban's J index, as shown by Randić and Pompe [58] and Randić et al. [59]). Thus different indices show different characteristics when partitioned to bond contributions, which is one of the reasons why different indices emerge more suitable for different properties.

8. On use and misuse of topological indices

We will here give several illustrations of misuse of topological indices, the connectivity indices in particular. It may appear to some that this "debate" is not appropriate for the paper, as one referee pointed out: "it sounds much more to me like a debate following a paper at a meeting. There is no opportunity for rebuttal in a manuscript. I say this even though I share some of the same sentiment as the author". We fully agree with this view, however we had no opportunity for rebuttal of a provocative statements on connectivity indices, such as cited in recent book of Kubinyi. If we would not challenge offensive assertions made by hostile critics those who are less familiar with chemical graph theory and use of mathematical descriptors for characterization of molecules could be mislead by claims that have been left unanswered.

1. A paper by Dunn and coworkers [60] has been cited as an early "warning" on use of the connectivity index. According to Taylor (see P. Taylor, as cited by Kubinyi in [9]) molecular connectivity indices, while they may be suitable for correlations of physico-chemical properties are not suitable for structure–activity correlations. According to Taylor, as quoted in a recent book by Kubinyi [9] one should apparently "... regard molecular connectivity as an irrelevance which has had the unfortunate effect of diverting attention from real work that need doing". Kubinyi apparently subscribe to such opinion by stating: "nothing can be added to this criticism. To whom it concerns".

   In our view the paper of Dunn et al. is an illustration how not to use connectivity indices and is misrepresented as an early "warning" on use of the connectivity index. The connectivity index, as is clear from its construction, is inherently a bond additive quantity. The *ortho–meta–para* effects, as is well-known are not bond additive. Hence, to use a descriptor designed for bond additivity to "explain" a property that is not bond additive, shows a lack of familiarity with the nature of this mathematical descriptor. To repeat this elementary misunderstanding some 20 years later [9] shows not only lack of familiarity with the mathematical characterization of chemical structure, but also a lack of willingness to understand the nature of mathematical characterization of chemical structure. To whom it concerns.

   We should however add that the connectivity indices were well-received in many other QSAR circles even if they were not widely used. For example, Franke [61] in his book on "Theoretical Drug Design Methods" writes: "the advantage of χ is that its calculation for practically any structure is easy and straightforward ... The use of connectivity can be recommended especially in such cases where relatively large structural variations are present in a series of compounds and where the primary goal of a QSAR analysis is data description rather than interpretation".
2. Often selection of descriptors from a large pool is constrained by excluding descriptors that show when used alone limited correlation with the property considered. However, such descriptors when combined with other descriptors may nevertheless lead to a high quality regression. An illustration is offered by MR of alkanes [62] which is not well-described either by ${}^1\chi$ or ${}^2\chi$, the coefficient of regression (r) being $r = 0.087$ and 0.187, respectively. However, when both "useless" descriptors are employed together we obtain a respectable regression the statistical parameters: $r = 0.971$.
3. Similarly, descriptors that show high correlation with already selected descriptors are often eliminated from structure–property–activity studies. They should not be. The *only* useful criterion for discarding a descriptor is its inability to reduce the standard error of the regression. For example, in several applications of connectivity indices, the second order connectivity index ${}^2\chi$ has been discarded because for compounds considered it shows close parallelism to the connectivity index ${}^1\chi$. But as we have seen in the previous illustration ${}^2\chi$ despite its parallelism to ${}^1\chi$ also *complement* it. That is, a part of ${}^2\chi$ which is *different* from ${}^1\chi$ (and which may be small) suffices to produce satisfactory regression.

   A referee has raised a serious objection to the above statement concerning employment of highly interrelated descriptors: "two highly correlated descriptors should not be used to establish a model based on statistical methods like regression. Such a model is generally not predictive, that is, when new compounds are predicted, their presence essentially alters the interrelation between the two descriptors, ${}^1\chi$ and ${}^2\chi$ in this example. A model cannot be judged as sound from the direct statistics alone. Some form of cross-validation or external validation must be used. Often when models using inter-correlated variables are used, they do not produce good validation statistics". While most of this criticism may hold in general, particularly the portion concerning a predicting power of MRLA, we would like to add that use of all available data in structure–property regressions is also a legitimate approach to modeling when one is not interested in prediction, but in discerning structural elements contributing or being responsible for a particular property. For example, if all the boiling points of smaller alkanes are known, there is nothing to be predicted. Cross-validation may be used here to test the robustness of the model, but once this has been established, if one is

Table 10
Experimental and calculated molar refractions for octane isomers based on highly interrelated connectivity indices $^1\chi$ and $^2\chi$

		MR		
		Experimental	Calculated	Cross-validation
1	*n*-Octane	39.194	39.238	39.259
2	2-Methylheptane	39.234	39.214	39.210
3	3-Methylheptane	39.102	39.071	39.067
4	4-Methylheptane	39.119	39.108	39.106
5	3-Ethylhexane	38.946	38.996	39.007
6	2,2-Dimethylhexane	39.255	39.293	39.306
7	2,3-Dimethylhexane	38.983	38.959	38.958
8	2,4-Dimethylhexane	39.132	39.063	39.058
9	2,5-Dimethylhexane	39.261	39.189	39.179
10	3,3-Dimethylhexane	39.011	39.034	39.036
11	3,4-Dimethylhexane	38.864	38.810	38.799
12	2-Methyl-3-ethylhexane	38.838	38.878	38.883
13	3-Methyl-3-ethylhexane	38.719	38.775	38.788
14	2,2,3-Trimethylpentane	38.927	38.936	38.939
15	2,2,4-Trimethylpentane	39.264	39.295	39.320
16	2,3,3-Trimethylpentane	38.764	38.798	38.809
17	2,3,4-Trimethylpentane	38.870	38.824	38.815

interested to know, for example, if the peripheral bonds like CH_3–CH_2, and CH_3–CH make bigger contributions than bonds like CH_2–CH_2 and CH_2–CH, why should one not use all the available data?

In order to respond to the main objection concerning use of highly inter-correlated descriptors as are $^1\chi$ and $^2\chi$ we show in Table 10 the experimental MR of octane isomers, the calculated MR by use of both $^1\chi$ and $^2\chi$ as descriptors as well as MR calculated by cross-validation. As we can see the calculated cross-validated MR values (obtained by leave-one-out procedure) are almost as good as the values obtained by using all the data on MR for a regression. The standard error for the plot of MR (calculated) against MR (experimental) is 0.044, while the standard error for the plot of MR (cross-validated) against MR (experimental) is 0.054. The corresponding coefficients of regressions are 0.9708 and 0.9569, respectively, which point to very good regressions, in view that we consider molecules of the same size. As we said before neither $^1\chi$ nor $^2\chi$ when used alone show any significant correlation, but only when combined produce a good correlation with MR. As is known the two connectivity indices $^1\chi$ and $^2\chi$ show high degree or inter-relation. In the case of octane isomers (which are molecules having the same size) the coefficient of regression is 0.9757. Hence, the opinion that highly interrelated descriptors do not produce good validation statistics simply does not hold, at least not in the case considered.

It is interesting to add that although neither $^1\chi$ nor $^2\chi$ are good descriptors when $^2\chi$ is made orthogonal to $^1\chi$ we obtain novel descriptor $^2\Omega$, which, as Wu [63] has recently demonstrated, gives a very good regression when used alone. The regression coefficient for MR of octanes when $^2\Omega$ is used is 0.9670 and the standard error is 0.048.

4. Use of *the same set of descriptors* may be advantageous in comparative studies of different properties for the same set of compounds [34,35]. Descriptors optimally selected for each property separately may not reveal that two properties are related. In such cases small variations in the experimental data, or absence of a few data in one set, may result in selecting a quite different set of descriptors as optimal for each case. This is illustrated by a study of Kier and Hall [64] who found different sets of connectivity indices as best for regression of the heat of formation ΔH_f and the heat atomization ΔH_a of alkanes:

$$\Delta H_f = 1.15\,^1\chi - 2.53\,^2\chi + 7.63\,^3\chi - 12.02\,^4\chi - 1.72\,^5\chi + 0.89\,^4\chi_{PC} - 1.46\,^5\chi_{PC} - 0.28$$

$$\Delta H_a = 286.15n - 12.08\,^1\chi - 0.92\,^4\chi + 1.50\,^5\chi - 2.44\,^5\chi_C + 0.86\,^4\chi_{PC} - 0.50\,^5\chi_{PC} - 1.42\,^6\chi_{PC} + 114.65$$

Even when only two descriptors, n and $^1\chi$, are used for regression no relationship between ΔH_a and ΔH_f is apparent from inspection of the respective regression equations despite that the same descriptors have been selected:

$$\Delta H_a = 283.33n - 6.321\,^1\chi + 115.72 \quad \text{and}$$
$$\Delta H_f = 7.649n - 3.286\,^1\chi + 11.70$$

It is only when we plot ΔH_a against ΔH_f that we see that points belonging to isomers lie on the same horizontal lines (Fig. 3). Hence, when we restrict regression to molecules having the same size we find that ΔH_a and ΔH_f are collinear. For octane isomers, for example, we obtain:

$$\Delta H_a = \Delta H_f + 2308.12$$

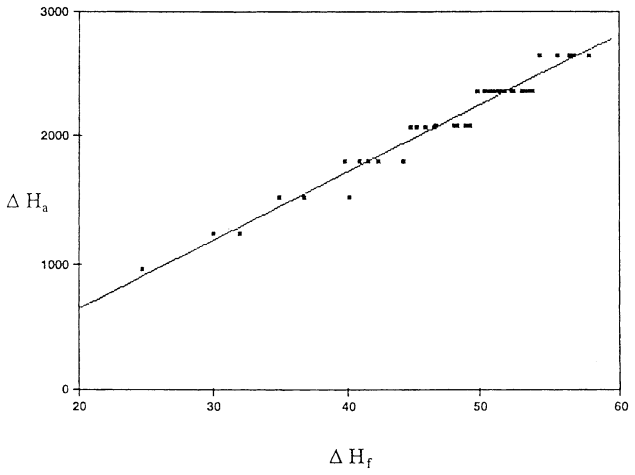

Fig. 3. Correlation of the heats of atomization against the heats of formation for smaller alkanes.

The prime purpose of discussing the above illustration is not to argue that relationship of ΔH_f and ΔH_a could not be established from data from combustion calorimetry, but to show that when different descriptors are used relationship between quantities considered, whether they are collinear or not, could be obscured.

5. When many descriptors are used multiple solutions may emerge. While this in itself is not necessarily a disadvantage, it leaves considerable room for subjective or arbitrary selection of the most "preferred" solution. In such situations one should re-examine the model used and restrict descriptors according the model under consideration *in advance*.

9. Remedies

Here we will outline possibilities for improving MLRA.

1. Restrict in advance the type of descriptors to be used, preferably to families of descriptors that have structurally related meaning. For example, use the path numbers, the walks of different length, the path/walk quotients, Kier's kappa indices of different degree, the connectivity indices of different order and such.
2. To arrive at stable regression equations construct orthogonal combinations of descriptors used [65–68]. The chaotic behavior of the coefficients of MLRA regression equations is caused by interdependence of descriptors.
3. Use as few descriptors as possible. One can sometimes reduce the number of descriptors by use of descriptors orthogonal to descriptors *not used* in MLRA. For example, as we have seen though ${}^1\chi$ and ${}^2\chi$ as single descriptors do not correlate with MR of octanes, when ${}^2\chi$ is orthogonalized to ${}^1\chi$ one obtains a new descriptor which alone gives simple regression with MR of high correlation coefficient [63].
4. Use retro-regression [69] to arrive at stable regression equations when different steps in a stepwise regression introduce different descriptors. This is the case with the illustration shown in Table 11. Suppose we consider the regression based on five descriptors (step 5 in Table 11) as a solution. View the process that has led from the steps (1) to (5) as the "history" of arriving at acceptable solution. Now consider the retro-regression by eliminating at each step the least important descriptor. This will lead to an ordering of descriptors that will allow subsequent descriptors to be orthogonalized. In this way stable regression equations will be obtained.

10. More on collinearity

1. If one uses instead of descriptors d_1 and d_2 their linear combinations $(d_1 + d_2)$ and $(d_1 - d_2)$ MLRA will give the same statistical results for both cases. However, linear combinations, such as $(d_1 + d_2)$ and $(d_1 - d_2)$, will have different structural interpretation than descriptors d_1 and d_2. One should keep this in mind when interpreting the results of a regression. Moreover, linear combination of descriptors may be of interest when one considers only some such combinations, and discards others. For example, Randić [70] found that the difference ${}^1\chi - {}^2\chi$ is a good descriptor for some properties of octanes for which ${}^1\chi$ and ${}^2\chi$ are not so satisfactory. Similarly, Kier and Hall [71] considered the differential molecular connectivity index, $D^m\chi$, defined as the difference between the connectivity index and the corresponding valence con-

Table 11
The results of a stepwise regression using the connectivity indices based on exhaustive search of the best combinations (due to Lučic and Trinajstic) (r is the coefficient of regression, s the standard error and F the Fisher ratio)

n	Descriptors	r	s	F
1	${}^2\chi$	0.776	3.92	50
2	${}^0\chi, {}^3\chi$	0.900	2.76	68
3	${}^1\chi, {}^4\chi, {}^5\chi$	0.956	1.88	110
4	${}^0\chi, {}^1\chi, {}^3\chi, {}^5\chi$	0.966	1.70	103
5	${}^0\chi, {}^3\chi, {}^4\chi, {}^5\chi, {}^6\chi$	0.968	1.67	85
6	${}^0\chi, {}^1\chi, {}^3\chi, {}^4\chi, {}^5\chi, {}^6\chi$	0.968	1.69	70
7	${}^0\chi, {}^1\chi, {}^2\chi, {}^3\chi, {}^4\chi, {}^5\chi, {}^6\chi$	0.968	1.71	58

nectivity index: ${}^{m}\chi - {}^{m}\chi^{v}$. They found that the information contained in this index is largely electronic, capable of modeling selected physical properties which are regiospecific within a molecule.

2. Linear combinations of linearly independent descriptors may produce collinear descriptors. Collinear descriptors fully duplicate regression results. Such descriptors may have different mathematical form and may capture distinct structural aspects of a molecule, and therefore it need not be apparent that they are collinear. This is illustrated by the Wiener index W and the "reverse Wiener" index M [23]. The reverse Wiener index is constructed by reversing the weights so that shorter paths have larger weight than longer paths. One way of computing the Wiener number is as the sum of weighted paths:

$$W = 1p_1 + 2p_2 + 3p_3 + 4p_4 + \cdots$$

where the weights 1, 2, 3, 4, ... increase as the path lengths increase. Symbol p_k gives the number of paths of length k in the molecule considered. In an analogous manner M can be defined as the sum of weighted paths:

$$M = np_1 + (n-1)p_2 + (n-2)p_3 + (n-3)p_4 + \cdots$$

This formula gives different numerical values for the characterization of the same molecule. However, when W and M are calculated for a set of isomeric molecules and tested in a regression, somewhat unexpectedly, one obtains the same regression statistics. One can show that indeed for isomers of octanes W and M are strictly collinear. Thus the two descriptors are interchangeable. This then also means that the apparent "success" of the Wiener index in a regression may be due to the reversed Wiener index M, which may be *crucial* for interpretation of the results.

3. Satisfactory regression equation may be based on descriptors associated with a preconceived molecular model in which *individual* descriptors d_i may have suitable interpretation. However, one should recognize that the set of descriptors $\{d_i\}$ that define regression equation also define a *structure–subspace* in which the considered property is characterized. Any linearly independent set of descriptors $\{c_1d_1 + c_2d_2 + c_3d_3 + \cdots\}$ that *span* the same structure–space can therefore equally well serve as the *basis* for the same subspace. Hence, alternative set of descriptors can with equal right be considered for interpretation of regression equations. Too much attention in the past in MLRA has been given to the *individual* molecular descriptors selected as the best. Instead, one should focus attention on the corresponding *subspace* for which one can find more than one alternative characterization.

11. Linear versus non-linear regressions

Many correlations when involving molecules of different size, show a departure from a linear model. In such cases a quadratic regression may offer a better description of the relationship than a linear model. There are no fundamental reasons that tell that a linear model should be given preference over quadratic or any other functional dependence of property on a descriptor. Hence, when comparison of results of structure–property–activity regressions based on different descriptors is considered it is unwarranted, and it may be misleading, to confine such comparison to linear models only. For example, recently Ren [72] introduced topological index, Xu, based on the adjacency matrix and the distance matrix. Using this index and the connectivity index he made a comparison for several physico-chemical properties of smaller alkane between the two using *linear regression*. In case of octane isomers Ren reported Xu as a better descriptor than ${}^{1}\chi$ for the quadratic mean radius, the density, the Pitzer eccentric factor, the critical volume and the octane numbers. For entropy, the heats of vaporization and the heats of formation, the connectivity index ${}^{1}\chi$ gave better results. However, when one compares results for molecules whose size may vary considerably the difference between *linear* and *non-linear* regression for *different* descriptors can be considerable. We will illustrate this for the boiling points in alkanes as reported by Ren [72]. The following results were obtained when Xu and ${}^{1}\chi$ indices were used in a linear model:

$$\text{bp} = -80.7898 + 6.6425\,\text{Xu};$$
$$r = 0.993, \quad s = 5.79, \quad F = 2615$$

$$\text{bp} = -130.3039 + 67.63286\,{}^{1}\chi;$$
$$r = 0.987, \quad s = 7.91, \quad F = 1428$$

It appear thus that Xu is a better descriptor for the boiling points of smaller alkanes than ${}^{1}\chi$. However, when we compare quadratic correlations for the same set of compounds and the same descriptors we obtain:

$$\text{bp} = -90.9607 + 68.8983\,\text{Xu} - 2.8276(\text{Xu})^2;$$
$$r = 0.995, \quad s = 5.01, \quad F = 1816$$

$$\text{bp} = -199.9927 + 125.2571\,{}^{1}\chi - 10.8299({}^{1}\chi^2);$$
$$r = 0.997, \quad s = 4.13, \quad F = 2680$$

Hence, as we see the connectivity index is better descriptor for the boiling points of alkanes when one goes beyond linear regressions.

12. On interpretation of linear combinations of descriptors

Linear combination of two descriptors represents a superposition of the characterizations originating from two descriptors. If we are able to partition contributions of individual descriptors used in a linear combination into various

Table 12
Contributions of individual C–C bonds in heptane isomers C_7H_{16} to the connectivity index constructed as various linear combinations of $^1\chi$ and $^2\chi$

	C_1–C_2	C_2–C_3	C_3–C_4	C_4–C_5	C_5–C_6	C_3–C_7
$1.0^1\chi + 0.0^2\chi$	0.70711	0.40825	0.40825	0.50000	0.70711	0.57735
$0.9^1\chi + 0.1^2\chi$	0.65681	0.42268	0.42268	0.48943	0.66140	0.56044
$0.8^1\chi + 0.2^2\chi$	0.60651	0.43712	0.43712	0.47887	0.61568	0.54353
$0.7^1\chi + 0.3^2\chi$	0.55621	0.45155	0.45155	0.46830	0.56997	0.52662
$0.6^1\chi + 0.4^2\chi$	0.50591	0.46598	0.46598	0.45773	0.52426	0.50971
$0.5^1\chi + 0.5^2\chi$	0.45562	0.48042	0.48042	0.44717	0.47856	0.49280
$0.4^1\chi + 0.6^2\chi$	0.40532	0.49485	0.49485	0.43660	0.43284	0.47589
$0.3^1\chi + 0.7^2\chi$	0.35502	0.50928	0.50928	0.42604	0.38713	0.45898
$0.2^1\chi + 0.8^2\chi$	0.30472	0.52372	0.52372	0.41547	0.34142	0.44207
$0.1^1\chi + 0.9^2\chi$	0.25442	0.53815	0.53815	0.40490	0.29571	0.42516
$0.0^1\chi + 1.0^2\chi$	0.20412	0.55259	0.55259	0.39434	0.25000	0.40825

bond contributions the superposition of these contributions would give contributions of the linear combination as a molecular descriptor. Hence, we can consider linear combination $(a\,^1\chi + b\,^2\chi)$ as a single descriptor, with various combinations of the coefficients a and b have been selected in advance.

In Table 12 we illustrate for C–C bonds of 3-methylhexane the contributions arising from linear combinations of $^1\chi$ and $^2\chi$ by varying the coefficients a and b in steps of 0.1. As we move from the top of the table to the bottom of the table we gradually see how $^1\chi$ transforms into $^2\chi$. The linear combinations of Table 12 can be viewed as a novel molecular descriptor for which we can use labels $^{1.1}\chi$, $^{1.2}\chi$, $^{1.3}\chi$, $^{1.4}\chi$ and so on, where the fractional superscript indicates the content of the admixture of the two descriptors. So $^{1.1}\chi$ is in magnitude and in bond partition close to $^1\chi$ while $^{1.9}\chi$ is close to $^2\chi$. Observe also, that despite that $^1\chi$ and $^2\chi$ are highly interrelated, a view that these indices practically duplicate each other is not correct, because the two indices show quite a distinct partitions of molecular indices into individual C–C bond contributions. Thus in the case of $^1\chi$ the peripheral bonds like C_1–C_2, C_5–C_6 and C_3–C_7 make the largest contributions, but in the case of $^2\chi$ the largest contributions comes from the internal bonds C_2–C_3 and C_3–C_4.

In Table 13 we have listed the variable linear combinations for the nine isomers of heptane and their boiling points. As we see, except for 2,2-dimethyl hexane and 2,2,3-trimethylpentane, all the variable indices decrease as we move from $^1\chi$ towards $^2\chi$, while the magnitudes of the variable indices for 2,2-dimethyl hexane and 2,2,3-trimethylpentane increase. In this way the relative positions of the points representing different isomers change during the transformation and the best linear combination (of a set of discrete linear combinations of descriptors considered) for a particular structure–property relationship can be found. In Table 14 we have listed the statistical parameters for all the linear combinations of Table 13 for a regression of the boiling points in heptane isomers to illustrate variations in the quality of the regression with the variations of the descriptors. Because $^2\chi$ gives a slightly better regression than $^1\chi$ one would think that by adding to $^1\chi$ a small amount of $^2\chi$ we will get a better regression. But as we see from Table 14 this is not the case. In varying $^1\chi$ the initial steps make worse the regression and this is because the slope of the regressions of bp for $^1\chi$ and $^2\chi$ are of opposite signs. Indeed, if we would vary $^1\chi$ in smaller steps we would come to a regression when the coefficient of the regression is zero (somewhere close to $^{1.25}\chi$). By continuing to vary $^1\chi$ we gradually improve the regression and reach the minimum at $^{1.6}\chi$.

As we have seen from Table 12 linear combination of topological indices can be viewed as a novel index which will assign to individual bonds weights that are different from those of the component descriptors. Thus if a property is bond additive and if descriptors can be partitioned into

Table 13
The variable connectivity index for heptane isomers C_7H_{16} based on various linear combinations of $^1\chi$ and $^2\chi$

	$^{1.1}\chi$	$^{1.2}\chi$	$^{1.3}\chi$	$^{1.4}\chi$	$^{1.5}\chi$	$^{1.6}\chi$	$^{1.7}\chi$	$^{1.8}\chi$	$^{1.9}\chi$	Boiling point
n-Heptane	3.27886	3.14350	3.00815	2.87279	2.73744	2.60208	2.46673	2.33137	2.19602	98.43
2-Methylhexane	3.19666	3.12326	3.04986	2.97646	2.90307	2.82967	2.75627	2.68287	2.60947	90.05
3-Methylhexane	3.20746	3.10687	3.00627	2.90567	2.80508	2.70448	2.60389	2.50329	2.40269	91.85
3-Ethylpentane	3.22054	3.09501	2.96948	2.84395	2.71842	2.59289	2.46736	2.34183	2.21630	93.48
2,2-Dimethylpentane	3.08513	3.11019	3.13524	3.16030	3.18536	3.21042	3.23548	3.26054	3.28560	79.20
2,3-Dimethylpentane	3.12562	3.07050	3.01538	2.96026	2.90514	2.85003	2.79491	2.73979	2.68467	89.78
2,4-Dimethylpentane	3.11565	3.10540	3.09515	3.08490	3.07464	3.06440	3.05415	3.04390	3.03365	80.50
3,3-Dimethylpentane	3.09632	3.07132	3.04632	3.02132	2.99632	2.97132	2.94632	2.92132	2.89632	86.06
2,2,3-Trimethylbutane	3.00111	3.05885	3.11658	3.17432	3.23205	3.28979	3.34752	3.40526	3.46299	80.88

Table 14
The statistical parameter r, s, and F, and the regression equation for heptane isomers C_7H_{16} obtained by using various linear combinations of $^1\chi$ and $^2\chi$

Descriptor	r	s	F	Equation
$^1\chi$	0.9208	2.753	39.01	$40.3948\,^1\chi - 41.3244$
$^{1.1}\chi$	0.8872	3.256	25.89	$68.7742\,^{1.1}\chi - 128.6620$
$^{1.2}\chi$	0.4585	6.273	1.86	$110.5354\,^{1.2}\chi - 254.6706$
$^{1.3}\chi$	0.9009	3.063	30.17	$-106.2104\,^{1.3}\chi + 411.6557$
$^{1.4}\chi$	0.9451	2.307	58.55	$-51.9734\,^{1.4}\chi + 243.7237$
$^{1.5}\chi$	0.9490	2.226	63.41	$-33.6924\,^{1.5}\chi + 187.2242$
$^{1.6}\chi$	0.9493	2.218	63.86	$-24.8450\,^{1.6}\chi + 159.8956$
$^{1.7}\chi$	0.9491	2.224	63.50	$-19.6590\,^{1.7}\chi + 143.8810$
$^{1.8}\chi$	0.9487	2.231	63.03	$-16.2579\,^{1.8}\chi + 133.3801$
$^{1.9}\chi$	0.9484	2.2385	62.59	$-13.8576\,^{1.9}\chi + 125.9698$
$^2\chi$	0.9481	2.245	62.21	$-12.0736\,^2\chi + 120.4630$

bond contributions a linear combination of such descriptors represents a novel descriptor which may leads to better relative weights for the contributing bonds.

13. Regressions using variable connectivity index

In order to come with visibly better regressions we need dramatic improvement in the design of molecular descriptors. One such promising direction has been outlined already about 10 years ago [49,50] but as is often the case, it has been overlooked for too long. Recently an effort was made to advertise the variable connectivity index by considering properties of several families of compounds in which performance of the variable connectivity index was illustrated [51–53,73–75]. In addition, also paths with variable weights have been considered [76–79], and a variable Balaban's index J [58]. What typifies these novel descriptors is the presence of molecular parameters that are determined *during* the search for best correlation. In this respect these descriptors are *fundamentally* different from hundreds of topological indices which are all numerically fixed quantities, once molecular structure is taken.

Usually one starts by selecting zero values for the variables $x, y, z, \ldots$ as the initial step. It is clear that use of variable topological indices can only improve correlations over the use of simple indices, since it is unlikely that $x =$

Table 15
Indicator variables, topological indices, and information theoretic indices available in CODESSA software

Indicator variables	Topological indices	Information theoretic indices
Number of atoms	Wiener index	Average information content (order 0)
Number of C atoms	Randić index (order 0)	Information content (order 0)
Relative number of C atoms	Randić index (order 1)	Average structural information content (order 0)
Number of H atoms	Randić index (order 2)	Structural information content (order 0)
Relative number of H atoms	Randić index (order 3)	Average complementary information content (order 0)
Number of O atoms	Kier and Hall index (order 0)	Complementary information content (order 0)
Relative number of O atoms	Kier and Hall index (order 1)	Average bonding information content (order 0)
Number of bonds	Kier and Hall index (order 2)	Bonding information content (order 0)
Number of single bonds	Kier and Hall index (order 3)	Average information content (order 1)
Relative number of single bonds	Kier shape index (order 1)	Information content (order 1)
Number of double bonds	Kier shape index (order 2)	Average structural information content (order 1)
Relative number of double bonds	Kier shape index (order 3)	Structural information content (order 1)
Molecular weight	Kier flexibility index	Average complementary information content (order 1)
Relative molecular weight	Balaban index	Complementary information content (order 1)
Gravitation index (all bonds)		Average bonding information content (order 1)
Gravitation index (all pairs)		Bonding information content (order 1)
		Average information content (order 2)
		Information content (order 2)
		Average structural information content (order 2)
		Structural information content (order 2)
		Average complementary information content (order 2)
		Complementary information content (order 2)
		Average bonding information content (order 2)
		Bonding information content (order 2)

Table 16
The comparison of the standard errors for several simple regressions based on the connectivity index $^1\chi$ and the variable connectivity index $^1\chi^f$

Compounds	Property	$^1\chi$	$^1\chi^f$
Alcohols	Boiling points	7.86	3.30
Amines	Boiling points	3.488	1.907
Smaller alkanes	Boiling points	2.928	2.481
Alkanes + cycloalkanes	Boiling points	4.15	3.18
Sulfides	Boiling points	2.71	1.326
Alcohols	Toxicity in mice	0.1297	0.0957
Alkane + alcohols	Retention indices	56.44	14.24

$y = z = \cdots = 0$ is an optimal solution. Hence, we expect that variable topological indices will improve the results based on the corresponding "fixed" (traditional) topological indices. They can only do better. How much better?

In a recent study Randić and Pompe [80] undertook to compare the performance of the variable connectivity index $^1\chi^f$ (not included in CODESSA) with topological indices available from CODESSA. CODESSA is a program that calculated several hundreds of topological, quantum chemical, and information theoretic descriptors developed by Katritzky et al. [81] (Table 15). We selected the boiling points of $n = 100$ alcohols as the property and examined the pool of 56 topological indices (listed in Table 12). CODESSA gives the following best results.

Number of descriptors	r	s	F
1	0.9615	8.480	1201
2	0.9857	5.232	1659
3	0.9870	5.024	1202
4	0.9901	4.410	1177
5	0.9945	3.389	1708

Using the variable connectivity index with weights $x = 0.10$ and $y = -0.92$ for carbon and oxygen atom, respectively, the following regression was obtained:

$$\text{bp} = 38.7599\,^1\chi^f - 40.4289$$

with the correlation coefficient $r = 0.9915$, the standard error $s = 4.018$, and Fisher ratio $F = 5691$. As we see a single (variable connectivity) index gave a result that is better than those based on the four best CODESSA descriptors. In Table 13 we show results obtained using variable molecular connectivity index applied to different properties of alkanes, cycloalkanes, alcohols, amines and sulfides. In Table 14 we show results obtained using variable molecular connectivity index applied to different properties of amino acids. Here we see that different properties do require different weights in calculations of the variable connectivity index. The infinite weight in fact means that the particular atom (e.g. oxygen for crystal density) does not make visible contribution to the correlation.

14. Concluding remarks

The connectivity index has passed several transformations since its early introduction. It became, together with other connectivity indices, a standard for comparison of performance of many newly proposed indices. There is no doubt that variable connectivity indices will do just the same as a new generation of variable molecular descriptors emerges (Table 16). The variable connectivity index, as was shown, can replace several "traditional" descriptors, such as are descriptors available by CODESSA (like those of Table 12) and other software. Use of fewer descriptors has important advantages when constructing regression equations, one of which is to facilitate interpretation of descriptors. In this respect it is interesting that Pompe and Randić [51] have recently shown that the relative weights given to oxygen in alcohols, ethers, esters and ketones in the construction of optimal variable connectivity indices correlated well with the relative charges on oxygen atoms as computed by a quantum mechanical scheme (Table 17).

By no means one should deduce that the "traditional" indices have no future and that variable descriptors will eventually replace them. This clearly will be the case in

Table 17
The optimal weight for the variable connectivity index for different properties of amino acids (top part) and the resulting statistical parameters (from Randić et al., to be published)

Property	x	y	z	w	n	r	s	r	s
Partial molar volume	−0.65	5.20	0.20	0.50	18	0.9776	5.861	0.9168	11.118
Crystal density	−0.98	∞	−0.94	0.14	10	0.9801	0.040	0.4184	0.184
Side-chain molecular volume	−0.48	19	+0.86	NA	18	0.9737	5.977	0.8731	12.79
Partition coefficient	−0.82	∞	−0.63	0.30	10	0.9439	0.269	0.672	0.5651
pH at the isoelectric point	−0.93	∞	−0.995	∞	20	0.7325	1.238	0.1925	1.784
Longitudinal relaxation rates	+0.24	3.50	−0.84	NA	8	0.9739	0.044	0.7583	0.125

Table 18
The range of values for a selection of topological indices used by Lahana and coworkers in screening combinatorial library of 280,000 fictitious compounds

Descriptor	Minimum	Maximum
Kappa alpha 2	26.1	44.3
Flexibility	22.5	40.3
Kier Chi v4	3.325	5.342
Balaban index	2.846	5.342
Kappa 1	56.1	93.0
Kappa 2	29.5	51.2
Kappa 3	22.1	41.8
Kappa alpha 1	51.8	86.9
Kappa alpha 3	19.3	37.2
Randić index	54.2	87.6
Wiener index	86872.0	312008.0
E-state	160.5	268.0

some application, but in some studies the traditional indices may maintain their advantage. One such illustration concerns screening of combinatorial libraries. Not long ago Lahana [82] and coworkers were able to screen a combinatorial library having over 280,000 virtual compounds based on a lead compound showing immunosuppressive activity (Table 18). They first set the lower and upper bound for some two dozen topological indices and molecular descriptors based on the leading compound (shown in Table 15). They were then able to narrow selection to about 20 compounds from which, after additional analysis four received special scrutiny and were synthesized and tested. Lahana and coworkers in this way found a compound that produced biological activity about 100 times higher than the initial lead compound. The complexity of this particular study lies in the size of the combinatorial library. It is clear that variable indices, just as other quantum chemical models, at least for now, may not be suitable for such analysis as they are more computer intensive. The simplicity of computation of topological indices, which perhaps today is not as critical a factor as it may have been 10–20 years ago, again is becoming an important factor, when considering combinatorial libraries. Hence, it seems that despite emergence of the variable connectivity index ${}^1\chi^f$, the "simple" connectivity index ${}^1\chi$ will continue to enjoy attention of researchers for some time in the future.

Acknowledgements

The author wishes to thank the referees who made numerous suggestions and raised some interesting questions, consideration of which has helped to improve the presentation of the work.

References

[1] E.B. Wilson, Introduction to Scientific Research, McGraw-Hill, New York, 1952.
[2] L.B. Kier, W.J. Murray, M. Randić, L.H. Hall, Molecular connectivity. I. Relationship to nonspecific local anesthesia, J. Pharm. Sci. 64 (1975) 1971–1974.
[3] C. Hansch, A. Leo, Exploring QSAR: Fundamentals and Applications in Chemistry and Biology, American Chemical Society, Washington, DC, 1995.
[4] M. Randić, On the characterization of molecular branching, J. Am. Chem. Soc. 97 (1975) 6609–6615.
[5] L.B. Kier, W.J. Murray, M. Randić, L.H. Hall, Molecular connectivity. V. Connectivity series applied to density, J. Pharm. Sci. 65 (1975) 1226–1230.
[6] L.B. Kier, L.H. Hall, Molecular connectivity. VII. Specific treatment of heteroatoms, J. Pharm. Sci. 65 (1975) 1806–1809.
[7] L.B. Kier, L.H. Hall, Molecular Connectivity in Chemistry and Drug Research, Academic Press, New York, 1876.
[8] L.B. Kier, L.H. Hall, Molecular Connectivity in Structure–Activity Analysis, Research Studies Press, Letchworth, 1986.
[9] H. Kubinyi, Hansch Analysis and Related Approaches, VCH, Weinheim, Germany, 1993.
[10] N. Trinajstić, Chemical Graph Theory, CRC Press, Boca Raton, 1992, pp. 225–273.
[11] C. Hansch, A. Leo, Exploring QSAR: Fundamentals and Applications in Chemistry and Biology, American Chemical Society, Washington, DC, 1995.
[12] D.J. Livingstone, The characterization of chemical structures using molecular properties: a survey, J. Chem. Inf. Comput. Sci. 40 (2000) 195–209.
[13] M. Randić, S.C. Basak, Multiple regression analysis with optimal molecular descriptors, SAR QSAR Environ. Res. 11 (2000) 1–23.
[14] M. Randić, Topological Indices, in: P.V.R. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.F. Schaefer III, P.R. Schreiner (Eds.), The Encyclopedia of Computational Chemistry, Wiley, Chichester, 1998, pp. 3018–3032.
[15] A.T. Balaban, Historical developments of topological indices, in: J. Devillers, A.T. Balaban (Eds.), Topological Indices and Related Descriptors in QSAR and QSPR, Gordon and Breach Science Publication, Amsterdam, 1999, pp. 21–57.
[16] S.C. Basak, Information theoretic indices of neighborhood complexity and their applications, in: J. Delivers, A.T. Balaban (Eds.), Topological Indices and Related Descriptors in QSAR and QSPR, Gordon and Breach Science Publication, Amsterdam, 1999, pp. 563–593.
[17] M. Randić, In search of structural invariants, J. Math. Chem. 9 (1992) 97–146.
[18] H. Wiener, Structural determination of paraffin boiling points, J. Am. Chem. Soc. 69 (1947) 17–20.
[19] H. Hosoya, Topological index: a newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons, Bull. Chem. Soc. Jpn. 44 (1971) 2332–2339.
[20] M.I. Skvortsova, I.I. Baskin, O.L. Slovokhotova, V.A. Palyulin, N.S. Zefirov, Inverse problem in QSAR/QSPR studies for the case of topological indices characterizing molecular shapes (Kier indices), J. Chem. Inf. Comput. Sci. 33 (1993) 630–634.
[21] V. Kvasnička, J. Poshipal, Canonical indexing and constructive enumeration of molecular graphs, J. Chem. Inf. Comput. Sci. 30 (1990) 99–105.
[22] L.B. Kier, L.H. Hall, J.W. Frazer, Design of molecules from quantitative structure–activity relationship models. 1. Information transfer between path and vertex degree counts, J. Chem. Inf. Comput. Sci. 33 (1993) 143–147.
[23] M. Randić, Linear combinations of path numbers as molecular descriptors, New J. Chem. 21 (1997) 945–951.
[24] A.T. Balaban, Highly discriminating distance-based topological index, Chem. Phys. Lett. 89 (1982) 399–404.
[25] A.T. Balaban, Topological index based on topological distances in molecular graphs, Pure Appl. Chem. 55 (1983) 199–206.

[26] M. Randić, X. Guo, T. Oxley, H. Krishnapriyan, Wiener matrix: source of novel graph invariants, J. Chem. Inf. Comput. Sci. 33 (1993) 709–716.
[27] M. Randić, X. Guo, T. Oxley, H. Krishnapriyan, L. Naylor, Wiener matrix invariants, J. Chem. Inf. Comput. Sci. 34 (1994) 361–367.
[28] M. Randić, Hosoya matrix — a source of new molecular descriptors, Croat. Chem. Acta 67 (1994) 415–429.
[29] M. Randić, Restricted random walks on a graph, Theor. Chim. Acta 92 (1995) 97–106.
[30] M.V. Diudea, O.M. Minailiuc, G. Katona, I. Gutman, Szeged matrices and related numbers, Math. Chem. 35 (1997) 129–143.
[31] M.V. Diudea, Cluj matrix: source of various graph descriptors, Math. Chem. 35 (1997) 300–305.
[32] M.V. Diudea, B. Parv, I. Gutman, Detour-Cluj and derived invariants, J. Chem. Inf. Comput. Sci. 37 (1997) 1101–1108.
[33] M. Randić, On molecular branching, Acta Chim. Slovenica 44 (1997) 57–77.
[34] D.E. Needham, I.-C. Wei, P.G. Seybold, Molecular modeling of the physical properties of alkanes, J. Am. Chem. Soc. 110 (1988) 4186–4194.
[35] M. Randić, P.G. Seybold, Molecular shape as a critical factor in structure–property–activity studies, SAR QSAR Environ. Res. 1 (1993) 77–85.
[36] J.R. Platt, Influence of neighbor bonds on additive bond properties in paraffins, J. Chem. Phys. 15 (1947) 419.
[37] J.R. Platt, Prediction of isomeric differences in paraffin properties, J. Phys. Chem. 56 (1952) 328–336.
[38] M. Randić, C.L. Wilkins, On graph theoretical basis for ordering of structures, Chem. Phys. Lett. 63 (1979) 332–336.
[39] M. Randić, Characterization of atoms, molecules, and classes of molecules based on path enumeration, Math. Chem. 7 (1979) 3–60.
[40] M. Randić, G.M. Brissey, R.G. Spencer, C.L. Wilkins, Search for all self-avoiding paths for molecular graphs, Comput. Chem. 3 (1979) 5–13.
[41] M. Randić, Chemical structure — what is she? J. Chem. Educ. 69 (1992) 713–718.
[42] L.B. Kier, Shape indexes or orders one and three from molecular graphs, Quant. Struct. Act. Relat. 5 (1986) 1–7.
[43] L.B. Kier, Distinguishing atom differences in a molecular graph shape index, Quant. Struct. Act. Relat. 5 (1986) 7–12.
[44] M. Randić, On characterization of shape of molecular graphs, J. Chem. Inf. Comput. Sci. 41 (2001).
[45] M. Randić, M. Razinger, On characterization of molecular shape, J. Chem. Inf. Comput. Sci. 35 (1995) 594–606.
[46] M. Randić, Molecular shape profiles, J. Chem. Inf. Comput. Sci. 35 (1995) 373–382.
[47] M. Randić, Molecular profiles — novel geometry dependent molecular descriptors, New J. Chem. 19 (1995) 781–791.
[48] M. Randić, Quantitative structure–property relationship: boiling points of planar benzenoids, New J. Chem. 20 (1996) 1001–1009.
[49] M. Randić, Novel graph theoretical approach to heteroatoms in QSAR, Chem. Intel. Lab. Syst. 10 (1991) 213–227.
[50] M. Randić, On computation of optimal parameters for multivariate analysis of structure–property relationship, J. Comput. Chem. 12 (1991) 970–980.
[51] M. Pompe, M.M. Randić, Variable molecular descriptor for oxygen containing molecules, J. Chem. Inf. Comput. Sci. 41 (2001).
[52] M. Randić, D. Plavšić, N. Lers, Variable connectivity index for cycle containing structures, J. Chem. Inf. Comput. Sci. 41 (2001).
[53] M. Randić, D. Mills, S.C. Basak, On use of variable connectivity index for characterization of amino acids, Int. J. Quant. Chem., in press.
[54] B. Lučić, N. Trinajstić, New developments in QSPR/QSAR modeling based on topological indices, SAR QSAR Environ. Res. 7 (1997) 45–62.
[55] M. Randić, J. Zupan, On structural interpretation of topological indices, in: Proceedings of the Harry Wiener International Memorial Conference on Topology in Chemistry, Athens, GA, 20–24 March 2001.
[56] L. Bytautas, D.J. Klein, Alkane isomer combinatorics: stereo-structure enumerations, graph invariants, and molecular–property distributions, J. Chem. Inf. Comput. Sci. 39 (1999) 803–818.
[57] M. Randić, J. Zupan, On interpretation of well-known topological indices, J. Chem. Inf. Comput. Sci. 41 (2001).
[58] M. Randić, M. Pompe, The variable descriptors based on distance related matrices, J. Chem. Inf. Comput. Sci. 41 (2001).
[59] M. Randić, A.T. Balaban, S.C. Basak, On structural interpretation of distance related topological indices, J. Chem. Inf. Comput. Sci. 41 (2001).
[60] R. Compadre, C.M. Compadre, R. Catillo, W.J. Dunn Jr., On the use of connectivity indexes in quantitative structure–activity studies, Eur. J. Med. Chem. 18 (1983) 569.
[61] R. Franke, Theoretical Drug Design Methods, Elsevier, Amsterdam, 1984, p. 133.
[62] M. Randić, On characterization of chemical structure, J. Chem. Inf. Comput. Sci. 37 (1997) 672–687.
[63] L. Xu, W.-J. Zhang, A comparison of different methods for variable selection, Reported at Chemometrics and Analytical Chemistry, CAC2000, Antwerpen, October 2000 (preprint).
[64] L.B. Kier, L.H. Hall, Molecular Connectivity in Structure–Activity Analysis, Wiley, New York, 1986, pp. 82–87.
[65] M. Randić, Orthogonal molecular descriptors, New J. Chem. 15 (1991) 517–525.
[66] M. Randić, Resolution of ambiguities in structure–property studies by use of orthogonal descriptors, J. Chem. Inf. Comput. Sci. 31 (1991) 311–370.
[67] M. Randić, Fitting of non-linear regressions by orthogonalized power series, J. Comput. Chem. 14 (1993) 363–370.
[68] M. Randić, Curve fitting paradox, Int. J. Quant. Chem., Quant. Biol. Symp. 21 (1994) 215–225.
[69] M. Randić, Retro-regression — another important multivariate regression improvement, J. Chem. Inf. Comput. Sci. 41 (2001).
[70] M. Randić, Comparative regression analysis: regressions based on a single descriptor, Croat. Chem. Acta 66 (1993) 289–312.
[71] L.B. Kier, L.H. Hall, A differential molecular connectivity index, Quant. Struct. Act. Relat. 10 (1991) 134–140.
[72] B. Ren, A new topological index for QSPR of alkanes, J. Chem. Inf. Comput. Sci. 39 (1999) 139–143.
[73] M. Randić, J.Cz. Dobrowolski, Optimal molecular connectivity descriptors for nitrogen-containing molecules, Int. J. Quant. Chem. 70 (1998) 1209–1215.
[74] M. Randić, High quality structure–property regressions: boiling points of smaller alkanes, New J. Chem. 24 (2000) 165–171.
[75] M. Randić, S.C. Basak, Construction of high quality structure–property–activity regressions: the boiling points of sulfides, J. Chem. Inf. Comput. Sci. 40 (2000) 899–905.
[76] M. Randić, S.C. Basak, M. Pompe, M. Novic, Prediction of gas chromatographic retention indices using variable connectivity index, Acta Chim. Slovenica, in press.
[77] M. Randić, D. Mills, S.C. Basak, L. Pogliani, On characterization of several physicochemical properties of amino acids, J. Chem. Inf. Comput. Sci., in press.
[78] M. Randić, M. Pompe, On characterization of CC double bond in alkenes, SAR QSAR Environ. Res. 10 (1999) 451–471.
[79] M. Randić, S.C. Basak, Optimal molecular descriptors based on weighted path numbers, J. Chem. Inf. Comput. Sci. 39 (1999) 261–266.
[80] M. Randić, M. Pompe, Variable molecular descriptor versus traditional molecular descriptors, J. Chem. Inf. Comput. Sci. 41 (2001).
[81] A.R. Katritzky, V. Lobanov, M. Karelson, CODESSA (COmprehensive DEscriptors for Structural and Statistical Analysis), University of Florida, Gainesville, FL.
[82] G. Grassy, B. Calas, A. Yasri, R. Lahana, J. Woo, S. Iyer, M. Kaczorek, R. Floc'h, R. Buelow, Computer-assisted rational design of immuno-suppressive compounds, Nat. Biotech. 16 (1998) 748–752.