Optimal Molecular Descriptors Based on Weighted Path Numbers

Milan Randić*,† and Subhash C. Basak‡

Department of Mathematics and Computer Science, Drake University, Des Moines, Iowa 50311, National Institute of Chemistry, Hajdrihova 19, P.O. Box 3430, Ljubljana, Slovenia, and Natural Resources Research Institute, Center for Water and Environment, University of Minnesota, Duluth, Minnesota 55811

Received April 15, 1998

We consider weighted path numbers as molecular descriptors for structure−property−activity studies. However, instead of using prescribed weights for paths we have optimized the weights so that the standard error in regression analysis is as small as possible. In particular we consider the boiling points of alcohols and use of weighted paths to differentiate an oxygen from a carbon atom.

INTRODUCTION

Despite the large number of molecular descriptors available for use in multiple regression analysis (MRA), we still have no good descriptors for many molecular properties of interest. Most descriptors are size dependent and can produce apparently good correlation when used for samples of compounds that include molecules of different size. However, when the same descriptors are used to describe variations in properties among molecules of the same size, they often show limited ability to characterize well the molecular shape. To illustrate the point, consider correlation of molar refraction R_m of alkanes with the connectivity indices ${}^1\chi$ [1] and ${}^2\chi$.[2] According to Kier and Hall[3] for 46 alkanes having from five to nine carbon atoms a correlation with single connectivity index gives:

$$R_m = 10.440\ {}^1\chi + 1.369 \qquad \text{with } r = 0.9520;\ s = 1.81 \tag{1}$$

where r is the coefficient of regression and s is the standard error. Superficially eq 1 appears fair if one recollects that for the set of compounds considered molar refraction varies from 25.267 (*n*-pentane) to 48.757 (2,2,5,5-tetramethylhexane). However, ${}^1\chi$ only correlates well with the molecular size. Within the 17 octane, R_m varies from 38.719 (3-methyl-3-ethylpentane) to 39.264 (2,2,4-trimethylpentane). This gives the maximal difference for R_m of only 0.545, which is well within the standard error of the regression based on ${}^1\chi$ ($s = 1.81$).

Use of two descriptors substantially improves the correlation as is evident by the decrease in the standard error by an order of magnitude when both ${}^1\chi$ and ${}^2\chi$ are used (ref 3, p 108):

$$R_m = 2.207\ {}^2\chi + 7.756\ {}^1\chi + 3.707 \qquad \text{with } r = 0.9997;\ s = 0.121 \tag{2}$$

It should not be surprising that when larger alkanes are considered no single descriptor will perform admirably. This is because as the size of molecules increases *additional* structural elements may play a role in addition to the connectivity and branching that simple indices consider. Using a *three*-parameter combination of the valence connectivity indices and the ordinary connectivity indices, Kier and Hall[3] reported the standard error $s = 0.23$ for the same data, and using a *six*-parameter combination of connectivity indices they obtained the standard error $s = 0.15$ for the molar refraction of alkenes.

These findings can be contrasted with the regression results based on the use of the path numbers as descriptors instead on the connectivity indices. The standard error $s = 0.17$ was obtained when a *single* path number was used as a descriptor, which was slightly improved to $s = 0.14$ when *four* path numbers were used.[4] Clearly, for molar refraction of alkenes, the path numbers outperform the connectivity indices even though the count of paths does not differentiate between C−C bonds and C=C bonds, which are differentiated by the valence connectivity indices. The qualification "outperform" was not based on the small differences in the standard errors between the alternative approaches that may not be so significant, but on the *number* of descriptors used.

If we consider some other property the opposite may be the case. This clearly points to the importance of descriptor selection. Different molecular descriptors and different topological indices will better describe different molecular properties. That different properties may require different descriptors is obvious from Figure 1, in which for C_8H_{18} octane isomers we plotted molar refraction against the boiling points (BPs). The two properties clearly do not depend on the same structural factors. If, however, each of the two properties is separately correlated with ${}^1\chi$, we see that molar refraction shows no correlation at all (Figure 2) but the connectivity index gives a fair correlation (Figure 3).

OPTIMIZED MOLECULAR DESCRIPTORS

In Table 1 we illustrate a number of properties of octanes and the best single descriptor. As we see from several

* To whom correspondence should be addressed.
† Drake University and National Institute of Chemistry.
‡ University of Minnesota.

10.1021/ci9800763 CCC: $18.00
Published on Web 02/05/1999

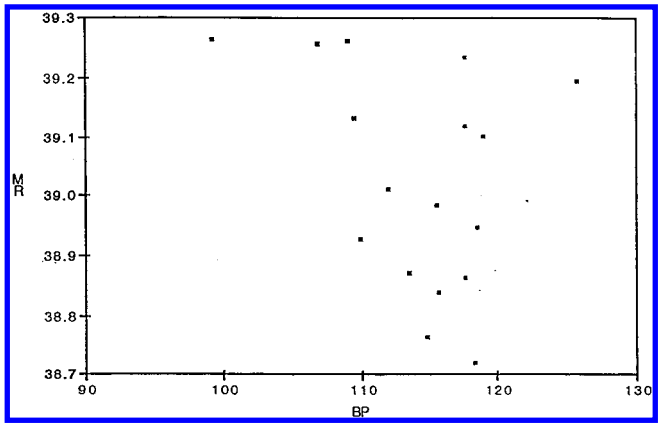

Figure 1. Correlation of molar refraction R_m against the boiling points (BPs) for *n*-octane isomers.

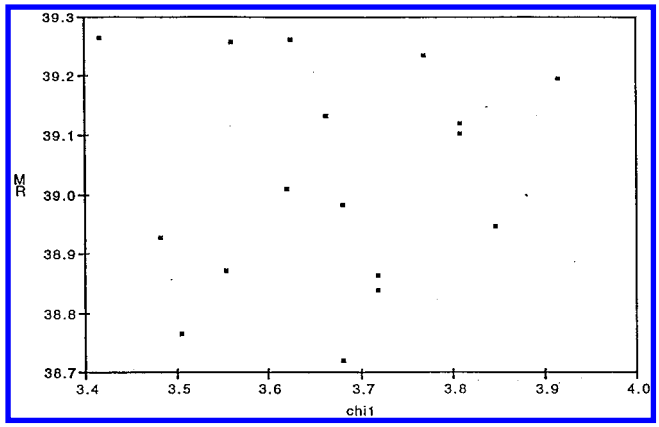

Figure 2. Plot of molar refraction R_m against the connectivity index $^1\chi$.

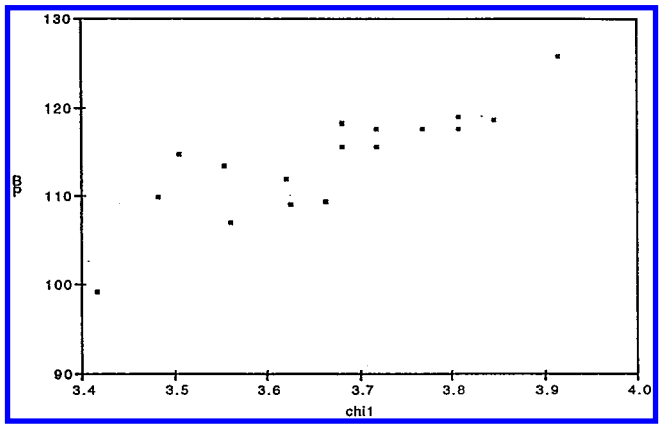

Figure 3. Plot of the BPs against the connectivity index $^1\chi$.

hundreds of descriptors reported in the literature, at most a dozen appear to emerge as the best *single* characterization of diverse physicochemical properties of octanes. Are the descriptors shown in Table 1 the best, or could there be descriptors that are even better, but we have hitherto been unsuccessful in finding them?

Two questions are considered in this paper. (1) How to search for descriptors (whether they are to be used as a single structural variable or are to be combined with other indices) that can best describe physicochemical properties of alkanes? (2) How to extend such descriptors to molecules having double bonds, triple bonds, aromatic CC bond, and heteroatoms?

Table 1. The Best Single Descriptors for Various Properties of Octanes C_8H_{18}

property	*r*	*s*	descriptor
eccentric factor	0.992	0.0039	$^2\chi$
density	0.979	0.0025	$^3\chi$
molecular volume	0.978	0.554	$^3\chi$
molar refraction	0.970	0.046	$^3\chi$
surface tension	0.964	0.241	$^2\chi-{}^3\chi$
motor octane number	0.959	7.27	I_{WD}
heat of vaporization	0.958	0.429	Z
entropy	0.954	1.40	$m^{1/2}$
heat of atomization	0.931	0.725	$1/^2\chi$
heat of formation	0.931	0.471	$1/^2\chi$
C^{13} chemical shift sum	0.929	19.1	W/Z
critical temperature	0.889	4.59	$^1\chi-{}^2\chi$
boiling points	0.888	2.90	Z
critical volume	0.849	8.67	χ^V
critical pressure	0.668	1.10	$1/^2\chi$

Because the pool of molecular descriptors has increased dramatically during the last decade, the problem of selecting molecular descriptors is a current topic of interest to many researchers.[5] For example, the computer program CODESSA[6] evaluates thousands of molecular descriptors, about a half of which are graph theoretical (topological indices) and the other half are quantum chemical descriptors. In contrast to the prevailing practice (which has its advantages and disadvantages), in which one selects a few descriptors from a large pool of descriptors, we will consider the other extreme: A few adjustable descriptors are selected and their variable part is optimized. In doing this, we will nevertheless require that descriptors have a direct, even if not necessarily transparent, structural meaning. Use of a minimal number of descriptors has not only the obvious statistical advantage but it may allow a simpler interpretation of the resulting regression equation. How can one plan to have very good regression with few descriptors where others used many? The answer is in optimization of molecular descriptors for the particular application.

The search for optimized molecular descriptors has been outlined in QSAR (Quantitative Structure−Activity Relationship),[3,7−9] but apparently has not yet received due attention. There are at least three distinctive routes to optimization of molecular descriptors.

(1) One may try to optimize the functional form.[7−9] For example, instead of using the connectivity index $^1\chi$,[1] the Wiener index W,[10] and the Hosoya index Z [11] as a single descriptor, one can consider various powers of such indices. For the BPs of smaller alkanes (from ethane to heptane isomers), the smallest standard deviations are found when one uses $\chi^{1/2}$ ($s = 2.83$), $W^{1/4}$ ($s = 4.42$), and $Z^{-1/3}$ ($s = 3.54$).[9]

(2) One may try to optimize the diagonal entries of an adjacency matrix to differentiate between atoms of different types.[12−14] For example, if one does not differentiate between carbon atom and oxygen atom when considering the BPs of hexanols, one obtains for the standard error $s = 7.86$ °C when the connectivity index is used as molecular descriptor. If, however, the connectivity index is constructed from the adjacency matrix in which the entries on the main diagonal are viewed as variables (one for carbon atoms and one for oxygen) the modified connectivity index leads to regression with the standard error of only 3.43 °C.[13]

(3) One can try to optimize off-diagonal entries of adjacency matrix, i.e., one can introduce variable weights

Table 2. Classification of Topological Indices

		reference
integers		
Wiener number	W	10
Hosoya index	Z	11
path numbers	p_1, p_2, p_3, ...	18
Centric index	C	19
Schultz index	MTI	20
real numbers		
Balaban's index	J	21
identification number	ID	22
the leading eigenvalue	λ_1	23
branching index	β	24
weighted ID number	WID	25
information theoretic		
Bonchev, Trinajstić		26
Sarkar et al.	IC	27
Basak et al.	SIC	28
Basak et al.	CIC	29
novel matrices		
expanded Wiener	Tratch et al.	30
total topological state	Hall	31
Wiener matrix index	Randić	32
Hosoya matrix index	Randić	33
detour matrix index	Trinajstić et al.	34
Szeged matrix index	Gutman et al.	35
Cluj matrix index	Diudea et al.	36
path matrix indices	Randić et al.	37

for bonds of different types.[15] For example, when paths of length two, three, and four are used as descriptor for correlation of R_m for 17 isomers of 1-heptene, the standard error for R_m is $s = 0.11$. If instead one introduces a variable weight for CC double bond in alkenes and then minimizes the standard error by adjusting the weight, the standard error further decreases to 0.08.[3]

CONSTRUCTION OF DESCRIPTORS SUITABLE FOR OPTIMIZATION

The first task when considering optimization of molecular descriptors is to find a generalized form for the descriptor that allows introduction of variables to be optimized. Most molecular descriptors have been designed "rigidly", i.e., the algorithm for their construction is fixed so that once the molecule is selected (including molecules having heteroatoms) the invariant of interest can be computed exactly. In Table 2 we list some better- and less-known molecular topological indices, and some more recent indices.[18−37] They can be classified as integers (the first-generation indices according to Balaban[16]), as real numbers (the second-order indices according to Balaban), as indices derived from information−theoretic considerations of chemical graphs,[17] and as indices derived from novel matrices constructed for molecular graphs. Some indices can be ordered in a natural way forming a sequence of structurally related molecular descriptors (Table 3[38−44]). To this list of diverse molecular descriptors we should add as a separate class indices that have inherent flexibility involving a variable part that can be optimized for different applications.

Topological indices are usually referred to as two-dimensional (2-D) indices. Several 3-D indices that have been introduced in recent years are listed in Table 4.[45−52] These indices can also be generalized to involve a variable part along the lines outlined in this paper for path numbers as 2-D topological indices.

Table 3

sequential descriptors		reference
path numbers	p_1, p_2, p_3, ...	38
connectivity indices	${}^1\chi$, ${}^2\chi$, ${}^3\chi$, ...	1, 2
valence connectivity	${}^1\chi^m$, ${}^2\chi^m$, ${}^3\chi^m$, ...	39
kappa indices	κ_1, κ_2, κ_3, ...	40
Hosoya type indices	${}^1Z$, ${}^2Z$, ${}^3Z$, ...	41, 42
path/walk indices	p_1/w_1, p_2/w_2, ...	43, 44

Table 4

3-D indices	reference
3-D Wiener index	45
3-D connectivity	46
molecular profiles	47
shape profiles	48
surface profile	49
3-D distance matrix indices	50
3-D charge based indices	51
3-D excluded volume	52

Table 5. The Count of Paths and Weighted Paths in 2-Methyl-3-pentanol and 2-Ethyl-3-methyl-1-butene (the Numbering of Carbon Atoms Is Based on 2-Methyl-3-pentanol)

carbon atom	p_1	p_2	p_3	p_4
1	1	2	$1 + x$	1
2	3	$1 + x$	1	
3	$2 + x$	3		
4	2	$1 + x$	2	
5	1	1	$1 + x$	2
6	1	2	$1 + x$	1
7	x	$2x -$	$3x$	
molecule	$5 + x$	$5 + 2x$	$3 + 3x$	2

WEIGHTED PATHS FOR ALCOHOLS

We have already mentioned use of weighted paths to describe the CC double bond in alkenes. Each time the CC double bond is involved in the count of paths it involves the weight x. In Table 5 we illustrate the count of paths for 2-ethyl-3-methyl-1-butene. If we assume $x = 1$, then the count represents the paths in 2,3-dimethylpentane; if we assume $x = 2$, then the count represent paths in 2-ethyl-3-methyl-1-butene. If x is left undefined, then we obtain the count of paths for 2-ethyl-3-methyl-1-butene or 2-methyl-3-pentanol if CC double bond is replaced by OH group for more generalized cases where x has yet to be determined. Hence, x can be viewed as a variable to be adjusted from regression analysis to obtain the smallest standard error for a regression that is considered.

By using weighted paths for molecules having heteroatoms (rather than a hetero-bond as was the case with CC double bond in alkenes[4]), we can enormously extend our approach of variable descriptors to molecules of different chemical composition. We will here re-examine the BPs of 58 alcohols already investigated in several structure−property studies.[53−58] In Table 6 we list the weighted paths of length one to length four for the molecules considered. We view the weight x as a variable to be freely adjusted and will seek the optimal weight for the C—O bond by minimizing the standard error in a stepwise multiple regression of the BPs of alcohols.

In Table 7 we show the regression coefficient (r), the standard error (s), and the Fisher ratio (F) for various values of x when from one to four weighted paths are used as descriptors. When p_1 is used as a single descriptor the

Table 6. The Weighted Path Numbers for Aliphatic Alcohols

compound	p_1	p_2	p_3	p_4
1 methanol	x	0	0	0
2 ethanol	$1+x$	x	0	0
3 1-propanol	$2+x$	$1+x$	x	0
4 2-propanol	$2+x$	$1+2x$	0	0
5 1-butanol	$3+x$	$2+x$	$1+x$	0
6 2-butanol	$3+x$	$2+2x$	$1+x$	0
7 2-methyl-1-propanol	$3+x$	$-3+x$	$2x$	0
8 2-methyl-2-propanol	$3+x$	$3+3x$	0	0
9 1-pentanol	$4+x$	$3+x$	$2+x$	$1+x$
10 2-pentanol	$4+x$	$3+2x$	$2+x$	$1+x$
11 3-pentanol	$4+x$	$3+2x$	$2+2x$	1
12 2-methyl-1-butanol	$4+x$	$4+x$	$2+2x$	x
13 3-methyl-1-butanol	$4+x$	$4+x$	$2+x$	$2x$
14 2-methyl-2-butanol	$4+x$	$4+3x$	$2+x$	0
15 3-methyl-2-butanol	$4+x$	$4+2x$	$2+2x$	0
16 2,2-dimethyl-1-propanol	$4+x$	$6+x$	$3x$	0
17 1-hexanol	$5+x$	$4+x$	$3+x$	$2+x$
18 2-hexanol	$5+x$	$4+2x$	$3+x$	$2+x$
19 3-hexanol	$5+x$	$4+2x$	$3+2x$	$2+x$
20 2-methyl-1-pentanol	$5+x$	$5+x$	$3+2x$	$2+x$
21 3-methyl-1-pentanol	$5+x$	$5+x$	$4+x$	$1+2x$
22 4-methyl-1-pentanol	$5+x$	$5+x$	$3+x$	$2+x$
23 2-methyl-2-pentanol	$5+x$	$5+3x$	$3+x$	$2+x$
24 3-methyl-2-pentanol	$5+x$	$5+2x$	$4+2x$	$1+x$
25 4-methyl-2-pentanol	$5+x$	$5+2x$	$3+x$	$2+2x$
26 2-methyl-3-pentanol	$5+x$	$5+2x$	$3+3x$	2
27 3-methyl-3-pentanol	$5+x$	$5+3x$	$4+2x$	1
28 2-ethyl-1-butanol	$5+x$	$5+x$	$4+2x$	$1+2x$
29 2,2-dimethyl-1-butanol	$5+x$	$7+x$	$3+3x$	x
30 2,3-dimethyl-1-butanol	$5+x$	$6+x$	$4+2x$	$2x$
31 3,3-dimethyl-1-butanol	$5+x$	$7+x$	$3+x$	$3x$
32 2,3-dimethyl-2-butanol	$5+x$	$6+3x$	$4+2x$	0
33 3,3-dimethyl-2-butanol	$5+x$	$7+2x$	$3+3x$	0
34 1-heptanol	$6+x$	$5+x$	$4+x$	$3+x$
35 3-heptanol	$6+x$	$5+2x$	$4+2x$	$3+x$
36 4-heptanol	$6+x$	$5+2x$	$4+2x$	$3+2x$
37 2-methyl-2-hexanol	$6+x$	$6+3x$	$4+x$	$3+x$
38 3-methyl-3-hexanol	$6+x$	$6+3x$	$5+2x$	$3+x$
39 3-ethyl-3-pentanol	$6+x$	$6+3x$	$6+3x$	3
40 2,3-dimethyl-2-pentanol	$6+x$	$7+3x$	$6+2x$	$2+x$
41 3,3-dimethyl-2-pentanol	$6+x$	$8+2x$	$6+3x$	$1+x$
42 2,2-dimethyl-3-pentanol	$6+x$	$8+2x$	$4+4x$	3
43 2,3-dimethyl-3-pentanol	$6+x$	$7+3x$	$6+3x$	2
44 2,4-dimethyl-3-pentanol	$6+x$	$7+2x$	$4+4x$	4
45 1-octanol	$7+x$	$6+x$	$5+x$	$4+x$
46 2-octanol	$7+x$	$6+2x$	$5+x$	$4+x$
47 2-ethyl-1-hexanol	$7+x$	$7+x$	$6+2x$	$4+2x$
48 2,2,3-trimethyl-3-pentanol	$7+x$	$10+3x$	$8+4x$	3
49 1-nonanol	$8+x$	$7+x$	$6+x$	$5+x$
50 2-nonanol	$8+x$	$7+2x$	$6+x$	$5+x$
51 3-nonanol	$8+x$	$7+2x$	$6+2x$	$5+x$
52 4-nonanol	$8+x$	$7+2x$	$6+2x$	$5+2x$
53 5-nonanol	$8+x$	$7+2x$	$6+2x$	$5+2x$
54 7-methyl-1-octanol	$8+x$	$8+x$	$6+x$	$5+x$
55 2,6-dimethyl-4-heptanol	$8+x$	$9+2x$	$6+2x$	$5+4x$
56 3,5-dimethyl-4-heptanol	$8+x$	$9+2x$	$8+4x$	$5+2x$
57 3,5,5-trimethyl-1-hexanol	$8+x$	$11+x$	$7+x$	$7+2x$
58 1-decanol	$9+x$	$8+x$	$7+x$	$6+x$

outcome of the regression does not depend on x, because x is an additive constant for all paths of length one. Clearly p_1 is not an adequate descriptor for BPs in alcohols, as could have been expected, because p_1 only reflects the molecular size, having the same value for all isomers. With two descriptors we see an impressive improvement in the regression statistics. The standard error then depends on the value of x and has decreased for the value $x = 1$ to half (6.64 °C) of the previous case (13.28 °C). The case $x = 1$ corresponds to treating carbon and oxygen atoms equally, hence represents a model in which we do not differentiate the presence of heteroatom. However, even this crude model has significantly smaller standard error than the regressions based on Wiener number W, the Schultz index MTI, or the valence connectivity index χ^v. These above indices were all considered in ref 53 in regressions based on a single descriptor and produced the standard errors above 9 °C. In contrast, we use two descriptors (p_1 and p_2), but on the other hand our descriptors do not differentiate heteroatoms, whereas the descriptors used in ref 53 were designed to differentiate oxygen and carbon.

Table 7. The Regression Coefficient (r), the Standard Error (s), and the Fisher Ratio (F) for Various Values of x when Weighted Paths Are Used as Descriptors

		r	s	F
single descriptor				
p_1		0.9294	13.277	355
two descriptors				
p_1, p_2	$x = 1$	0.96530	6.6424	794
	$x = 2$	0.9931	4.269	1961
	$x = 2.2$	0.9935	4.131	2096
	$x = 2.5$	0.9938	4.045	2188
	$x = 2.6$	0.9938	4.039	2193
	$x = 2.7$	0.9938	4.044	2188
	$x = 2.8$	0.9937	4.056	2175
	$x = 3$	0.9936	4.098	2130
three descriptors				
p_1, p_2, p_3	$x = 1$	0.96647	6.5290	549
	$x = 2$	0.9931	4.295	1292
	$x = 2.2$	0.9936	4.128	1400
	$x = 2.5$	0.9941	3.979	1508
	$x = 2.6$	0.9942	3.949	1531
	$x = 2.7$	0.9942	3.926	1549
	$x = 2.8$	0.9943	3.910	1562
	$x = 3$	0.9943	3.893	1576
	$x = 3.1$	0.9943	3.891	1578
	$x = 3.5$	0.9943	3.912	1561

Now we will consider x as a variable. From Table 7 we see that as x increases the standard error decreases and has minimum for $x = 2.6$. Figure 4 shows the variation of the standard error with x in the interval $x = 1$ to $x = 3$ (fitted to a quartic function).

The regression equations using first only p_1, then p_1, p_2, and finally p_1, p_2, p_3 are

$$BP = 17.65758\, p_1 + 10.62758 \quad (6)$$

$$BP = 25.01704\, p_1 - 6.00248\, p_2 + 11.03094 \quad (7)$$

$$BP = 25.68984\, p_1 - 6.02031\, p_2 - 0.42308\, p_3 + 8.72600 \quad (8)$$

which after the descriptors have been orthogonalized are, respectively

$$BP = 17.65758\, \Omega_1 + 10.62758 \quad (9)$$

$$BP = 17.65758\, \Omega_1 - 6.00248\, \Omega_2 + 10.62758 \quad (10)$$

$$BP = 17.65758\, \Omega_1 - 6.00248\, \Omega_2 - 0.44723\, \Omega_3 + 10.62758 \quad (11)$$

Here Ω_1 is p_1, Ω_2 is the part of p_2 not paralleling p_1 (i.e., the residual of the correlation of p_2 against p_1), and Ω_3 is the part of p_3 not paralleling both p_1 and p_2 (more correctly not paralleling Ω_1 and Ω_2). The orthogonalization process was described in refs 59−65.

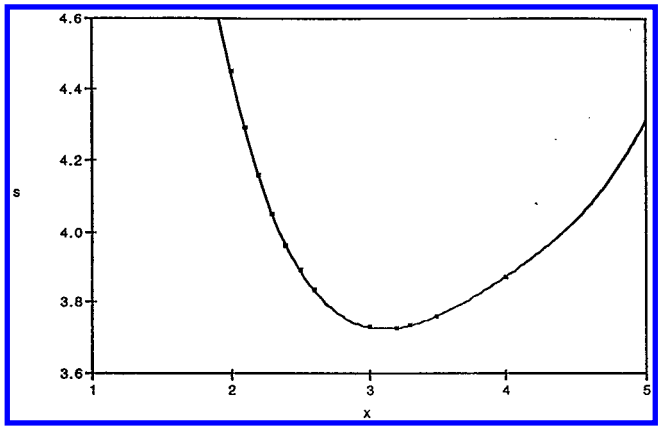

Figure 4. The variation of the standard error with the weight x (in the interval $1 \leq x \leq 3$) fitted to quartic polynomial.

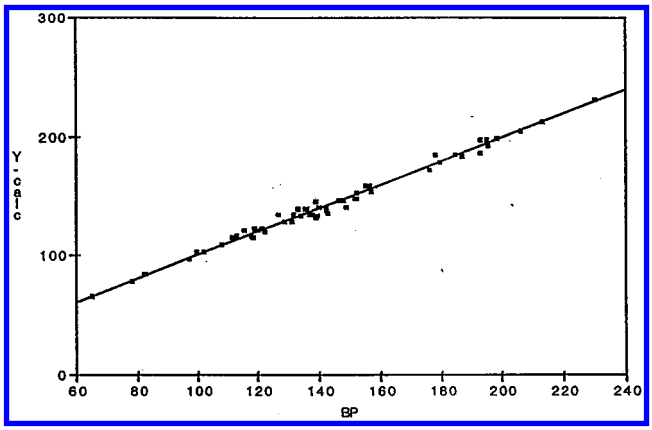

Figure 5. The regression of calculated against the experimental BP for alcohols of Table 5.

As we see from the above discussion, the introduction of p_3 improves the correlation somewhat but apparently not considerably. Hence, eq 7 or its equivalent, eq 10, can be taken as the best characterization of the correlation of the BPs of aliphatic alcohols with weighted path numbers. Figure 5 shows the regression of the calculated BP versus the experimental BP, and Table 8 lists the experimental and calculated BP.

CONCLUSION

The examples given clearly show the high-quality results based on optimal molecular descriptors which on one hand use fewer descriptors and on the other hand give correlation with significantly if not dramatically reduced standard error. When the multiple regression based on optimal descriptors is combined with the orthogonalization of the descriptors, one can expect results that will not only give satisfactory structure−property−activity relationship but will also lead to meaningful interpretation of the results—something that is currently missing in structure−property relationship studies. To further illuminate the structure−property−activity relationship, it seems appropriate not only to use orthogonalized molecular descriptors but to use whenever possible the same set of descriptors that serve as a basis for structure−property relationship that is valid for a wider pool of structures and a wider range of properties.

Table 8. The Experimental and Calculated Boiling Points

	descriptors p_1, p_2		
compound	BP exp.	BP calc.	residual
1 methanol	64.7	65.24	−0.54
2 ethanol	78.3	77.69	0.61
3 1-propanol	97.2	96.42	0.77
4 2-propanol	82.3	84.11	−1.81
5 1-butanol	117.7	115.67	2.03
6 2-butanol	99.6	102.43	−2.83
7 2-methyl-1-propanol	107.9	109.15	−1.25
8 2-methyl-2-propanol	82.4	84.52	−2.12
9 1-pentanol	137.8	134.92	2.88
10 2-pentanol	119.0	121.68	−2.68
11 3-pentanol	115.3	120.75	−5.45
12 2-methyl-1-butanol	128.7	127.97	0.73
13 3-methyl-1-butanol	131.2	128.90	2.30
14 2-methyl-2-butanol	102.0	102.41	−0.41
15 3-methyl-2-butanol	111.5	114.72	−3.22
16 2,2-dimethyl-1-propanol	113.1	115.84	−2.74
17 1-hexanol	157.0	154.17	2.83
18 2-hexanol	139.9	140.92	−1.02
19 3-hexanol	135.4	139.99	−4.59
20 2-methyl-1-pentanol	148.0	147.22	0.78
21 3-methyl-1-pentanol	152.4	147.72	4.68
22 4-methyl-1-pentanol	151.8	148.15	3.65
23 2-methyl-2-pentanol	121.4	121.66	−0.25
24 3-methyl-2-pentanol	134.2	133.55	0.65
25 4-methyl-2-pentanol	131.7	134.90	−3.20
26 2-methyl-3-pentanol	126.5	134.31	−7.81
27 3-methyl-3-pentanol	122.4	120.30	2.10
28 2-ethyl-1-butanol	146.5	146.79	−0.29
29 2,2-dimethyl-1-butanol	136.8	134.37	2.43
30 2,3-dimethyl-1-butanol	149.0	140.77	8.23
31 3,3-dimethyl-1-butanol	143.0	136.11	6.89
32 2,3-dimethyl-2-butanol	118.6	114.28	4.32
33 3,3-dimethyl-2-butanol	120.0	121.00	−1.00
34 1-heptanol	176.3	173.41	2.87
35 3-heptanol	156.8	159.24	−2.44
36 4-heptanol	155.0	159.24	−4.24
37 2-methyl-2-hexanol	142.5	140.90	1.60
38 3-methyl-3-hexanol	142.4	139.55	2.85
39 3-ethyl-3-pentanol	142.5	138.37	4.13
40 2,3-dimethyl-2-pentanol	139.7	133.11	6.59
41 3,3-dimethyl-2-pentanol	133.0	139.67	−6.57
42 2,2-dimethyl-3-pentanol	136.0	139.32	−3.32
43 2,3-dimethyl-3-pentanol	139.0	132.18	6.82
44 2,4-dimethyl-3-pentanol	138.8	145.34	−6.54
45 1-octanol	195.2	192.58	2.62
46 2-octanol	179.8	179.33	0.47
47 2-ethyl-1-hexanol	184.6	185.29	−0.69
48 2,2,3-trimethyl-3-pentanol	152.2	152.78	−0.57
49 1-nonanol	213.1	211.91	1.19
50 2-nonanol	198.5	198.66	−0.16
51 3-nonanol	194.7	197.73	−3.03
52 4-nonanol	193.0	197.73	−4.73
53 5-nonanol	195.1	197.73	−2.63
54 7-methyl-1-octanol	206.0	205.46	0.54
55 2,6-dimethyl-4-heptanol	178.0	185.69	−7.69
56 3,5-dimethyl-4-heptanol	187.0	183.83	3.17
57 3,5,5-trimethyl-1-hexanol	193.0	186.98	6.02
58 1-decanol	230.2	231.15	−0.95

ACKNOWLEDGMENT

We thank the Ministry of Science and Technology of Slovenia for partial support of this work through the grant J1-8901-0104-97.

REFERENCES AND NOTES

(1) The connectivity index $^1\chi$ was first introduced in: Randić, M. On the characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609−6615.
(2) The connectivity index $^2\chi$ and higher order connectivity indices $^m\chi$ were first introduced in: Kier, L. B.; Murray, W. J.; Randić, M.; Hall,

L. H. Molecular connectivity V: Connectivity series applied to density. *J. Pharm. Sci.* **1975**, *65*, 1226−1230.
(3) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press, New York, 1976.
(4) Randić, M.; Pompe, M. On characterization of CC double bond in alkenes, SAR and QSAR. In press.
(5) Lučić, B.; Trinajstić, N.; Sild, S.; Karelson, M.; Katritzky, A. R. A new efficient approach for variable selection based on multiregression: Prediction of gas chromatographic retention ties and response factors. *J. Chem. Inf. Comput. Sci*., submitted for publication.
(6) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. QSPR: The correlation and quantitative prediction of chemical and physical properties from structure. *Chem. Soc. Rev*. **1995**, *24*, 279−287.
(7) 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.
(8) Estrada, E. Graph theoretical invariant of Randić revisited. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1022−1025.
(9) Amić, D.; Bešlo, D.; Lučić, B.; Nikolić, S.; Trinajstić, N. The vertex-connectivity index revisited. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 819−822.
(10) Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc*. **1947**, *69*, 17−20.
(11) 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.
(12) Randić, M. Novel graph theoretical approach to heteroatom in quantitative structure-activity relationship. *Chemom. Intell. Lab. Syst*. **1991**, *12*, 970−980.
(13) Randić, M. On computation of optimal parameters for multivariate analysis of structure-property relationship. *J. Comput. Chem.* **1991**, *12*, 970−980.
(14) Randić, M.; Dobrowolski, J. Cz. Optimal molecular connectivity descriptors for nitrogen containing molecules. *Int. J. Quant. Chem.: Quant. Biol. Symp*. **1998**, *70*, 1209−1215.
(15) Grosman, S. C.; Jerman-Blažič Džonova, B.; Randić, M. A graph theoretical approach to quantitative structure-activity relationship. *Int. J. Quant. Chem.: Quant. Biol. Symp.* **1985**, *12*, 123−139.
(16) Balaban, A. T. Using real numbers as vertex invariants for third generation topological indices. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 23−28.
(17) Bonchev, D. *Information Theoretic Characterization of Chemical Structures*; Ellis Horwood: Chichester, England, 1983.
(18) Platt, J. R. Influence of neighbor bonds on additive bond properties in paraffins. *J. Chem. Phys*. **1947**, *15*, 419−420.
(19) Balaban, A. T. Chemical Graphs. XXXIV. Five new topological indices for the branching of tree-like graphs. *Theor. Chim. Acta* **1979**, *53*, 335−375.
(20) Schultz, H. P. Topological organic chemistry. 1. Graph theory and topological indices of alkanes. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 277−288.
(21) Balaban, A. T. Highly discriminating distance-based topological index. *Chem. Phys. Lett*. **1982**, *89*, 399−404.
(22) Randić, M. On molecular identification numbers. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 164−175.
(23) Lovasz, L.; Pelikan. On the eigenvalues of trees. *J. Period. Math. Hung*. **1973**, *3*, 17−1825.
(24) Randić, M. On molecular branching. *Acta Chim. Slov.* **1997**, *44*, 57−77.
(25) Szymanski, K.; Mülleer, W. R.; Knop, J. V.; Trinajstić, N. *Int. J. Quant. Chem.: Quant. Chem. Symp.* **1986**, *20*, 173.
(26) Bonchev, D.; Trinajstić; N. Information theory, distance matrix and molecular branching. *J. Chem. Phys*. **1977**, *67*, 4517−4533.
(27) Sarkar, R.; Roy, A. B.; Sarkar, P. K. *Math. Biosci*. **1978**, *39*, 299.
(28) Basak, S. C; Roy, A. B.; Ghosh, J. J. *Proceedings of the II International Conference on Mathematical Modelling*; University of Missouri-Rola: Missouri, 1979; Vol. 2, p 851.
(29) Raychaudhury, C.; Basak, S. C.; Ray, S. K.; Ghosh, J. J. Abstract. In *Proceedings of the 19th Annual Meeting*; *Society of Engineering Sciences Inc.*; University of Missouri-Rola: Missouri, 1982.
(30) Tratch, S. S.; Stankevich, M. V.; Zefirov, N. S. Combinatorial models and algorithms in chemistry. The expanded Wiener number - a novel topological index. *J. Comput. Chem.* **1990**, *11*, 899−908.
(31) Hall, L. H. Computational aspects of molecular connectivity and its role in structure−property modeling. In *Computational Graph Theory*; Rouvray, D. H., Ed.; Nova Publishers: New York, 1990; pp 202−233.
(32) Randić, M. Novel molecular descriptor for structure−property studies. *Chem. Phys. Lett*. **1993**, *211*, 478−483.
(33) Randić, M. Hosoya matrix - a source of novel molecular descriptors. *Croat. Chem. Acta* **1994**, *67*, 415−429.
(34) Amić, D.; Trinajstić, N. On the detour matrix. *Croat. Chem. Acta* **1995**, *68*, 53−62.
(35) Diudea, M. V.; Minailiuc, O.; Katona, G.; Gutman I. Szeged matrices and related numbers. *MATCH* **1997**, *35*, 119−143.
(36) Diudea, M. V. Cluj matrix, C_{ju}: Source of various graph descriptors. *MATCH* **1997**, *35*, 163−183.
(37) Randić, M.; Plavšić, D.; Razinger, M. Double invariants. *MATCH* **1997**, *35*, 243−259.
(38) Randić, M. Characterization of atoms, molecules, and classes of molecules based on path enumerations. *MATCH* **1979**, *7*, 3−60.
(39) Kier, L. B.; Hall, L. H. Molecular connectivity VII: Specific treatment of heteroatoms. *J. Pharm. Sci.* **1976**, *65*, 1806−1809.
(40) Kier, L. B. Indexes of molecular shape from chemical graphs. *Med. Res. Rev.* **1987**, *4*, 417−440.
(41) Hermann, A.; Zinn, P. List operations on chemical graphs. 6. Comparative study of combinatorial topological indexes of the Hosoya type. *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 551.
(42) Randić, M.; Morales D. A.; Araujo, O. Higher order Fibonacci numbers. *J. Math. Chem*. **1996**, *20*, 79−94.
(43) Randić, M. On characterization of the shape of molecular graphs. *J. Mol. Model*., In press.
(44) Randić, M. Novel shape descriptors for molecular graphs. *Chemom. Intell. Lab. Syst.*, submitted for publication.
(45) Bogdanov, B.; Nikolić, S.; Trinajstić, N. On the three dimensional Wiener number. *J. Math. Chem*. **1988**, *3*, 299−309.
(46) Randić, M. Molecular topographic descriptors. *Stud. Phys. Theor. Chem.* **1988**, *54*, 101−108.
(47) Randić, M. Molecular profiles: Novel geometry-dependent molecular descriptors. *New J. Chem*. **1994**, *19*, 781−791.
(48) Randić, M. Molecular shape profiles. *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 373−382.
(49) Randić, M.; Krilov, G. On characterization of molecular surfaces. *Int. J. Quant. Chem.* **1997**, *65*, 1065−1076.
(50) Diudea, M. V.; Horvath, D.; Graovac, A. Molecular topology. 15. 3D distance matrices and related topological indices. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 129−135.
(51) Estrada, E. Three-dimensional descriptors based on electron charge density weighted graphs. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 708−713.
(52) Tominaga, Y.; Fujiwara, I. Novel 3D descriptors using excluded volume: Application to 3D quantitative structure−activity relationships. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1158−1161.
(53) Nikolić, S.; Trinajstić, N.; Mihalic, Z. Molecular topological index. *J. Math. Chem*. **1993**, *12*, 251−264.
(54) Seybold, P. G.; May, M.; Bagel, U. A. Molecular structure-property relationships. *J. Chem. Educ.* **1987**, *64*, 575−581.
(55) Amidon, G. L.; Yalkovsky, S. H.; Leung, S. *J. Pharm. Sci.* **1974**, *63*, 1858.
(56) Smeeks, F. C.; Jurs, P. C. *Anal. Chim. Acta* **1990**, *233*, 111−119.
(57) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. Normal boiling points for organic compounds: Correlation and prediction by quantitative structure-property relationship. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 28−41.
(58) Magnuson, V. R.; Harriss, D. K.; Basak, S. C. Topological indices based on neighboring symmetry: Chemical and biological applications. In *Chemical Applications of Topology and Graph Theory*; King, R. B., Ed.; Studies in Physical and Theoretical Chemistry 28; Elsevier: The Netherlands, 1983; pp 178−191.
(59) Randić, M. Orthogonal molecular descriptors. *New J. Chem.* **1991**, *15*, 517−525.
(60) Randić, M. Resolution of ambiguities in structure-property studies by use of orthogonal descriptors. *J. Chem. Inf. Comput. Sci*. **1991**, *31*, 311−320.
(61) Randić, M. Fitting non-linear regressions by orthogonalized power series. *J. Comput. Chem*. **1993**, *14*, 363−370.
(62) Randić, M. Curve fitting paradox. *Int. J. Quant. Chem.: Quant. Biol. Symp.* **1994**, *21*, 215−225.
(63) Amić, D.; Davidović-Amić, D.; Jurić, A.; Lučić, B.; Trinajstić, N. Structure-activity correlation of flavone derivatives for inhibition of cAMP phosphodiesterase. *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 1034−1038.
(64) Lučić, B.; Nikolić, S.; Trinajstić, N.; Jurić, D. The structure-property models can be improved using the orthogonalized descriptors. *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 532−538.
(65) Šoškić, M.; Plavšić, D.; Trinajstić, N. Link between orthogonal and standard multiple linear regression models. *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 829−832.

CI9800763