

ARTICLES

“Anticonnectivity”: A Challenge for Structure–Property–Activity StudiesMatevž Pompe^{*,†} and Milan Randić[‡]

Faculty of Chemistry and Chemical Technology, University of Ljubljana, Aškerčeva 5, 1000 Ljubljana, Slovenia, 3225 Kingman Rd., Ames, Iowa 50014

Received April 12, 2005

The higher-order variable connectivity indices were introduced to account for the combination of positive and negative relative contributions of atoms and bonds in the construction of the quantitative structure–property relationships or quantitative structure–activity relationships models. The coding capabilities of modified descriptors were presented on the modeling of the atmospheric reaction rate constants of selected organic compounds with OH radicals. The optimization of diagonal weights of the augmented adjacency matrix pointed out the significant enhancing effect of oxygen and the suppressive effect of chlorine on the overall molecular atmospheric reactivity of organic compounds with OH radicals. The linear regression model, using a single structural descriptor, that is, a variable connectivity index of order one, produced a root-mean-square error of 0.343 log units. Although the obtained calculation error was higher than in previously reported multiple linear regression models, the new model offered important insight into the role of the individual structural components that are influencing the reactivity of organic compounds.

1. INTRODUCTION

The most important problem of quantitative structure–property relationship (QSPR) and quantitative structure–activity relationship (QSAR) studies remains to be a mathematical representation of chemical structure, that is, a representation of molecular structure with mathematical descriptors and the translation of molecular structure to the computer-readable form. For the creation of the prediction model, one has to recognize and extract critical structural information relevant to the considered structure–property relationship. Significant progress has been reported during recent years in the development of various topological, geometric, electrostatic, and quantum chemical indices, to be used as molecular descriptors.^{1–4} Because of the simplicity of topological structural representation, these indices are sometimes preferred to more complicated geometric, electrostatic, and quantum chemical descriptors, especially in cases where their use significantly reduces the computation time. One example is the screening of combinatorial libraries that may contain several hundred thousand virtual compounds. For an illustration of this novel use of topological indices, we may point to the work of Lahana and co-workers. They have selected around two dozen candidates from the virtual library having some 280 000 decapeptides based on the screening using topological indices. Out of them, five were examined more closely and were synthesized. The described study resulted in a candidate with an immunosuppressive activity 2 orders of magnitude higher than that of the initial leading compound.⁵

Nevertheless, the majority of the topological indices still suffer from two drawbacks: (1) a lack of structural interpretation of the obtained models and (2) an inability to differentiate adequately between heteroatoms and, therefore, an inability to encode polar interactions between molecules.

For example, despite that the Wiener number⁶ was introduced more than 50 years ago, and has been widely used and studied, its structural interpretation remains somewhat still unclear. Recently, however, some progress was made in clarifying the interpretation of a selection of common topological indices by partitioning their magnitude to bond additive contributions.^{7–9} On the other hand, Kier and Hall were the first to introduce different “corrections” to the role of heteroatoms in molecular graphs by constructing valence connectivity.¹⁰ Later, Ivanciuc et al.¹¹ modified the Balaban J index¹² to similarly be able to discriminate between heteroatoms. More recently, Zefirov and co-workers introduced a solvation connectivity index, which prescribed different contributions of atoms depending on their position in the periodic table.^{13,14}

It is important to emphasize that all mentioned numerical modifications of topological indices produce “fixed” descriptors; that is, once selected, they do not change from application to application. They ought to be viewed as empirical in nature, and despite that they tend to be “justified” by some ad hoc structural/electronic arguments, they often lack a deeper theoretical justification. They all start with the premise that the relative weights that encode the contribution of individual heteroatoms are the same regardless of the modeled property. It is difficult to believe that, for instance, different atoms will have the same contribution to the boiling point of particular compounds as they will to their molar refraction or activity. There were

* Author to whom correspondence should be addressed. Fax: ++386 1 2419 220. E-mail: matevz.pompe@fkkt.uni-lj.si.

† University of Ljubljana.

‡ 3225 Kingman Rd.

clear signs that such assumptions are incorrect, but these early indications were overlooked until the variable connectivity index was “invented”.

For instance, in a number of illustrations of structure–property–activity correlation in the book of Kier and Hall on molecular connectivity in chemistry and drug research,¹⁰ there are correlations in which the ordinary connectivity index works better than the valence connectivity index, instead of the opposite. This means that the “fixed” weights associated with the valence connectivity of heteroatoms in molecules produce worse regression for the corresponding properties than when “unweighted” atomic contributions are used. However, because such a phenomenon is observed for some properties and not others, this means that different weights better suit different properties.

All mentioned ambiguities were solved by introduction of the variable topological indices,^{15,16} which solved the problem of differentiation between heteroatoms as well as led to the insightful structural interpretation of the optimal variable weights.^{17,18} Optimal weights represent a relative contribution of individual heteroatoms and are modified during the modeling procedure to best suit the modeled physicochemical property.^{19–27} Not only did such an approach enable construction of the high-quality regression models, but it offered novel structural interpretation for the descriptors used. The relative values of weights may offer an insight into which part of the molecule is responsible for the particular property. An alternative route to incorporate some flexibility to topological indices is to consider modification of the functional form of descriptors, such as variation of the exponent^{28–31} in the expression for descriptors. However, the so-constructed indices still belong to the class of “fixed” topological indices, when discrimination between heteroatoms is considered, because they remain constant during the search for best regression equation.

Several properties exist where the presence of a particular atom or functional group in the molecule suppresses the observed property. The possibility of such a phenomenon was previously overlooked. Hundreds and hundreds of topological indices hitherto used were constructed with positive local contributions to the overall molecular index. However, with the introduction of the variable connectivity index, for the first time, it was realized that atoms or bonds can be associated with the “zero” contribution. Moreover, only recently, one of the present authors introduced a modification of the variable connectivity index, which takes into account not only the positive but also possible *negative* contributions of atoms or bonds in structure–property–activity relationships.³² The negative region of influences was named the “anticonectivity” region. The so-modified variable connectivity index of order zero was used for modeling flame ionization detector (FID) response factors and resulted in a better regression than a correlation using only positive additive contributions in the construction of topological indices.

The aim of this work is to extend the definition of the modified connectivity index of zero order to the higher-order connectivity indices. The coding capabilities of newly defined indices will be illustrated in modeling of the atmospheric reaction rate constants of organic compounds with OH radicals. This example, however, is not an isolated case of a property that involves the notation of anticonectivity. Our

Table 1. Augmented Adjacency Matrix of *n*-Butanol

	1	2	3	4	5	row sum
1	<i>x</i>	1	0	0	0	1 + <i>x</i>
2	1	<i>x</i>	1	0	0	2 + <i>x</i>
3	0	1	<i>x</i>	1	0	2 + <i>x</i>
4	0	0	1	<i>x</i>	1	2 + <i>x</i>
5	0	0	0	1	<i>y</i>	1 + <i>y</i>

preliminary examination shows that a similar situation occurs with the pK_a values of selected organic acids as well as with selected properties of amino acids. Hence, anticonectivity may be a more widespread phenomenon in structure–property relationships than one may have anticipated.

2. THEORY

The variable connectivity index (${}^1\chi^f$) developed from the connectivity index of order one³³ was the first variable topological index and was introduced almost 15 years ago.^{15,16} It was calculated from the augmented adjacency matrix illustrated in Table 1 for the case of *n*-butanol, where the zero diagonal elements of the adjacency matrix were replaced by variables *x* and *y*. The row sum of the augmented adjacency matrix (δ_i^f) represents atomic augmented valence. The new diagonal elements of the matrix, that is, the variables *x* and *y*, represent, in the calculation of the molecular connectivity index, atomic attributes for carbon and oxygen atoms, respectively. The expression for the first-order variable connectivity index (${}^1\chi^f$) is as follows:

$${}^1\chi^f = \sum_{j=1}^m {}^1\chi_j^f = \sum_{j=1}^m \prod_{i=1}^2 (\delta_i^f)^{-0.5} \quad (1)$$

The summation is made over all edges *m* (the paths of length one); therefore, ${}^1\chi_j^f$ represents the contribution of edge *j* to the connectivity index, and δ_i^f is the row sum of the augmented adjacency matrix, which can be calculated as the sum of the atom valence (δ_i) and corresponding diagonal weight (w_i). A generalization of eq 1 is as follows:

$${}^k\chi^f = \sum_{j=1}^m {}^k\chi_j^f = \sum_{j=1}^m \prod_{i=1}^{k+1} (\delta_i^f)^{-0.5} \quad (2)$$

where ${}^k\chi_j^f$ represents the contribution of *j*th path of length *k* to the variable connectivity index of the same order and *m* is the number of paths of length *k*. The contributions to the variable connectivity index are defined for real numbers; therefore, the weights *x* and *y* have to be varied in such a way that the row sum of the augmented adjacency matrix remains positive ($\delta_i^f = \delta_i + x_i > 0$). In this way, the contribution of an individual atom or a bond to the total sum varied only in the interval from zero to infinity (the upper right curve in Figure 1).

Despite the flexibility of the variable connectivity index, which was successfully used for the construction of several high-quality structure–property regressions,^{19,26} the variable index also has some limitations. In a few instances, the optimal weights for some atoms were found to be close to infinity, which means that such atoms or functional groups hardly have influence on the modeled property. But what if the true influence of these atoms is not zero but negative, which would mean that a certain atom or functional group

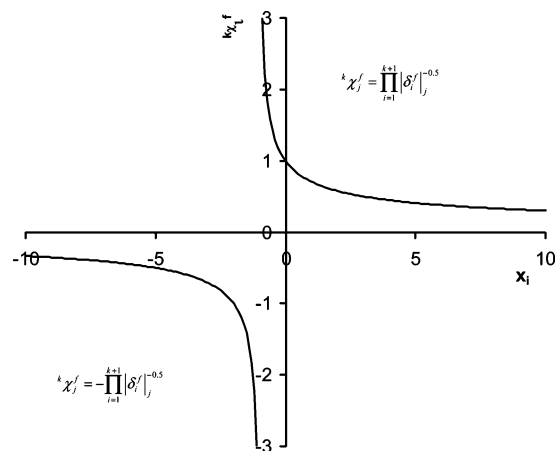


Figure 1. Dependence of $k\chi^f$ on the individual weight of the augmented adjacency matrix.

suppresses the observed property? This challenge was very recently successfully addressed by broadening the concept of connectivity to include the notion of anticonnectivity. The first such use was illustrated by modifying the connectivity index of zero order,³² which was applied to modeling of the FID response factors in gas chromatography.³⁴ The study has shown that the optimization of the so-modified variable connectivity index of zero order correctly predicts negative influence of oxygen on the FID response.

One can generalize the same modification also to higher-order connectivity indices, which could be used for the modeling of bond additive molecular properties. The contribution of a selected path of length k to the connectivity index of the same order is, therefore, expressed as

$$^k\chi_j^f = \pm \prod_{i=1}^{k+1} \delta_{ij}^{f-0.5} \quad (3)$$

The negative contribution of an individual bond is shown in the lower left curve of Figure 1. To stress that contributions are negative and suppress the property considered, we refer to this region of variables as the “anticonnectivity”.

3. CALCULATION OF MODIFIED VARIABLE CONNECTIVITY INDEX

Optimization of the modified variable connectivity index of order k ($^k\chi^f$) was performed by using the Simplex procedure.³⁵ Initially, random weights were assigned to the diagonal elements of the augmented adjacency matrix; however, they were subject to certain limits to avoid the occurrence of complex numbers. During the optimization procedure, the Simplex algorithm was used to optimize weights that would minimize the standard error of a linear regression equation: Property = $a \times ^k\chi^f + b$. At the beginning, all the weights were initialized in the “connectivity” region. If the diagonal weight for a certain atom exceeded the value of 10^8 , the algorithm used the negative solution of eq 3 to find the optimal value of the particular weight in the “anticonnectivity” region.

It is known that the Simplex algorithm may lead to a local minimum rather than the global one. The chances of selecting a local minimum were reduced by starting the optimization procedure several times using different initial weights. The

calculation ability of the developed model was evaluated by root-mean-square (RMS) error.

4. MODELING OF k_{OH} VALUES

In this work, we will show the use of modified variable connectivity indices in the modeling of the tropospheric reaction rate constants for the reaction of OH radicals with different organic compounds. The reactions of volatile organic compounds (VOCs) play an important role in different atmospheric photochemical processes. Up to now, experimental reaction rate constants are available only for around 500 VOCs.^{36–38} To reduce the analysis time and costs of such measurements, it would be useful to develop a theoretical prediction model to estimate these values. Several QSPR models based on the fragment contribution technique,³⁹ the bond dissociation energy,^{40,41} the ionization potentials,⁴² the molecular orbital calculations,^{43–46} and various molecular descriptors^{47–51} are already available. The comprehensive overviews of these methods together with their partial evaluations were recently published.^{52,53} Although the present modeling techniques offer moderate prediction capabilities, the development of a simple prediction model, which enables a good structural interpretation, continues to be of interest and would be beneficial.

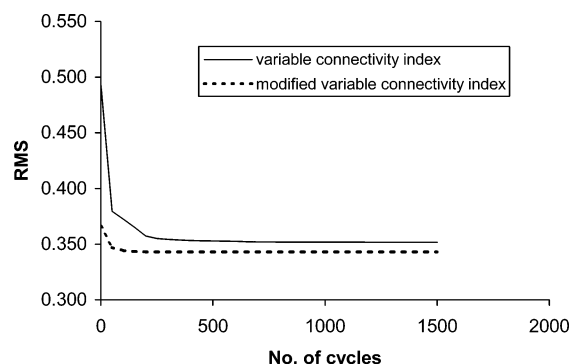
The studied data set was taken from the literature.⁵⁴ It has 39 organic compounds that contain carbon, oxygen, and chlorine atoms (Table 2). The dominant pathway for their reaction with OH radicals is hydrogen atom abstraction. It is also known that the presence of chlorine atoms reduces the reaction rate with OH radicals. The chemical structures were encoded by the variable connectivity index not only because it enables the construction of simple linear regression models but also because a recent study has shown that its modification can also be used for modeling of the presence of “anticonnectivity” effects³² such as the possible “negative” influence of chlorine atoms.

Separate weights x , y , and z were assigned to diagonal elements which represent the contributions of C, O, and Cl atoms, respectively, to the overall $-\log k_{OH}$ value of the molecule. By assuming a linear regression model, the Simplex method was used for the optimization of mentioned weights to get the lowest RMS error for the retrieved data. Both the variable connectivity indices of orders zero and one were tested for modeling capabilities. The variable connectivity index of order one has shown better results; therefore, it was used in the further study. The changes of RMS error value during the optimization procedure are shown in Figure 2 (the upper full curve). The optimization lowered the RMS error by 20–30% and reached the value of 0.352 log units. The optimal weights found for C, O, and Cl atoms were $x = 0.161$, $y = -0.773$, and $z = 10^8$, respectively. That the variable weight for carbon atoms is close to zero means that the relative contributions of carbon atoms continue to parallel their graph theoretical valence, meaning that the simple connectivity index $^1\chi$ would be a fair descriptor. On the other hand, the negative weight for oxygen points to the fact that the overall reactivity of the molecule is significantly increased if oxygen atoms are present. The negative value of y decreases the denominator and, thus, increases the contribution to the overall connectivity index of C–O bonds. It is very interesting that the optimal

Table 2. Tropospheric $-\log k_{\text{OH}}$ Values for Different Organic Compounds, Anticonnectivity Calculated $\log k_{\text{OH}}$ Values, and Values Obtained by the Linear Regression Models of Bakken and Jurs (B & J) and CODESSA^a

ID	name	exptl	calcd	B & J	CODESSA
1	ethane	12.57	12.32	12.30	12.53
2	<i>n</i> -butane	11.60	11.86	11.55	11.96
3	2,2-dimethylpropane	12.07	11.77	12.50	11.89
4	<i>n</i> -pentane	11.40	11.62	11.46	11.65
5	2-methylpropane	11.63	11.92	11.92	12.08
6	<i>n</i> -propane	11.94	12.09	11.99	12.28
8	2-methylbutane	11.41	11.67	11.44	11.74
9	2-methylpentane	11.25	11.43	11.31	11.42
10	3-methylpentane	11.24	11.43	11.35	11.40
11	2,2-dimethylbutane	11.63	11.52	11.68	11.54
12	2,3-dimethylbutane	11.21	11.48	11.24	11.51
13	2,4-dimethylpentane	11.29	11.25	11.21	11.18
14	2,2,3-trimethylbutane	11.37	11.33	11.54	11.30
15	2,2,4-trimethylpentane	11.43	11.10	11.30	10.97
16	2,2,3,3-tetramethylbutane	11.97	11.17	11.95	11.05
25	chloromethane	13.36	12.86	12.43	13.01
27	dichloromethane	12.85	12.99	12.90	13.02
32	trichloromethane	12.99	13.09	13.51	12.97
34	chloroethane	12.41	12.54	12.32	12.65
37	1,1-dichloroethane	12.59	12.70	12.88	12.67
38	1,2-dichloroethane	12.66	12.75	12.57	12.69
42	1,1,1-trichloroethane	13.92	12.82	13.71	12.69
43	1,1,2-trichloroethane	12.50	12.90	12.92	12.71
56	1-chloropropane	12.10	12.30	11.98	12.33
57	2-chloropropane	12.37	12.31	12.30	12.37
58	2,2-dichloropropane	13.10	12.47	13.13	12.39
59	1,2-dichloropropane	12.36	12.51	12.33	12.40
60	1,2,3-trichloropropane	12.37	12.72	12.76	12.43
61	1,3-dichloropropane	12.22	12.51	12.48	12.38
62	methanol	12.05	11.79	11.61	11.78
63	ethanol	11.54	11.65	11.45	11.74
64	propanol	11.28	11.41	11.17	11.43
65	2-propanol	11.21	11.53	11.49	11.65
66	2-methyl-2-propanol	11.96	11.42	11.84	11.50
67	butanol	11.12	11.18	11.12	11.11
73	dimethyl ether	11.53	11.78	11.33	11.90
74	diethyl ether	10.88	11.35	10.90	11.39
75	dipropyl ether	10.76	10.88	10.72	10.74
76	<i>tert</i> -butyl methyl ether	11.55	11.26	11.14	11.24

^a The entries in the last two columns that are bold are within ± 0.21 of the calculated anticonnectivity values.

**Figure 2.** Optimization of the variable connectivity index $1\chi^f$ for the modeling of $-\log k_{\text{OH}}$ values.

value of the weight that represents chlorine atoms was optimized toward infinity. Very large values of weights for particular atom means that the bonds which are formed by this atom do not make a significant contribution toward the molecular descriptor and are not important for the modeled property. However, it is possible that this atom even suppresses the modeled property, but to see if that is the case, one has to consider the possibility of anticonnectivity.

Table 3. Organic Compounds Considered and the Calculated Connectivity Indices

ID	name	1χ	$1\chi^v$	$1\chi^f$
1	ethane	1.00000	1.00000	0.44020
2	<i>n</i> -butane	1.91421	1.91421	1.03927
3	2,2-dimethylpropane	2.00000	2.00000	1.15588
4	<i>n</i> -pentane	2.41421	2.41421	1.34493
5	2-methylpropane	1.73205	1.73205	0.96305
6	<i>n</i> -propane	1.41421	1.41421	0.73362
8	2-methylbutane	2.27006	2.27006	1.27634
9	2-methylpentane	2.77006	2.77006	1.58199
10	3-methylpentane	2.80806	2.80806	1.58963
11	2,2-dimethylbutane	2.56066	2.56066	1.47451
12	2,3-dimethylbutane	2.64273	2.64273	1.51817
13	2,4-dimethylpentane	3.12590	3.12590	1.81906
14	2,2,3-trimethylbutane	2.94338	2.94338	1.71967
15	2,2,4-trimethylpentane	3.41650	3.41650	2.01723
16	2,2,3,3-tetramethylbutane	3.25000	3.25000	1.92351
25	chloromethane	1.00000	0.37796	-0.25124
27	dichloromethane	1.41421	0.53452	-0.41871
32	trichloromethane	1.73205	0.65465	-0.54966
34	chloroethane	1.41421	0.97437	0.15746
37	1,1-dichloroethane	1.73205	1.01379	-0.04542
38	1,2-dichloroethane	1.91421	1.03452	-0.11306
42	1,1,1-trichloroethane	2.00000	1.06695	-0.20581
43	1,1,2-trichloroethane	2.27006	1.11195	-0.30830
56	1-chloropropane	1.91421	1.47437	0.46311
57	2-chloropropane	1.73205	1.37292	0.45881
58	2,2-dichloropropane	2.00000	1.37796	0.24808
59	1,2-dichloropropane	2.27006	1.47108	0.19594
60	1,2,3-trichloropropane	2.80806	1.56924	-0.06694
61	1,3-dichloropropane	2.41421	1.53452	0.19260
62	methanol	1.00000	0.44721	1.12358
63	ethanol	1.41421	1.02333	1.30306
64	propanol	1.91421	1.52333	1.60871
65	2-propanol	1.73205	1.41290	1.46140
66	2-methyl-2-propanol	2.00000	1.72361	1.60448
67	butanol	2.41421	2.02333	1.91437
73	dimethyl ether	1.41421	0.81650	1.14261
74	diethyl ether	2.41421	1.99156	1.68573
75	dipropyl ether	3.41421	2.99156	2.29704
76	<i>tert</i> -butyl methyl ether	2.56066	2.11237	1.81325

That is why we have tested if the optimization of the variable connectivity index of order one in the anticonnectivity region improves the calculation capabilities of the linear regression model.

The definition of the variable index was extended also to the negative solution of eq 3. Simplex optimization found the minimal RMS error when weights for carbon and oxygen atoms were optimized in the “connectivity” region and the contribution of the chlorine atom entered the “anticonnectivity” region. The corresponding optimal weights for carbon, oxygen, and chlorine atoms were 1.272, -0.6513, and (-)5.974, respectively. The negative sign in parentheses indicates that the weight is in the “anticonnectivity” region. In Table 3, we have listed the optimal variable connectivity indices associated with the above optimal weights. For comparison, we also show the calculated values for the connectivity index 1χ of Randić and the valence connectivity index $1\chi^v$ of Kier and Hall. The dashed line in Figure 2 shows how the RMS error decreased when the anticonnectivity term was included in the calculation. The final regression model was $-\log k_{\text{OH}} = -0.777 \times 1\chi^f + 12.7$. When allowing for the presence of the negative contribution of chlorine atoms, the RMS error dropped to 0.343 log units. The calculated $-\log k_{\text{OH}}$ values versus the experimental values are presented in Figure 3.

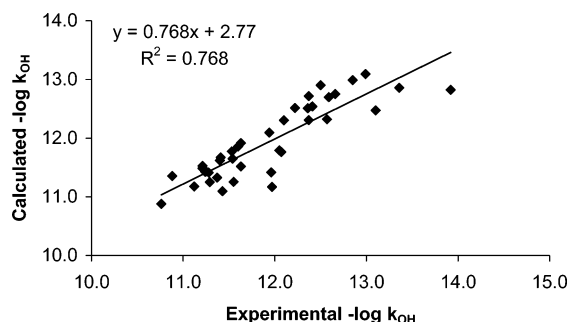


Figure 3. Calculated vs experimental $-\log k_{\text{OH}}$ values.

Although the reduction of RMS error is not dramatic, we can conclude that the new model represents a substantial improvement of its quality since we get more structural information about the molecular features that control the modeled property. The new anticonnectivity model correctly predicts relative influences of the individual atoms present in the molecule on the overall molecular reactivity with OH radicals. The conclusions, which can be extracted from the anticonnectivity model, are in full agreement with the Kwok and Atkinson's fragment contribution QSPR technique.⁵⁵

5. DISCUSSION

Let us revisit Table 3, in which we have listed the connectivity indices $^1\chi$, $^1\chi^v$, and $^1\chi^f$ for the 39 organic compounds. If we compare the connectivity indices with the valence connectivity indices, of course, for compounds 1–16 (alkanes), there is no difference. Both the valence connectivity indices for chlorine- and oxygen-containing compounds show a visible decrease in magnitudes. In the case of oxygen-containing compounds, the decrease is approximately constant (0.40 ± 0.10), while for chlorine compounds, the decrease is roughly proportional to the number of chlorine atoms. This in itself is interesting, as it shows how the valence connectivity index indirectly parallels the indicator variable that counts chlorine atoms, as used, for instance, in CODESSA software.^{56,57} More striking is the difference between the relative magnitude of the variable connectivity index and the other two connectivity indices. The variable connectivity indices for alkanes (compounds 1–16 in Table 3) are now all considerably smaller, which indicates that the role of carbon atoms for the considered property ($-\log k_{\text{OH}}$) has decreased. On the other hand, the role of oxygen has increased sufficiently to counterbalance the reduced role of carbon atoms in the case of alcohols and ethers (compounds 62–76 in Table 3). Most interesting and most significant are the changes in the relative magnitudes of the variable connectivity values for chlorine-containing compounds. The new values for the variable connectivity indices have dramatically reduced their magnitudes and, in half of the compounds, even assumed negative values, because the anticonnectivity role of chlorine atoms in such compounds overcomes the positive contributions of carbon atoms. The variable connectivity index should be given the credit for the considerable improvement of the simple regression in comparison with the regression based on the ordinary connectivity index and the valence connectivity index (Table 4). By inclusion of the anticonnectivity term, a further improvement follows, which, even though being minor, is significant, because we are obviously close to the limit of encoding capabilities of the single topological index.

Table 4. Regression Parameters for Linear Regression Models Using Various Modifications of the Connectivity Index of Order One

model	description	b_1	b_0	RMS	r	F
$^1\chi$	simple connectivity	−0.549	13.1	0.620	0.493	11.9
$^1\chi^v$	valence connectivity	−0.581	13.0	0.530	0.668	29.8
$^1\chi^f$	variable connectivity	−0.633	13.0	0.352	0.869	114.4
$^1\chi^a$	variable anticonnectivity	−0.777	12.7	0.343	0.876	122.5

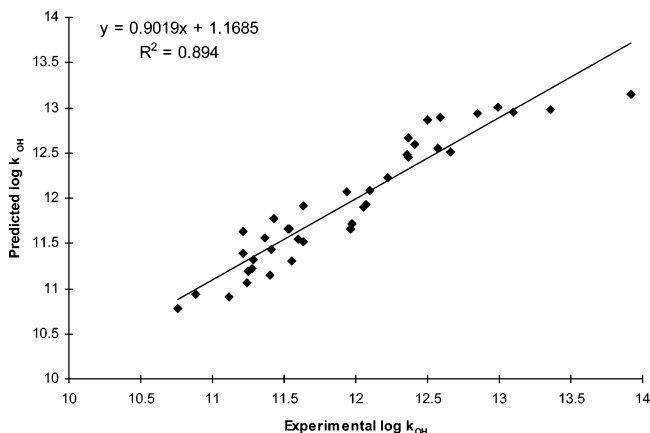
There were previous arguments that the variable connectivity model is not a single-descriptor model but contains “hidden” variables.¹⁴ Therefore, a regression model with three variables should be compared with the three-parameter MLR model. This statement is not completely correct. There were previous attempts to differentiate between heteroatoms by introducing fixed weights for individual atoms to the different types of connectivity indices.^{10,13,14} It was soon clear that such selection of the weights was suitable just for the modeling of certain properties. Our optimization procedure could be replaced also by the calculation of a very large number of fixed descriptors where all combinations of the variables are taken into account. Therefore, in our opinion, we could still claim that our model is a single-parameter model, but we must be careful about the number of selected variables of the augmented connectivity matrix. By increasing the number of variables, we are screening a very large modeling space; therefore, we are increasing the possibility to obtain a chance correlation. But, we think that if one is careful about the selection of a suitable training set, that is, there are at least 10 examples in each group of compounds and that these groups are approximately equally represented in the training set, then the chances of obtaining a random model are substantially reduced.

In Table 2, we have listed calculated $-\log k_{\text{OH}}$ values based on the anticonnectivity values shown in the last column of Table 3 and have compared them with the corresponding values as reported by Bakken and Jurs (column B & J) and values obtained by using the program CODESSA. As we see, most of the calculated values are similar. We have shown in bold all those values of the last two columns that exceed ± 0.21 of the values given by the anticonnectivity index. The threshold 0.21 was selected as the minimum RMS error obtained by any of these models. In the case of CODESSA, this includes 38 out of 39 compounds, the exception being 1,2,3-trichloropropane. Both the Bakken and Jurs approach (based on 10 descriptors) and the use of CODESSA (based on five descriptors) gave better regressions, which is not surprising in view of the fact that we are comparing a simple regression with linear regressions using five descriptors. We should also add that the sample of Bakken and Jurs had 312 compounds from which we extracted 39, so their result (amplified by the consideration of nonlinear regression models) is of much broader scope than our task considered in this work, which is more focused on illustrating the concept of anticonnectivity, rather than being an effort to find the best regression using several molecular descriptors.

The comparison of our results with similar results obtained by CODESSA is also instructive. In Table 5, we show the selected CODESSA descriptors for the best MLR model ($r = 0.963$, RMS = 0.21); the calculated versus experimental $-\log k_{\text{OH}}$ values obtained using the five-parameter MLR model are illustrated in Figure 4. Observe the presence of

Table 5. Selected Structural Descriptors for the Best MLR Model ($r = 0.963$, $F = 84$, $RMS = 0.21$)

number	coefficient	type of descriptor
0	7.5475	intercept
1	-2.0830	Randić index (order one)
2	0.03364	molecular weight
3	0.02713	Wiener index
4	8.0667	relative number of H atoms
5	5.3081	relative number of Cl atoms

**Figure 4.** Calculated versus experimental $-\log k_{OH}$ values obtained using five parameters.

three size-dependent descriptors (Randić index of order 1, which is the simple connectivity index $^1\chi$, the molecular weight, and the Wiener index). In addition, we have two indicator variables, the relative number of hydrogen atoms and the relative number of chlorine atoms. This information, on its own, speaks in favor of the anticonnectivity index, which, as we have seen, was able somehow to combine and compact the same information into a single numerical entry. It is, therefore, not surprising that it produces a good, possibly the best, single-variable regression. In fact, the anticonnectivity index gives better regression than CODESSA using two descriptors, but CODESSA with three descriptors gives a better RMS than the anticonnectivity index model. It will be of considerable interest to see if CODESSA can be “upgraded” to include the variable connectivity and anticonnectivity index and, when so-upgraded, to see which are to be the dominant molecular descriptors for $-\log k_{OH}$ and other similar properties that require the anticonnectivity descriptor.

6. CONCLUSIONS

The aim of this work was to extend the definition of the modified variable connectivity index of order zero, which accounts for a combination of positive and negative bond and atom contributions to the higher connectivity indices. A generalization of the zero-order connectivity index was introduced. The encoding capabilities of the zero- and first-order connectivity indices were presented in the modeling of the atmospheric reaction rate constants of selected organic compounds with OH radicals. The data set was constructed from 39 organic compounds containing carbon, oxygen, and chlorine atoms. The RMS error was significantly reduced when the variable connectivity index was used instead of the fixed connectivity index or the valence connectivity index. Although the introduction of the negative contribution

of chlorine atoms further lowered the RMS error to 0.343 log units, the main advantage of the recently introduced modified connectivity index remains its aptitude for structural interpretation. The modified variable connectivity index of order one correctly predicted the positive influence of carbon and oxygen atoms on modeled reaction rate constants as well as the suppressive influence of chlorine atoms. It must be mentioned that the modified connectivity index still cannot provide better correlation than the QSAR model using linear or nonlinear multivariate models, but it can offer a much better structural interpretation because the user can gain direct knowledge about which part of the molecule is an enhancer and which is a suppressor of the modeled property.

ACKNOWLEDGMENT

The authors acknowledge financial support by the Ministry of Higher Education and Science of the Republic of Slovenia (Grant L1-6709).

REFERENCES AND NOTES

- (1) Todeschini, R.; Consonni, V. *The Handbook of Molecular Descriptors, in the Series of Methods and Principles in Medicinal Chemistry*; Mannhold, R., Kubinyi, H., Timmerman, H., Eds.; Wiley-VCH: New York, 2000; Vol. 11.
- (2) Katritzky, A. R.; Gordeeva, E. V. Traditional topological indexes vs electronic, geometrical, and combined molecular descriptors in QSAR/QSPR research. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 835–857.
- (3) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-chemical descriptors in QSAR/QSPR studies. *Chem. Rev.* **1996**, *96*, 1027–1043.
- (4) Devillers, J.; Balaban, A. T. *Topological Indices and Related Descriptors in QSAR and QSPR*; Gordon and Breach: Amsterdam, 1999.
- (5) Grassy, G.; Calas, B.; Yasri, A.; Lahana, R.; Woo, J.; Iyer, S.; Kaczorek, M.; Floc'h, R.; Buelow, R. Computer-assisted rational design of immunosuppressive compounds. *Nat. Biotechnol.* **1998**, *16*, 748.
- (6) Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **1947**, *69*, 17–20.
- (7) Randić, M.; Zupan, J. On interpretation of well-known topological indices. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 550–560.
- (8) Randić, M.; Zupan, J. On the structural interpretation of topological indices. In *Topology in Chemistry—Discrete Mathematics of Molecules*; Rouvray, D. H., King, R. B., Eds.; Horwood Publishers: Chichester, England, 2002; pp 249–291.
- (9) Randić, M.; Pompe, M. The variable molecular descriptors based on distance related matrices. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 575–581.
- (10) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: London, 1976.
- (11) Ivanciuc, O.; Balaban, A. T. Design of Topological indices. Part 5. Precision and error in computing graph theoretical invariants for molecules containing heteroatoms and multiple bonds. *MATCH* **1994**, *30*, 117–139.
- (12) Balaban, A. T. Highly discriminating distance-based topological index. *Chem. Phys. Lett.* **1982**, *89*, 399–404.
- (13) Antipin, I. S.; Arslanov, N. A.; Palyulin, V. A.; Kononov, A. I.; Zefirov, N. S. Solvation topological index. Topological description of dispersion interaction. *Dokl. Akad. Nauk. SSSR* **1991**, *316*, 925–927.
- (14) Zefirov, N. S.; Palyulin, V. A. QSAR for boiling points of “Small” sulfides. Are the “High-quality structure–property–activity regressions” the real high quality QSAR models? *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1022–1027.
- (15) Randić, M. Novel graph theoretical approach to heteroatoms in quantitative structure–activity relationships. *Chemom. Intell. Lab. Syst.* **1991**, *10*, 213–227.
- (16) Randić, M. On computation of optimal parameters for multivariate analysis of structure–property relationship. *J. Comput. Chem.* **1991**, *12*, 970–980.
- (17) Randić, M.; Pompe, M. On characterization of the CC double bond in alkenes. *SAR QSAR Environ. Res.* **1999**, *10*, 451–471.
- (18) Randić, M.; Basak, S. C. Construction of high quality structure–property–activity regressions: the boiling points of sulfides. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 899–905.

- (19) 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.
- (20) Liu, D.; Zhong, C. Modeling of the heat capacity of polymers with the variable connectivity index. *Polym. J.* **2002**, *34*, 954–961.
- (21) Krenkel, G.; Castro, E. A.; Toropov, A. A. Improved molecular descriptors based on the optimization of correlation weights of local graph invariants. *THEOCHEM* **2001**, *542*, 107–113.
- (22) Zhong, C.; He, J.; Xue, C.; Li, Y. A QSAR study on inhibitory activity of 1-phenylbenzimidazoles against the platelet-derived growth factor receptor. *Bioorg. Med. Chem.* **2004**, *12*, 4009–4015.
- (23) Hu, Q.-N.; Liang, Y.-Z.; Wang, Y.-L.; Xu, C.-J.; Zeng, Z.-D.; Fang, K.-T.; Peng, X.-L.; Hong, Y. External factor variable connectivity index. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 773–778.
- (24) Hu, Q.-N.; Liang, Y.-Z.; Peng, X.-L.; Yin, H.; Fang, K.-T. Structural interpretation of a topological index. 1. External factor variable connectivity index (EFVCI). *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 437–446.
- (25) Kezele, N.; Klasinc, L.; von Knop, J.; Ivaniš, S.; Nikolić, S. Computing the variable vertex-connectivity index. *Croat. Chem. Acta* **2002**, *75*, 651–661.
- (26) Pompe, M.; Veber, M.; Randić, M.; Balaban, A. T. Using variable and fixed topological indices for the prediction of reaction rate constants of volatile unsaturated hydrocarbons with OH radicals. *Molecules* **2004**, *9*, 1160–1176.
- (27) Randić, M.; Pompe, M.; Mills, D.; Basak, S. C. Variable connectivity index as a tool for modeling structure–property relationship. *Molecules* **2004**, *9*, 1177–1193.
- (28) Lučić, B.; Miličević, A.; Nikolić, S.; Trinajstić, N. On variable Wiener index. *Indian J. Chem., Sect. A* **2003**, *42*, 1279–1282.
- (29) Miličević, A.; Nikolić, S. On variable Zagreb indices. *Croat. Chem. Acta* **2004**, *77*, 97–101.
- (30) Perdih, A.; Perdih, B. Topological indices derived from the G(a,b,c) matrix, useful as physicochemical property indices. *Acta Chim. Slov.* **2004**, *51*, 598–609.
- (31) Randić, M.; Hansen, P. J.; Jurs, P. C. Search for useful graph theoretical invariants of molecular-structure. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 60–68.
- (32) Pompe, M. Variable connectivity index as a tool for solving ‘anti-connectivity’ problem. *Chem. Phys. Lett.* **2005**, *404*, 296–299.
- (33) Randić, M. Characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.
- (34) Scanlon, J. T.; Willis, D. E. Calculation of flame ionization detector relative response factors using the effective carbon number concept. *J. Chromatogr. Sci.* **1985**, *23*, 333–340.
- (35) Massart, D. L. *Chemometrics: A textbook, Data handling in science and technology*; Elsevier: Amsterdam, 1988; Vol. 2.
- (36) Atkinson, R. Gas-Phase Reactions of the Hydroxyl Radicals. *Chem. Rev.* **1986**, *86*, 69–201.
- (37) Atkinson, R. Kinetics and mechanisms of the gas-phase reactions of the hydroxyl radical with organic compounds. *J. Phys. Chem. Ref. Data* **1989**, *Monograph 1*, 1–246.
- (38) Atkinson, R. Gas-phase tropospheric chemistry of organic compounds. *J. Phys. Chem. Ref. Data* **1994**, *Monograph 2*, 1–216.
- (39) Atkinson, R. Estimation of Gas-Phase Hydroxyl Radical Rate Constants for Organic Chemicals. *Environ. Toxicol. Chem.* **1988**, *7*, 435–442.
- (40) Heickler, J. The correlation of Rate Coefficients for H-Atom Abstraction by OH Radicals with C–H Bond Dissociation Enthalpies. *Int. J. Chem. Kinet.* **1981**, *13*, 651–665.
- (41) Jolly, G. S.; Paraskevopoulos, G.; Singleton, D. L. Rates of OH radical reactions. XII. The reaction of OH with $c\text{-C}_3\text{H}_6$, $c\text{-C}_5\text{H}_{10}$, and $c\text{-C}_7\text{H}_{14}$. Correlation of hydroxyl rate constants with bond dissociation energies. *Int. J. Chem. Kinet.* **1984**, *17*, 1–10.
- (42) Grosjean, D.; Williams, E. L., II. Environmental persistence of organic compounds estimated from structure–reactivity and linear free-energy relationships—unsaturated aliphatics. *Atmos. Environ.* **1992**, *26A*, 1395–1405.
- (43) Sekušak, S.; Güsten, H.; Sabljčić, A. An ab initio study on reactivity of chloro ethane with hydroxyl radical: Application of G2 theory. *J. Phys. Chem.* **1996**, *100*, 6212–6224.
- (44) Klamt, A. Estimation of gas-phase hydroxyl radical rate constants of oxygenated compounds based on molecular orbital calculations. *Chemosphere* **1996**, *32*, 717–726.
- (45) Melissas, V. S.; Truhlar, D. G. Interpolated variational transition-state theory and semiclassical tunneling calculations of the rate constant of the reaction $\text{OH} + \text{C}_2\text{H}_6$ at 200–3000 K. *J. Phys. Chem.* **1994**, *98*, 875–886.
- (46) King, M. D.; Canosa-Mas, C. E.; Wayne, R. P. A structure–activity relationship (SAR) for predicting rate constants for the reaction of NO_3 , OH and O_3 with monoalkenes and conjugated dienes. *Phys. Chem. Chem. Phys.* **1999**, *1*, 2239–2246.
- (47) Tosato, M. L.; Chiorboli, C.; Eriksson, L.; Jonsson, L. Multivariate modelling of the rate constant of the gas-phase reaction of haloalkanes with the hydroxyl radical. *Sci. Total Environ.* **1991**, *109/110*, 307–325.
- (48) Eriksson, L.; Rännar, S.; Sjöström, M.; Hermens, J. L. M. Multivariate QSARs to model the hydroxyl radical rate constant for halogenated aliphatic hydrocarbons. *Environmetrics* **1994**, *5*, 197–208.
- (49) Medven, Z.; Güsten, H.; Sabljčić, A. Comparative QSAR study on hydroxyl radical reactivity with unsaturated hydrocarbons: PLS versus MLR. *J. Chemom.* **1996**, *10*, 135–147.
- (50) Bakken, G. A.; Jurs, P. C. Prediction of hydroxyl radical rate constants from molecular structure. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1064–1075.
- (51) Gramatica, P.; Pilutti, P.; Papa, E. Validated QSAR prediction of OH tropospheric degradation of VOCs: Splitting into training – test set and consensus modeling. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1794–1802.
- (52) Güsten, H.; Medven, Z.; Sekušak, S.; Sabljčić, A. A Predicting Tropospheric Degradation of Chemicals: From Estimation to Computations. *SAR QSAR Environ. Res.* **1995**, *4*, 197–209.
- (53) Güsten, H. Predicting the abiotic degradability of organic pollutants in the troposphere. *Chemosphere* **1999**, *38*, 1361–1370.
- (54) Bakken, G. A.; Jurs, P. C. Prediction of hydroxyl radical rate constants from molecular structure. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1064–1075.
- (55) Kwok, E. C.; Atkinson, R. Estimation of hydroxyl radical reaction rate constants for gas-phase organic compounds using a structure–reactivity relationship: an update. *Atmos. Environ.* **1995**, *29*, 1685–1695.
- (56) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-chemical descriptors in QSAR/QSPR studies. *Chem. Rev.* **1996**, *96*, 1027–1043.
- (57) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *CODESSA Training Manual*; University of Florida: Gainesville, FL, 1995.

CI050125T