

On Structural Interpretation of Several Distance Related Topological Indices

Milan Randić,^{*,†,‡} Alexandru T. Balaban,^{§,#} and Subhash C. Basak

Natural Resources Research Institute, University of Minnesota at Duluth, 5013 Miller Trunk Highway, Duluth, Minnesota 55811-1442, National Institute of Chemistry, Ljubljana, Slovenia, Ames Laboratory-DOE, Iowa State University, Ames, Iowa 50011, Polytechnic University, Bucharest, Romania, and Department of Oceanography and Marine Sciences, Texas A&M University at Galveston, Galveston, Texas 77553

Received July 22, 2000

We consider the role that individual bonds play in bond-additivities in order to better understand the structural basis of various topological indices. In particular we consider indices closely related to the Wiener index (W) and the distance matrix and search for optimal weights of terminal and interior CC bonds in alkanes for a selection of physicochemical properties. It is interesting to note that different properties are associated with different relative roles of the exterior and the interior CC bonds.

INTRODUCTION

Mathematical characterization of molecules has led to hundreds of molecular descriptors, and their number continues to grow.^{1–6} These descriptors, often referred to as topological indices (TI), play an important role in structure–property and structure–activity studies, particularly when multivariate regression analysis, artificial neural networks, and pattern recognition are used as statistical tools. Their advantage over the “traditional” molecular descriptors used in the so-called Hansch Analysis,⁷ or the descriptors derived from Quantum Chemical approaches, is that they are easily available and can be quickly computed for existing or virtual structures. The apparent disadvantage of topological indices that may have contributed to their misunderstanding by less informed critics is a lack of studies on their interpretation. Even a cursory browsing through the abundant literature on the development and applications of topological indices shows that very few publications deal with the interpretation of topological indices. Without going into a thorough search it is possible to miss some such contributions, but we only came across rather few.^{8–16} Why is there such disparity between the use of indices and the efforts to clarify their meaning? An apparent answer is that interpretation of TI is difficult. This is well illustrated already for the Wiener index, the first nontrivial topological index introduced in chemistry over 50 years ago.¹⁷ Platt,¹⁸ who suggested that W be called Wiener index, tried to interpret this somewhat mysterious index “defined as the sum of the distances between any two carbon atoms in a molecule, in terms of carbon–carbon bonds. Brief method of calculation: Multiply the number of carbon atoms on one side of any bond by those on the other side; W is the sum of these values for all bonds.”¹⁷ According to Platt “ $\sqrt[3]{W}$ is a sort of mean molecular diameter”.

We recognize that the problem of interpretation of topological indices is very important and as such should not

be avoided. Recently, Randić and Zupan¹⁹ initiated an effort for a systematic search for interpretation of topological indices. This paper is a continuation of this effort.

ON INTERPRETATION OF MOLECULAR DESCRIPTORS

First we have to clarify what we understand by “interpretation” of theoretical quantities such as topological indices. Often those who are critical of topological indices have no valid perception of what “interpretation” is to represent, when asking “what is the physical meaning of such and such index”? If a TI has a clear-cut physical meaning so much the better, but when considering an interpretation of a quantity that is classified as a nonobservable, what is required is that the quantity has a meaning *within* the model considered. Such quantity needs not have meaning in some more general area of chemistry or physical chemistry. However, we should add that criticisms were right when raising the question of interpretation of topological indices, because many topological indices have no clear meaning even within the chemical graph theory. This is particularly the case with indices that aim at a very high power of discrimination. Such TI may be of interest for the graph isomorphism problem and may be of lesser interest for structure–property and structure–activity studies.

We should add further that one should distinguish a *definition* of an index from an *interpretation* of an index. By being mathematical constructions all topological indices are formally well defined, as we have just seen with the definition of the Wiener index. It is the meaning of such constructions within chemical models that need closer attention.

In Table 1 we have listed a number of nonobservables used in chemistry just to point out that topological indices are nothing special. Consider now two well-known concepts often mentioned in quantum chemical papers: the HOMO–LUMO separation and the Kekulé valence structures. Chemists are very familiar with both of these concepts, which do not need to be defined here. The HOMO–LUMO gap plays an important role in chemistry as has been initially docu-

[†] National Institute of Chemistry.

[‡] Iowa State University.

[§] Polytechnic University.

[#] Texas A&M University at Galveston.

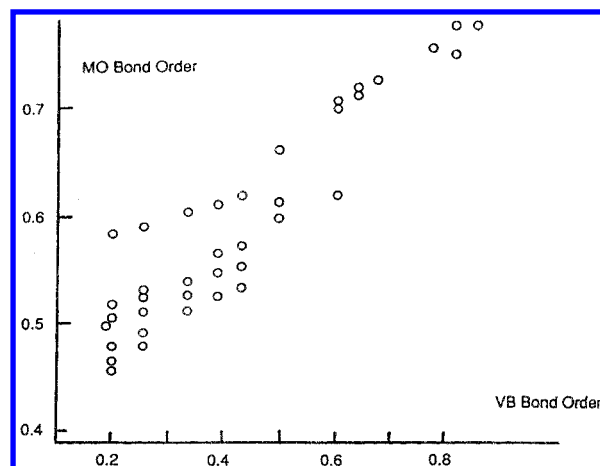
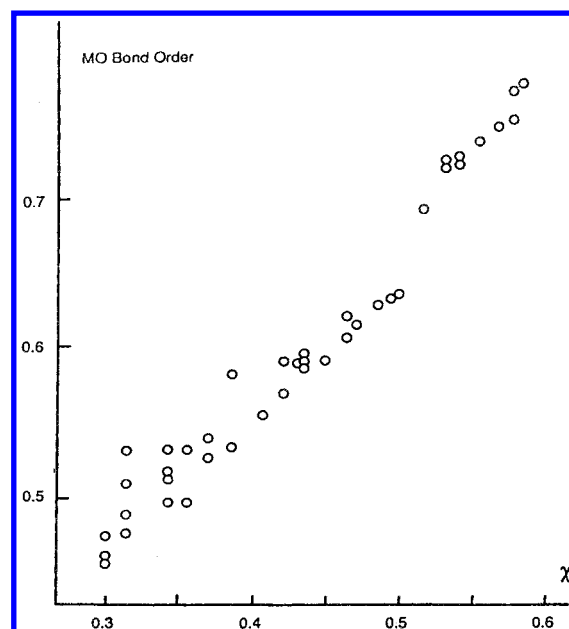
Table 1. List of a Number of Nonobservables Used in Chemistry

nonobservable	nonobservable
hybridization	topological indices
molecular orbitals	substructure
Kekulé valence structure	resonance energy
Clar structure	conjugated circuits
Fries valence structure	HOMO–LUMO
potential energy	localized orbitals
bond dipoles	

mented by the “Frontier Orbital Approach” of Fukui²⁰ (work which was awarded a Nobel prize in the year 1981). The so-called “zero gap” (when HOMO and LUMO orbitals are degenerate) is of interest, for example, for superconductivity of polymers, and plays a role in the chemistry of radicals and biradicals. The HOMO–LUMO separation has been evoked in a number of QSAR studies as a bona fide molecular descriptor, often by persons who frown upon topological indices without realizing that both quantities are equally legitimate or equally illegitimate (incidentally we think the former to be the case). Hence, clearly HOMO–LUMO is very important concept, but it has only meaning in MO models of quantum chemistry. HOMO–LUMO plays no role and has no interpretation in VB models. Similarly, nobody could deny the importance of Kekulé valence structures in chemistry. There are a number of structural concepts closely tied to Kekulé valence structures, e.g.: The Fries rule for characterization of the “most important” Kekulé valence structures,²¹ the Clar’s aromatic π -sextet,²² the conjugated circuits,^{23,24} and the innate degree of freedom.^{25,26} Yet, Kekulé valence structures have no place in MO models. This, of course, does not mean that some nonobservables of VB theory cannot be interpreted in MO theory and vice versa. An illustration is provided by work of Polansky and Derflinger²⁷ on interpretation of the aromatic sextets of Clar using MO calculations to compute local ring indices.

There are concepts of quantum chemistry that have a direct interpretation in both the MO model and VB model, which is an advantage of such concepts. These, for instance, include hybridization and localized orbitals. Because of the different nature of the MO and the VB models sometimes even the quantities that play the corresponding role in these models are not simply related. Consider, for instance, the Coulson bond order²⁸ and the Pauling bond order.²⁹ They both are expected to relate to experimental bond length of conjugated systems having several Kekulé valence structures. In Figure 1 we illustrate a plot of Pauling bond orders against Coulson bond orders for smaller benzenoid systems. Although there is clearly a visible correlation, there are also visible departures for many CC bonds included in the correlation. It may now be surprising to many to find out that a plot shown in Figure 2, in which Pauling bond orders have been plotted against the connectivity index $^1\chi$, shows a much better correlation than the correlations of the two quantum chemical bond orders.^{30,31} Apparently the connectivity index “has nothing to do” with quantum chemistry, or perhaps it has?

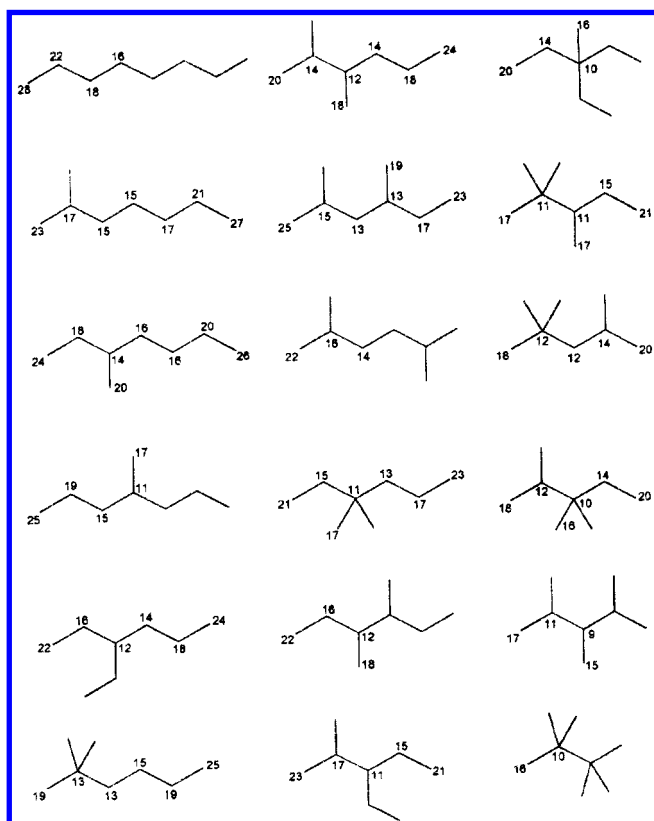
In the same spirit our basic position when considering interpretation of TIs is that we insist that topological indices have an interpretation *within* structural chemistry and not necessarily within physical chemistry. Justification for such an attitude is that models of structural chemistry and models of physical chemistry often involve different underlying

**Figure 1.** Correlation of the Pauling bond orders against Coulson bond orders for smaller benzenoid hydrocarbons.**Figure 2.** Correlation of the Pauling bond orders against the connectivity index $^1\chi$ for the same set of smaller benzenoid hydrocarbons considered in Figure 1.

concepts and hence need not be simply related or related at all. Randić and Zupan¹⁹ in their paper on interpretation of topological indices consider paths and walks as the *elementary* structural components, concepts assumed to be understood and not needing further explanations. Atoms can be formally viewed as paths of length zero, while chemical bonds represent paths (or walks) of length one. Hence, we would accomplish our task of interpreting topological indices if we could relate them to atoms and bonds, paths and walks, and in more general situations paths and walks of longer lengths. In the first paper of this series Randić and Zupan considered the Wiener index,¹⁷ the Hosoya index,³² and the connectivity index.³³ They have pointed out how one can partition topological indices associated with a graph matrix into bond contributions. There are many such indices, some examples are listed in Table 2. While the partition of an index into bond contributions could be difficult to obtain, we should add that partitioning of any topological index that is associated with sum of entries above the main diagonal of a symmetric matrix into atom contributions is straightfor-

Table 2. Selection of Topological Indices Associated with a Graph Matrix^a

matrix	index	symbol
adjacency matrix	the connectivity index	${}^1\chi$
the distance matrix	Balaban's J index	J
adjacency + distance	Schultz index	MTI
wiener matrix	Hyper-Wiener index	WW
reversed distance matrix	reversed Wiener	M
reciprocal distance matrix	Harary index	H
Hosoya matrix	modified Hosoya index	Z'
edge adjacency matrix	edge connectivity index	ϵ

^a The partition of which into bond contributions is possible.**Figure 3.** Distasums (row sums in distance matrices) for carbon atoms of 18 isomers of octane.

ward: assign to each atom the value obtained by summing all entries in the row, the row sum. In the case of distance matrix the row sums were referred to as distasums.³⁴ In Figure 3 we illustrate distasums for carbon atoms in the 18 isomers of nonane.

FROM ATOM PARTITION TO BOND PARTITION

The row sums of a matrix associated with a molecular graph can serve as a source for construction of a topological index that characterizes individual edges. To obtain bond contribution one considers the row sums for adjacent atoms. The row sums R_i and R_j associated with the bond (i, j) leads to contribution $1/\sqrt{(R_i \cdot R_j)}$, that is analogous to bond contributions used for construction of the first-order connectivity index ${}^1\chi$. Additional indices can be constructed by considering other types of row operations like $(R_i \cdot R_j)$, $1/(R_i \cdot R_j)$, and $\sqrt{(R_i \cdot R_j)}$. For example, Balaban³⁴ designed his topological index J by using the algorithm $1/\sqrt{(R_i \cdot R_j)}$, which originated with the connectivity index, and applied it

to the row sums of the distance matrix. Since then this algorithm has been used for constructing many analogous indices, such as the hyper-Wiener index and its extensions, including collection of indices derived from the Szeged and Cluj matrices.^{35–42} When the same algorithm is applied to the adjacency matrix one obtains the connectivity index, as was for the first time pointed out by Balaban.³⁴

The partitioning of W, ${}^1\chi$, and Z indices into bond contributions has shown an important distinction: while both ${}^1\chi$ and Z indices give greater weight to the terminal CC bonds of octane isomers, the W index does just the opposite: it assigns a greater role to the interior or “buried” CC bonds. This appears counterintuitive, because the more “exposed” terminal bonds are associated with larger portions of molecular surface and are expected to make a greater contribution to a selection of physicochemical properties. That indeed this is to be expected has been confirmed by construction of a modified W index, the W* index, which shows a better regression for the boiling points of octane isomers than W. The W index is constructed by taking the product of the number of atoms on each side of a bond.¹⁷ The W* index is obtained by taking the product of the reciprocal number of atoms on each side of a bond.

In summary, if we want to construct TI that will give greater weight to terminal bonds if the row sums belonging to terminal vertices of a graph matrix are smaller than the average row sums of the interior vertices, one should employ the products of the row sums given by the reciprocal square root, i.e., one should use $1/\sqrt{(R_i \cdot R_j)}$ for construction of bond contributions. This operation has been recently examined in more details and referred to as the “Balaban–Ivanciuc operation”.⁴² On the other hand, when the row sums belonging to terminal vertices of a graph matrix are greater than the average row sums of the interior vertices, then for obtaining bond contributions one should use $\sqrt{(R_i \cdot R_j)}$. With this insight we will consider distasums for construction of a novel topological index R^* . An apparent advantage of the R^* index (which is the consequence of its construction) is that it has its bond partitioning quite straightforward. The bond contributions are given either by $\sqrt{(R_i \cdot R_j)}$ or by $1/\sqrt{(R_i \cdot R_j)}$, depending on the average values of the row sums belonging to terminal and interior bonds.

An alternative method for obtaining bond contributions from the vertex partitioning of an index is to assign to a bond (i, j) the average value of the corresponding row sums: $(R_i + R_j)/2$. In Figure 4 we show the values of the average row sums for carbon atoms of octanes. By summing the so obtained bond contributions we obtain an alternative novel index that we will designate as R^+ .

COMPARISON OF R^* AND R^+ FOR OCTANE ISOMERS

In Table 3 we have listed the values of R^* and R^+ for the 18 isomers of octane. A simple regression shows that indeed the two indices are highly correlated and will not be expected to produce significant differences in applications. That this is the case is confirmed in Table 4 in which we listed correlations for seven physicochemical properties of octanes using R^* and R^+ as descriptors. We have deliberately shown the statistical parameters r , s , and F on several decimal places in order to illustrate minor differences between the two descriptors. The physicochemical properties considered are listed in Table 5.

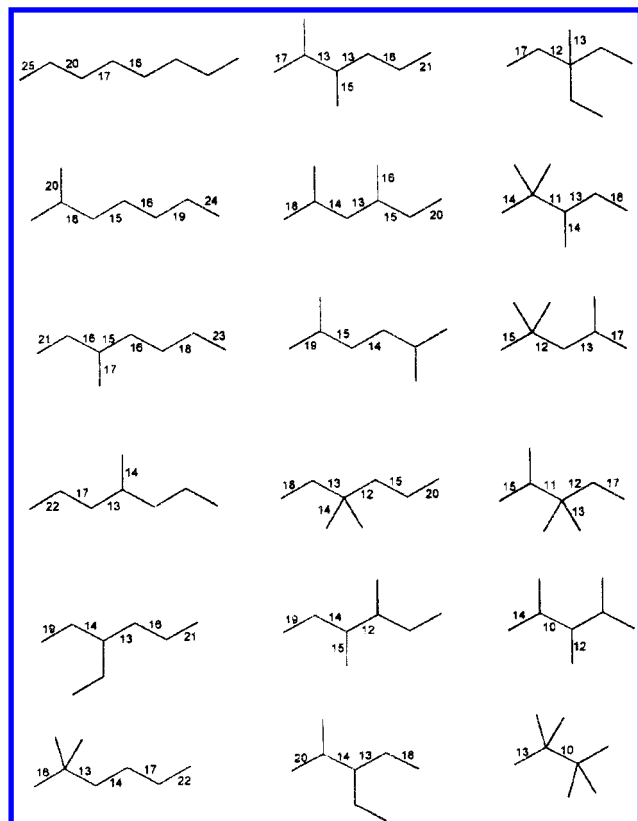


Figure 4. The average vertex contributions to individual CC bonds (the average row sums in distance matrices for adjacent carbon atoms) for carbon atoms of 18 isomers of octane. Bonds devoid of numbers have symmetrically equivalent bonds elsewhere in the graph.

Table 3. List of the Values of R^* and R^+ for the 18 Isomers of Octane

isomer	$(R_i^* R_j)$	$(R_i + R_j)/2$
<i>n</i> -octane	139.3795	20.00000
2M	129.1911	18.57143
3M	125.0510	18.00000
4M	116.7181	16.85714
3E	114.8567	16.57143
22MM	112.7894	16.28571
23MM	110.7454	16.00000
24MM	112.8167	16.28571
25MM	118.9799	17.14286
33MM	104.5411	15.14286
34MM	106.6300	15.42857
2M3E	114.4092	16.57143
3M3E	98.3452	14.28571
223MMM	96.2926	14.00000
224MMM	102.5187	14.85714
233MMM	94.2119	13.71429
234MMM	86.2179	12.57143
2233MMMM	85.8947	12.57143

In view of the strong overlap of the two descriptors R^* and R^+ we will in the following consider only R^+ , which is simpler to calculate. Before illustrating properties of R^+ descriptor let us comment on comparison of regressions and different methodologies, such as the multivariate regression analysis (MRA) and artificial neural networks (ANN). The purpose of such comparisons should not be to show which index or which method "outperforms" the other index or the other method. The issue is to obtain useful models and to better understand such models. The competing indices or the competing models should therefore be viewed as comple-

Table 4. Correlation of R^* and R^+ with Selected Physicochemical Properties of Octane Isomers^a

property		r	s	F
BP	R^*	0.8064	3.853	13.94
	R^+	0.8042	3.873	13.73
MV	R^*	0.9173	1.086	31.14
	R^+	0.9173	1.085	37.15
MR	R^*	0.9349	0.0693	48.55
	R^+	0.9349	0.0693	48.55
HV	R^*	0.8456	0.8263	17.57
	R^+	0.8439	0.8306	17.32
CT	R^*	0.6363	7.997	5.10
	R^+	0.6344	8.014	5.05
CP	R^*	0.6660	1.141	5.98
	R^+	0.6601	1.141	5.98
ST	R^*	0.8320	0.520	15.74
	R^+	0.8315	0.5206	16.69

^a The second descriptor is P, the number of paths of length three.

Table 5. Physicochemical Properties Considered in This Article

isomer	BP	MV	MR	HV	CT	CP	ST
<i>n</i> -octane	125.67	162.59	39.1922	41.48	296.20	24.64	21.76
2M	117.65	163.66	39.2316	39.68	288	24.8	20.60
3M	118.93	161.83	39.1001	39.83	292	25.6	21.17
4M	117.71	162.11	39.1174	39.67	290	25.6	21.00
3E	118.53	160.07	38.9441	39.40	292	25.74	21.51
22MM	106.84	164.29	39.2525	37.29	279	25.6	19.60
23MM	115.61	160.40	38.9808	38.79	293	26.6	20.99
24MM	109.43	163.09	39.1300	37.76	282	25.8	20.05
25MM	109.10	164.70	39.2596	37.86	279	25	19.73
33MM	111.97	160.88	39.0087	37.93	290.84	27.2	20.63
34MM	117.73	158.81	38.8453	39.02	298	27.4	21.64
2M3E	115.65	158.79	38.8362	38.52	295	27.4	21.52
3M3E	118.26	157.03	38.7171	37.99	305	28.9	21.99
223MMM	109.84	159.53	38.9249	36.91	294	28.2	20.67
224MMM	99.24	165.08	39.2617	35.13	271.15	25.5	18.77
233MMM	114.76	157.29	38.7617	37.22	303	29	21.56
234MMM	113.47	158.85	38.8681	37.61	295	27.6	21.14
2233MMMM	106.47				270.80	24.5	

mentary to one another and not as excluding each other. As Kohn⁴³ has pointed out the following:

"Modeling is a ubiquitous and often misunderstood enterprise ... It is unusual for only one model to be compatible with experimental observations. Often data are not sufficiently extensive to discriminate among rival models and new experiments must be designed to answer the outstanding questions. The statistical, graph theoretical, and sensitivity analysis methods ... can identify the areas for further investigation that are likely to produce significant new results."

ON COMPARISON OF REGRESSIONS

The correlations in Table 4 represent, at best, fair regressions and should not be considered necessarily satisfactory. However, a judgment of the quality of a regression is often subjective. Often authors do not consider a direct comparison of different descriptors on the same set of compounds and on the same data for such compounds. In addition, novel descriptors are compared sometimes with not the best descriptions available. It is important when comparing the performance of different indices to take into account the following:

(1) Different indices may have a different dependence on the size or even the shape. Hence, one should not restrict comparison of indices by restricting regressions to linear

Table 6. Illustration of Reported Fallacious Conclusion Concerning the Relative Merits of Different Indices When One Restricts the Correlations to Linear Model

Linear Model			
BP = -80.7898 + 6.6425 Xu	$r = 0.993$	$s = 5.791$	$F = 2615$
BP = -130.3039 + 67.63286 $^1\chi$	$r = 0.987$	$s = 7.908$	$F = 1428$
Quadratic Model			
BP = -90.9607 + 68.8983 Xu - 2.8276 (Xu) ²	$r = 0.995$	$s = 5.007$	$F = 1816$
BP = -199.9927 + 125.2571 $^1\chi$ - 10.8299 ($^1\chi$) ²	$r = 0.997$	$s = 4.128$	$F = 2680$

Table 7. Correlations of Normal Boiling Points Involving Shape Descriptors Applied to a Set of Compounds Containing Molecules of the Same or Different Size Illustrated for Nonanes and C₂-C₉ Alkanes, Respectively

C ₉ , n = 35	descriptors	coeff of regression	standard error, °C
1-variable	W*	0.2041	5.43
	H'	0.1243	5.70
	H	0.1032	5.76
	W	0.0919	5.80
2-variables	W, P	0.9278	1.66
	W*, P	0.9024	1.93
	H, P	0.8445	2.44
	H', P	0.8005	2.76
Alkanes			
C ₂ - C ₉ , n = 74	descriptors	coeff of regression	standard error, °C
1-variable	N	0.9712	7.91
	H'	0.9277	12.54
	H	0.9148	13.61
	W	0.8410	18.59
	W*	0.5405	31.61
2-variables	W* and N	0.9793	6.75
	H', P	0.9288	12.53
	H, P	0.9158	13.63
	W, P	0.9019	14.70
	W*, P	0.9203	13.25
	W*, P and N	0.9850	5.80

models, because some of the indices considered may depend quadratically (or another functional form) on molecular size. In Table 6 we illustrate fallacious conclusions that have been reported⁴⁴ concerning relative merits of indices when the above recommendation was not followed. Indeed, there is nothing in the theory of the structure-property-activity relationship that *demands* that regressions be linear!

(2) Some indices are shape descriptors, which by definition implies that they are not size-dependent. Their performance should be compared to other (size-dependent descriptors) only when the set of compounds considered are of the same size. The most simple size-dependent descriptor is *N*, the number of non-hydrogen atoms in a structure. If shape descriptors are to be applied to a set of compounds containing molecules of a different size, then *N* should also be included as a variable. We illustrate such a comparison in Table 7 on use of index W* on nonanes as a subset consisting of molecules of the same size, taken from a recent comparative study of distance-related indices.⁴⁵ The numerical values of all the indices used in this comparison are shown only for octane isomers in Table 8.

(3) Indices can have misleading labels or names or labels that only partially describe their nature. For example, Kier's kappa shape indices⁴⁶⁻⁵¹ do have some size dependence besides a shape dependence and therefore are not 100% shape indices. (The indices path/walks quotients^{52,53} and W* appear less dependent on *N*.) Similarly, $^0\chi$, the zero-order con-

Table 8. Numerical Values for p₃, W, H, H', and W* for Octane Isomers

isomer	p ₃	W	H	H'	W*
C ₈	5	84	13.7429	9.5015	.64821
2M - C ₇	5	79	14.1000	9.7311	.70774
3M - C ₇	5	76	14.2667	9.8144	.72440
4M - C ₇	6	75	14.3167	9.8369	.72857
3E - C ₆	7	72	14.4833	9.9203	.74524
22MM - C ₆	5	71	14.7667	10.1756	.78393
23MM - C ₆	7	70	14.7333	10.1078	.78810
24MM - C ₆	6	71	14.6500	10.0592	.78393
25MM - C ₆	5	74	14.4667	9.9656	.76726
33MM - C ₆	5	67	15.0333	10.3178	.80476
34MM - C ₆	8	68	14.8667	10.1789	.80060
2M3E - C ₅	8	67	14.9167	10.2914	.80476
3M3E - C ₅	9	64	15.2500	10.4375	.82143
223MMM - C ₅	8	63	15.4167	10.5764	.86012
224MMM - C ₅	5	66	15.1667	10.4306	.84345
233MMM - C ₅	9	62	15.5000	10.6250	.86429
234MMM - C ₅	8	65	15.1667	10.3889	.84762
2233MMMM - C ₄	9	58	16.0000	11.0000	.91964

nectivity index of Kier and Hall,⁵⁴ is an atomic index and not a bond index. Hence, when one considers bond additive properties one should not use $^0\chi$. Instead $^1\chi$ should be employed in such situations.

In a recent review, Bonchev⁵⁵ compared his topological complexity indices mK with connectivity indices $^m\chi$ using one to five variable models. According to Bonchev "In one-variable regressions, the 0K index, which is simply the number of C atoms, shows better statistics than the more complicated zero-order molecular connectivity for the seven of the nine properties examined." Bonchev substantiated this claim by showing one-variable regression between the number of C atoms (0K) and $^0\chi$. However, most of the properties considered there were bond additive and one should have used $^1\chi$ in one-variable regression, not $^0\chi$.

(4) We also would like to point to dangers of generalized statements, when considering correlations using few properties only. In his review, Bonchev refers to his "overall connectivity" and the connectivity indices as follows:⁵⁵

"Not dependent on subjectively selected mathematical function, like the inverse square root of molecular connectivities, the overall connectivities (i.e., Bonchev's mK) are ... constructed on a sound theoretical basis, ... and were shown to outperform molecular connectivities in modeling physicochemical properties of alkanes ..."

However, one should consider more than nine molecular properties if such a strong statement is to be made. First, in this study alkane properties, such as entropy, density, and motor octane numbers, were not included to mention a few. In support for his claim Bonchev states the following:

"The overall balance of the comparison made shows that in 39 of 45 pairs of models the new indices outperform molecular connectivities."

Table 9. Normalized Values for R^+ for $m = 2$ through $m = 6$ for the 18 Isomers of Octane

isomer	$k =$				
	2	3	4	5	6
C_8	58.85714	178.34402	555.53020	1774.23336	5791.2401
2M - C_7	50.48980	140.67055	401.41358	1171.84804	3494.2406
3M - C_7	47.34694	127.52187	351.73511	992.82410	2863.5761
4M - C_7	42.44898	111.54519	304.49479	858.41542	2483.8129
3E - C_6	40.40816	101.34111	260.74636	686.13173	1840.3980
22MM - C_6	38.89796	95.59767	242.13328	632.37544	1701.7184
23MM - C_6	37.51020	90.23907	222.74302	563.71464	1460.8090
24MM - C_6	38.65306	93.51603	230.35818	576.91462	1324.0352
25MM - C_6	42.65306	107.66764	275.28113	711.66597	1857.1710
33MM - C_6	33.75510	77.60933	183.97751	448.90320	1124.3165
34MM - C_6	34.85714	80.71137	191.36193	463.82031	1146.7847
2M3E - C_5	40.44898	101.46356	260.51229	681.83174	1812.2347
3M3E - C_5	29.95918	64.48980	142.16243	319.94764	732.6664
223MMM - C_5	28.53061	59.28863	125.71512	272.10139	601.1844
224MMM - C_5	31.95918	69.60933	153.35860	341.40322	767.1957
233MMM - C_5	27.38776	55.73177	115.48105	243.41501	521.2962
234MMM - C_5	23.02041	42.86880	80.96626	154.70506	298.3802
2233MMMM - C_4	22.73469	41.34694	75.53769	138.49932	254.6631

Such a statement is of no use to someone interested in the seven properties in which the opposite is the case. It would be more useful to find out why some combinations of indices are better in one case and others give superior results for other properties. In addition, there is no basis for not “mixing” indices, one does not have to use either mK or ${}^m\chi$, but one can combine indices that give best models. The above statement is misleading also by implying that other indices (here the molecular connectivity indices) have no “sound theoretical basis”. The selection of the inverse square root function in the construction of the connectivity indices in order to obtain bond contributions to ${}^1\chi$, as has been outlined in the paper on molecular connectivity indices, is based on finding a solution to a set of inequalities, hence, hardly deserves an attribute of a “subjectively selected mathematical function”. Moreover, recently Gutman, Araujo, and Rada⁵⁶ have arrived at a quite strong mathematical argument in favor of the choice of $-1/2$ as the exponent for bond contributions.

(5) In comparing topological indices and selecting one set of indices over another it is also of importance to consider interpretation of such indices as well as composition of the linear combinations of indices appearing in the regression equations. In addition, in some applications computational complexity may also be important factor. If one considers applications of topological indices in combinatorial chemistry as outlined by Lahana and co-workers⁵⁷ in recently reported synthesis of immunosuppressive compounds computational complexity becomes critical factor. In such applications many hundreds of thousands of virtual compounds have to be analyzed, and the selection of descriptors may depend on efforts needed for their evaluation. Indices that require construction of all subgraphs, which is for instance the underlying basis of Bonchev's complexity indices, or indices that are optimized during the statistical procedure, such as are the variable connectivity indices and variable paths,^{58–70} may be less attractive in such applications.

HOW TO IMPROVE CORRELATIONS

Our interest here is not to promote indices R^* and R^+ but along the quote of M. C. Kohn to “identify the areas for further investigation that are likely to produce significant

new results” with respect to the construction of novel topological indices. One often gets an impression, which may only in part be true, that most indices are constructed in an ad hoc manner. This in itself is not to be considered a deficiency, but by having so many indices around one would like to see development of methods that can improve the existing indices, rather than to come with another set of indices by trial and error. We can mention two schemes that allow construction of “improved” indices that have been outlined in the literature:

(1) Variations have been considered in the format of functional dependence of indices. For example, to obtain bond contributions when constructing the connectivity index one uses the algorithm $(mn)^{-1/2}$, where m, n are the degrees of vertices corresponding to atoms forming a bond. However different exponents can also be considered.

(2) Variables can be introduced directly when defining an index, and then during the regression analysis optimal variables are sought. For example, when considering physicochemical properties of alcohols one can introduce variables x and y to discriminate between the relative role of carbon and oxygen atoms in CC and CO bond contributions.

In the present paper we will explore yet another general approach to derive “improved” descriptors that resulted from the recent work on interpretation of well-known topological indices.¹⁷ We have seen that R^+ indices of octane isomers apparently satisfy our goal of having larger contributions for terminal bonds and smaller contributions for interior bonds. Nevertheless, when compared with several other indices that also satisfy the above requirement (e.g., the connectivity index ${}^1\chi$ and the Hosoya Z index), they are inferior. We have to conclude that this must be due to the fact that the relative weights for the terminal CC bonds and the interior CC bonds in octanes are better represented by ${}^1\chi$ and Z. So, if we want to improve the performance of R^+ index we have to consider modifications of the relative roles of the terminal and the peripheral CC bonds. There is a general way how to accomplish that: Consider instead of the weights $(R_i + R_j)/2$ different powers $\{R_i + R_j\}/2^m$, where m can take values different from one.

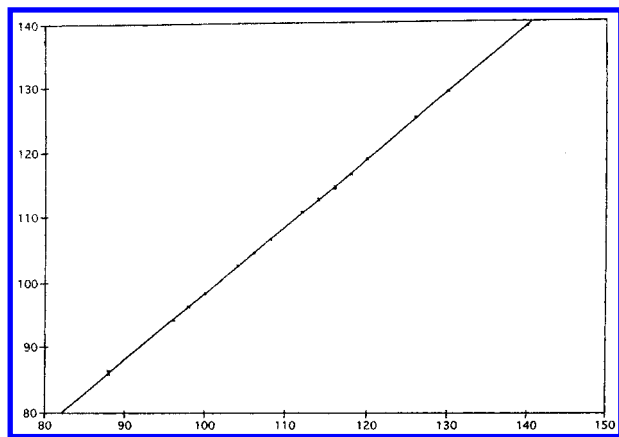


Figure 5. The plot of R^* against R^+ for 18 isomers of octane.

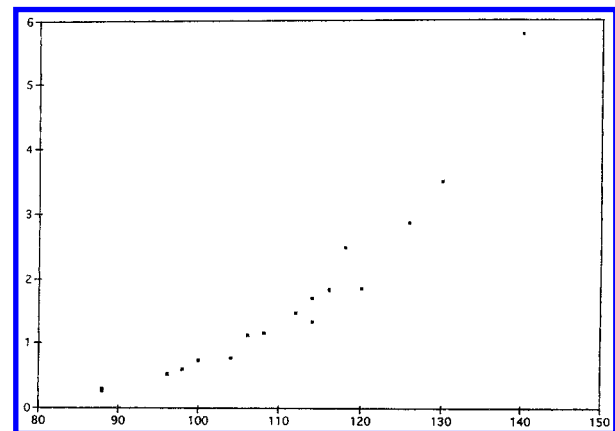


Figure 6. The plot of R^+ for $m = 6$ against R^+ for $m = 1$ for 18 isomers of octane.

SEARCH FOR OPTIMAL FORM OF R^+

In Table 9 we show normalized values for R^+ for $m = 2$ to $m = 6$ for the 18 isomers of octane. The normalization factor is $1/7^m$ (because octane isomer have seven bonds), that is, instead of the sum over all the bond contributions as given by R^+ , we now take the average bond contribution in order to restrain the fast growth of bond contributions. The normalization factor is introduced for convenience and does not involve conceptual novelty. By looking at the individual entries of Table 9 it is not immediately clear in which way R^+ values are modified, except for the apparent increase in magnitudes of R^+ , and the increase in the range of the values it takes as the exponent m increases. The range is given by the difference between the R^+ values for n -octane (the largest value) and 2,2,3,3-tetramethylbutane (the smallest value). In Figure 6 we show a plot of R^+ for $m = 6$ against R^+ ($m = 2$) from which we can see that “movements” of different octane isomers is nonuniform. Such nonuniform relative displacements will influence the quality of a regression analysis, because different isomers have been differently affected.

That indeed statistical parameters of regression equations are affected by the exponent m , which alters the relative weights of individual terminal and interior CC bonds, is evident from Table 10. We have summarized results for the seven physicochemical properties of octanes using modified descriptors $[R^+]^m$ with $m = 1$ to $m = 6$, and as the second descriptor P , the paths of length three. In each case we report the coefficient of the regression r , the standard error of

Table 10. Results for Correlations of the Seven Properties of Octanes Using Modified R^+ with $n = 1$ to $m = 6$

BP	r	s	F
$m = 1$	0.8042	3.873	13.73
$m = 2$	0.8293	3.6409	16.53
$m = 3$	0.8444	3.4916	18.63
$m = 4$	0.8492	3.4413	19.40
$m = 5$	0.8449	3.4857	18.71
$m = 6$	0.8349	3.5866	17.26
MV	r	s	F
$m = 1$	0.9173	1.085	37.15
$m = 2$	0.9183	1.079	37.65
$m = 3$	0.9193	1.073	38.17
$m = 4$	0.9202	1.067	38.65
$m = 5$	0.9209	1.063	39.06
$m = 6$	0.9218	1.057	39.58
MR	r	s	F
$m = 1$	0.9349	0.0693	48.55
$m = 2$	0.9349	0.0693	48.55
$m = 3$	0.9349	0.0693	48.56
$m = 4$	0.9349	0.0693	48.59
$m = 5$	0.9349	0.0693	48.59
$m = 6$	0.9349	0.0693	48.61
HV	r	s	F
$m = 1$	0.8439	0.8306	17.32
$m = 2$	0.8688	0.7607	21.54
$m = 3$	0.8834	0.7254	24.89
$m = 4$	0.8882	0.7112	26.17
$m = 5$	0.8841	0.7234	25.06
$m = 6$	0.8728	0.7556	22.39
CT	r	s	F
$m = 1$	0.6344	8.014	5.05
$m = 2$	0.6374	7.988	5.13
$m = 3$	0.6368	7.994	5.11
$m = 4$	0.6328	8.027	5.01
$m = 5$	0.6263	8.082	5.84
$m = 6$	0.6201	8.133	4.69
CP	r	s	F
$m = 1$	0.6601	1.141	5.98
$m = 2$	0.6707	1.135	6.13
$m = 3$	0.6750	1.129	6.28
$m = 4$	0.6789	1.123	6.41
$m = 5$	0.6818	1.119	6.52
$m = 6$	0.6832	1.117	6.57
ST	r	s	F
$m = 1$	0.8315	0.5206	15.69
$m = 2$	0.8505	0.4929	18.30
$m = 3$	0.8649	0.4704	20.78
$m = 4$	0.8741	0.4704	22.66
$m = 5$	0.8783	0.4552	23.61
$m = 6$	0.8800	0.4450	24.03

deviation s , and the Fisher ratio F . The first thing to observe in Table 10 is that for *different* properties, *different* values of the exponent m emerge as the best. This is in agreement with recent results obtained using as descriptors variable molecular paths and variable connectivity indices. This finding points out the limitations of “fixed” molecular descriptors which are constant during the regression analysis. The best regressions found when varying the exponent m have been emphasized. In the case of the boiling points (BP) and heats of vaporization (HV) of octanes, this happens when $m = 4$; for the critical temperatures (CT), the optimal case is $m = 2$; and for ST, the optimal exponent is $m = 3$. In the case of molar volumes (MV), molar refractivity (MR), and critical pressures of octanes (CP), optimal values of m have

Table 11. Average Values of Bond Contributions Found in Octane Isomers for Different Bond Types

bond type	bond contributions
1, 2	25, 24, 23, 22, 21, 20, 19, 18, 17
1, 3	20, 19, 18, 17, 15, 14, 12
1, 4	16, 15, 14, 13
2, 2	20, 19, 18, 17, 16, 15, 14
2, 3	15, 14, 13
2, 4	13, 12
3, 3	14, 13, 12, 10
3, 4	11
4, 4	10

not been found as the statistical parameters r , s , and F continue to improve slightly even after $m = 6$ has been tested.

We did not wish to pursue the search of optimal exponent beyond $m = 6$ for two reasons:

(1) The main goal of this contribution is to point out to a novel general approach for improvement of the topological indices that can be partitioned into atom and bond contributions.

(2) The selected indices of type R^* and R^+ have some limitations when compared to several other topological indices mentioned in this paper. In particular they appear inferior to $^1\chi$, Z , and W^* . Hence, it is doubtful that even the best R^+ index will lead to optimal differences for the relative role of the terminal CC bonds and the "buried" CC bonds.

The outcome of the analysis, summarized in the Table 10, may at first look disappointing. Because we have molecules of the same size the coefficient of regression r as well as the Fisher ratio F may serve as indicators of the relative quality of individual correlations considering different properties. The successive improvements of the topological index R^+ are significant even if not dramatic. Not all properties are equally affected by modification of the exponent m in $[R^+]^m$. For example, in the case of MR, there is hardly any improvement as m increases, which points out to limitations of R^+ for describing MR. In fact, when one considers only P as a descriptor in a simple regression for MR one obtains

$$r = 0.93487; s = 0.06699, \text{ and } F = 104.03$$

Hence, the use of R^+ in this case makes the regression rather worse than better.

Knowing the desirable qualities of a descriptor, such as giving larger weights to terminal bonds rather than to interior bonds, is helpful in the construction of novel molecular descriptors. We can certainly claim for properties such as the boiling points of octanes that the above considerations represent the necessary condition for the successful construction of a topological index. But from Table 10 it is clear that this is also not a sufficient condition, because R^+ did not perform as well as some other topological indices satisfying the same necessary conditions.

It may be instructive to investigate why R^+ indices have not reached a better performance. In Table 11 we have summarized the average values of bond contributions for different bond types as already shown in Figure 4. As we see from Table 11, as the valence of bond atoms increases the bond contributions decrease, and the interior bonds have on average smaller contributions than the terminal bonds (as desired). However, bond contributions for individual bonds shows quite wide variations, leading to overlaps not only among the terminal bond types but also between the terminal

and the interior bond types. Although within a single molecule terminal bonds mostly have greater contributions than internal CC bonds, there are few exceptions. For example, in 3-methylheptane and 4-methylheptane there is a terminal bond that has smaller weight than an internal CC bond. Additional cases include 2,2-dimethylhexane, 3,3-dimethylhexane, and 2,3-dimethylhexane. We believe that it is this overlapping of relative weights of bonds of different types that limit the adaptability of the R^+ index to produce better correlation for several physicochemical properties of octanes.

ACKNOWLEDGMENT

This is contribution number 287 from the Center for Water and the Environment of the Natural Resources Research Institute. Research reported in this paper was supported in part by Grants F49620-98-1-0015 and F49620-01-1-0098 from the United States Airforce. Authors wish to express their thanks to Dr. Ernesto Estrada and Professor Ivan Gutman for their constructive comments on the manuscript.

REFERENCES AND NOTES

- (1) Randić, M. Topological Indices. 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, 1998; pp 3018–3032.
- (2) Balaban, A. T. Historical developments of topological indices. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach Sci. Publ: Amsterdam, 1999; pp 21–57.
- (3) Basak, S. C. Information theoretic indices of neighborhood complexity and their applications. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach Sci. Publ: Amsterdam, 1999; pp 563–593.
- (4) Randić, M.; Novič, M.; Vračko, M. *Molecular Descriptors, New and Old, Lecture Notes in Chemistry*; Submitted for publication.
- (5) Randić, M.; Vračko, M. In *QSAR/QSPR Studies by Molecular Descriptors*; Diudea, M. V., Ed.; Nova Sci. Publ.: Commack, NY.
- (6) Randić, M. On characterization of chemical structure. *J. Chem. Inf. Comput. Sci.* **1997**, 37, 672–687.
- (7) Hansch, C.; Leo, A. *Exploring QSAR. Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995.
- (8) Motoc, I.; Balaban, A. T. Topological indices: Intercorrelations, physical meaning, correlation ability. *Rev. Roum. Chim.* **1981**, 265, 593–600.
- (9) Kier, L. B.; Hall, L. H. Derivation and significance of valence molecular connectivity. *J. Pharm. Sci.* **1981**, 70, 583–589.
- (10) Labanowski, J. K.; Motoc, I.; Dammkoehler, R. A. The physical meaning of topological indices. *Computer Chem.* **1991**, 15, 47–53.
- (11) Bögel, H.; Dettmann, J.; Randić, M. Why is the topological approach in chemistry so successful? *Croat. Chem. Acta* **1997**, 70, 827–840.
- (12) Hosoya, H.; Gotoh, M.; Murakami, M.; Ikeda, S. Topological index and thermodynamic properties. 5. How can we explain the topological dependency of thermodynamic properties of alkanes with the topology of graphs. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 192–196.
- (13) Kier, L. B.; Hall, L. H. The nature of structure–activity relationships and their relation to molecular connectivity. *Eur. J. Med. Chem. Chim. Therapeutica* **1977**, 12, 307–312.
- (14) Kier, L. B.; Hall, L. H. Intermolecular accessibility: The meaning of molecular connectivity. *J. Chem. Inf. Comput. Sci.* In press.
- (15) Estrada, E. Connectivity polynomial and long-range contributions in the molecular connectivity model. *Chem. Phys. Lett.* **1999**, 312, 556–560.
- (16) Stankevich, I. V.; Skvortsova, M. I.; Zefirov, N. S. On a quantum, chemical interpretation of molecular connectivity indices for conjugated hydrocarbons. *J. Mol. Struct. (THEOCHEM)* **1995**, 342, 173–179.
- (17) Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **1947**, 69, 17–20.
- (18) Platt, J. R. Prediction of isomeric differences in paraffin properties. *J. Phys. Chem.* **1952**, 56, 328–336.
- (19) Randić, M.; Zupan, J. On the interpretation of well-known topological indices. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 550–560.

- (20) Fukui, K. The path of chemical reactions—The IRC approach. *Acc. Chem. Res.* **1981**, *14*, 363–368.
- (21) Kekulé, A. *Bull. Soc. Chim. France* **1865**, *3*, 98.
- (22) Clar, E. *The Aromatic Sextet*; Wiley: New York, 1972.
- (23) Randić, M. Conjugated circuits and resonance energies of benzenoid hydrocarbons. *Chem. Phys. Lett.* **1976**, *38*, 68–70.
- (24) Randić, M. Aromaticity and conjugation. *J. Am. Chem. Soc.* **1977**, *99*, 444–450.
- (25) Randić, M.; Klein, D. J. Kekule valence structure revisited. Innate degrees of freedom of pi-electron coupling. In *Mathematical and Computational Concepts in Chemistry*; Ellis Horwood Ltd: Chichester England, 1986; pp 274–282.
- (26) Klein, D. J.; Randić, M. Innate degree of freedom of a graph. *J. Comput. Chem.* **1987**, *8*, 516–521.
- (27) Polansky, O. E.; Derflinger, G. Zur Clar'schen Theorie Lokaler Benzoïder Gebiete in Kondensierten Aromaten. *Int. J. Quantum Chem.* **1967**, *1*, 379–401.
- (28) Coulson, *Valence*; Oxford University Press: 1961.
- (29) Pauling, L. *The Nature of the Chemical Bond*, 3rd ed.; Cornell University Press: Ithaca, NY, 1960.
- (30) Randić, M.; Guo, X. Generalized bond orders. *Int. J. Quantum Chem.* **1994**, *49*, 215–237.
- (31) Gupta, S. P.; Singh, P. The relationship of p-binding energy with molecular connectivity in hydrocarbons. *Bull. Chem. Soc. Jpn.* **1979**, *52*, 2745–2746.
- (32) Hosoya, H. Topological index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332–2339.
- (33) Randić, M. On the characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.
- (34) Balaban, A. T. Highly discriminating distance-based topological index. *Chem. Phys. Lett.* **1982**, *89*, 399–404.
- (35) Randić, M.; Guo, X.; Oxley, T.; Krishnapriyan, H. Wiener matrix: Source of novel graph invariants. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 709–716.
- (36) Randić, M.; Guo, X.; Oxley, T.; Krishnapriyan, H.; Naylor, L. Wiener matrix invariants. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 361–367.
- (37) 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.
- (38) Estrada, E.; Rodriguez, L.; Gutierrez, A. Matrix algebraic manipulations of molecular graphs. 1. Distance and vertex-adjacency matrix, *MATCH* **1997**, *35*, 145–156.
- (39) Estrada, E.; Ivanciuc, O.; Gutman, I.; Gutierrez, A.; Rodriguez, L. extended Wiener indices. A new set of molecular descriptors for quantitative structure–property studies. *New J. Chem.* **1998**, 819–822.
- (40) Diudea, M. V.; Minailiuc, O.; Katona, G.; Gutman I. Szeged matrices and related numbers. *MATCH* **1997**, *35*, 119–143.
- (41) Diudea, M. V. Cluj matrix, C_{ju} : Source of various graph descriptors. *MATCH* **1997**, *35*, 163–183.
- (42) Ivanciuc, O.; Ivanciuc, T.; Balaban, A. T. Design of topological indices. Part 10. Parameters based on electronegativity and covalent radius for computation of molecular graph descriptors for heteroatom-containing molecules. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 395–401.
- (43) Kohn, M. C. Strategies for computer modeling. *Bull. Math. Biol.* **1986**, *48*, 417–426.
- (44) Ren, B. A new topological index for QSPR of alkanes. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 139–143.
- (45) Trinajstić, N.; Nikolic, S.; Randić, M. Work in progress.
- (46) Kier, L. B. A shape index from molecular graphs. *Quant. Struct. – Act. Relat.* **1985**, *4*, 109–116.
- (47) Kier, L. B. Shape indexes or orders one and three from molecular graphs. *Quant. Struct. – Act. Relat.* **1986**, *5*, 1–7.
- (48) Kier, L. B. Distinguishing atom differences in a molecular graph shape index. *Quant. Struct. – Act. Relat.* **1986**, *5*, 7–12.
- (49) Kier, L. B. Indexes of molecular shape from chemical graphs. *Acta Pharm. Jugosl.* **1986**, *36*, 171–188.
- (50) Kier, L. B. Indexes of molecular shape from chemical graphs. *Med. Res. Rev.* **1987**, *7*, 417–440.
- (51) Kier, L. B.; Hall, L. H. *The kappa indices for modeling molecular shape and flexibility*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach Sci. Publ: Amsterdam, 1999; pp 491–562.
- (52) Randić, M. Novel shape descriptors for molecular graphs. *J. Chem. Inf. Comput.* **2001**, *41*, 607–613.
- (53) Randić, M. On Characterization of shape of molecular graphs. *SAR QSAR Environ. Sci.* Submitted for publication.
- (54) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Analysis*; John Wiley & Sons: New York, 1986.
- (55) Bonchev, D. Overall connectivity and topological complexity for QSAR/QSPR. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Delivers, J., Balaban, A. T., Eds.; Gordon and Breach Sci. Publ: Amsterdam, 1999; pp 361–401.
- (56) Gutman, I.; Araujo, O.; Rada, J. An identity for Randić's connectivity index and its applications. *ACH Mod. Chem.* **2000**, *137*, 653–658.
- (57) Grassy, G.; Calas, B.; Yasri, A.; Lahana, R.; Woo, J.; Iyer, S.; Kaczorek, M.; Floc'h, R.; Buelow, R. Computer-assisted rational design of immuno-suppressive compounds. *Nature Biotechn.* **1998**, *16*, 748–752.
- (58) Bošnjak, N.; Mihalić, Z.; Trinajstić, N. Application of topological indices to chromatographic data: Calculation of retention indices of alkanes. *J. Chromatogr.* **1991**, *540*, 430–440.
- (59) Pogliani, L. From molecular connectivity indices to semiempirical terms. Recent trends in graph theoretical descriptors. *Chem. Rev.* **2000**, *100*, 3827–3858.
- (60) Randić, M. Novel graph theoretical approach to heteroatoms in QSAR. *Chemometrics Intel. Lab. Syst.* **1991**, *10*, 213–227.
- (61) Randić, M. On computation of optimal parameters for multivariate analysis of structure–property relationship. *J. Comput. Chem.* **1991**, *12*, 970–980.
- (62) Randić, M.; Dobrowolski, J. Cz. Optimal molecular connectivity descriptors for nitrogen-containing molecules. *Int. J. Quantum Chem.* **1998**, *70*, 1209–1215.
- (63) Randić, M.; Mills, D.; Basak, S. C. On use of variable connectivity index for characterization of amino acids. *Int. J. Quantum Chem.* **2000**, *80*, 1199–1209.
- (64) Randić, M.; Mills, D.; Basak, S. C.; Pogliani, L. On characterization of several physicochemical properties of amino acids. *New J. Chem.* To be submitted.
- (65) Randić, M.; Basak, S. C. On construction of high quality structure–property regressions: The boiling points of Sulfides. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 899–905.
- (66) Randić, M.; Basak, S. C. Optimal molecular descriptors based on weighted path numbers. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 261–266.
- (67) Randić, M.; Pompe, M. On characterization of CC double bond in Alkenes. *SAR QSAR Environ. Res.* **1999**, *10*, 451–471.
- (68) Randić, M. High quality structure–property regressions. Boiling point of smaller alkanes. *New J. Chem.* **2000**, *24*, 165–171.
- (69) Randić, M.; Pompe, M. Variable molecular descriptor versus traditional molecular descriptors. *J. Chem. Inf. Comput. Sci.* In press.
- (70) Randić, M.; Hansen, P. J.; Jurs, P. C. Search of useful graph theoretical invariants of molecular structure, *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 60–68.

CI000105L