On Use of the Variable Connectivity Index $^1\chi^f$ in QSAR: Toxicity of Aliphatic Ethers

Milan Randić*,† and Subhash C. Basak‡,§

NRRI, University of Minnesota at Duluth, Duluth, Minnesota 55811, Department of Mathematics and Computer Science, Drake University, Des Moines, Iowa 50311, Ames Laboratory − DOE, Iowa State University, Ames, Iowa 50011, and Laboratory for Chemometrics, National Institute of Chemistry, Hajdrihova 19, Ljubljana, Slovenia

Received August 12, 2000

The number of non-hydrogen atoms in a molecule, *N*, appears as a very good molecular descriptor for the toxicity of aliphatic ethers, despite the fact that it does not differentiate among isomers. The regression based on *N* as a descriptor for the toxicity in mice of 21 alkyl ethers was reported to yield the regression coefficient $r = 0.9751$. The simple connectivity index $^1\chi$ produced for the same data a less satisfactory regression: $r = 0.9548$. To see if variable connectivity index $^1\chi^f$ can improve the regression characterized by *N* we examined the same data using the variable connectivity index $^1\chi^f$ as a molecular descriptor. By varying *x*, *y*, the variables that discriminate between carbon and oxygen atoms, we obtained regression which approaches in quality the best reported regression using weighted paths as descriptor and which is marginally better than the regression based on *N*.

INTRODUCTION

In this article we will use the variable connectivity index $^1\chi^f$, a generalization of $^1\chi$, in predicting toxicity of aliphatic ethers in mice. This particular structure−activity relationship has been studied previously by Trinajstić, Knop, and coworkers,[1] who found that *N*, the number of non-hydrogen atoms, already gives a very good correlation, $r = 0.9751$, $s = 0.098$, and $F = 174$, where *r*, *s*, and *F* are the regression coefficient, the standard error, and the Fisher ratio, respectively. Then they used the weighted identification numbers (WID numbers) as a descriptor and were able to get slightly better regression: $r = 0.9761$, $s = 0.095$, and $F = 181$. The following question can be raised: Should one try to find a better correlation when a very good correlation already exists. We will try to answer this question.

TOXICITY OF ALIPHATIC ETHERS

The above-mentioned statistical parameters clearly show that the regression of the ether toxicity has been well characterized either by *N* or WID. This is reflected in the standard error and the Fisher ratio *F* which are reliable indicators of the quality of a regression. The high value of *r*, the coefficient of the regressions, is not necessarily a reliable indictor of the quality of a regression when the sample considered includes molecules of very different sizes. The range of the toxicity, measured as pC, from the compound having the smallest value, dimethyl ether (1.43), to ethyl *tert*-butyl ether (3.15) which has the largest value, gives the relative error about 5.8% and 5.6% for *N* and WID, respectively. When considering biological data the relative standard error of about 5−10% may appear satisfactory, but it is not a guarantee that the distribution of the residuals is at random, which is desirable for a good regression. Moreover, while there is no problem in interpreting *N*, the size of compounds, this parameter does not differentiate among isomers. On the other hand, WID parameters have no clear structural interpretation. Thus, despite the fact that both give very good regressions, there is room for improvement, particularly with respect to structural interpretation of such correlations.

Table 1. List of 21 Ethers, Their pC Values, the Corresponding WID Values, and the Connectivity Indices $^1\chi$

	ether	pC	WID	(0, 0)
1	dimethyl	1.43	3.29255	1.41421
2	methyl ethyl	1.74	4.12444	1.91421
3	diethyl	2.22	5.05815	2.41421
4	methyl isopropyl	2.26	5.07386	2.27006
5	methyl propyl	2.45	5.05815	2.42421
6	ethyl propyl	2.60	6.03157	2.91421
7	ethyl isopropyl	2.60	6.03723	2.77006
8	methyl butyl	2.70	6.03157	2.92421
9	methyl isobutyl	2.79	6.03985	2.77006
10	methyl *sec*-butyl	2.79	6.03723	2.80806
11	methyl *tert*-butyl	2.79	6.05011	2.56066
12	methyl pentyl	2.88	7.01917	3.41421
13	ethyl butyl	2.82	7.01917	3.41421
14	ethyl isobutyl	2.82	7.02315	3.27006
15	ethyl *sec*-butyl	2.85	7.02154	3.30806
16	ethyl *tert*-butyl	2.92	7.02712	3.06066
17	dipropyl	2.79	7.01917	3.41421
18	propyl isopropyl	2.82	7.02154	3.27006
19	diisopropyl	2.82	7.02449	3.12590
17	ethyl pentyl	3.00	8.01257	3.91421
20	ethyl *tert*-pentyl	3.15	8.01811	3.62132

In Table 1 we have listed 21 ethers and their pC values, the corresponding connectivity indices $^1\chi$, and the WID values, as reported by Trinajstić at al. In Figure 1 we show the regression of ether toxicity against *N*. A closer look at WID values shows that the criticism raised against *N*, that it does not discriminate among isomers, holds to a great extent

* Corresponding author phone: (515)292-7411; fax: (512)292-8629; e-mail: milan.randic@drake.edu.
† Drake University.
‡ University of Minnesota at Duluth.
§ E-mail: sbasak@NRRI.UMN.edu.

10.1021/ci000114u CCC: $20.00 © 2001 American Chemical Society
Published on Web 04/07/2001

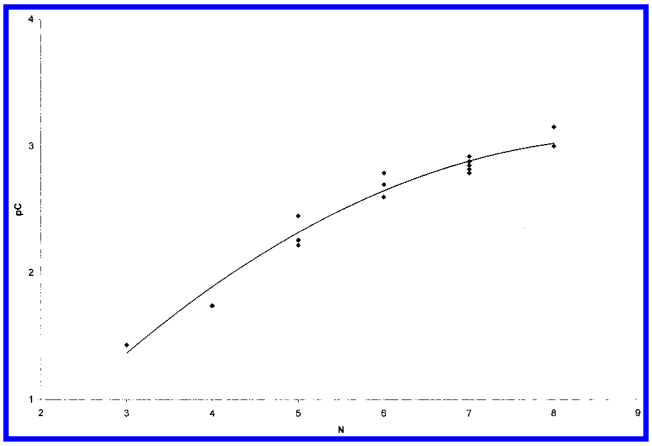

Figure 1. Correlation of ether toxicity (pC) against N.

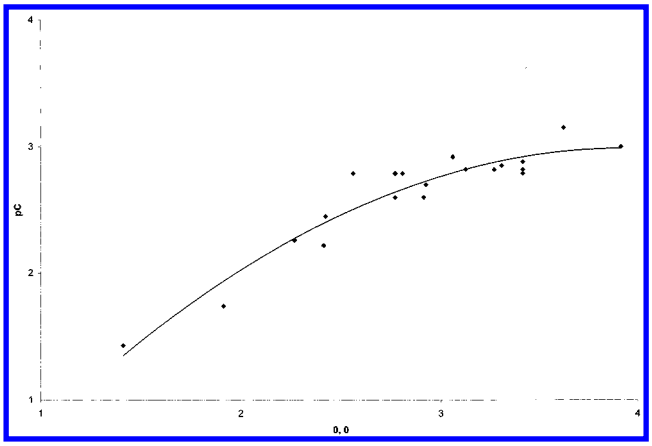

Figure 2. Correlation of ether toxicity (pC) against the connectivity index.

also for WID. Nevertheless, both descriptors are adequate because the experimental pC values (listed in Table 1) do not differ greatly among the isomers. For instance, we have

$$N = 5 \quad 2.22 > \text{pC} > 2.45$$

$$N = 6 \quad 2.60 > \text{pC} > 2.79$$

$$N = 7 \quad 3.00 > \text{pC} > 3.15$$

Observe also that the WID values, except for the two smallest compounds, are all within 1% of the value of *N*! Therefore, we can conclude that the toxicity of ether can alternatively be viewed as an atom additive property, although in the case of one-variable regression from the statistical point of view the distinction between an atom and bond additivity is blurred. The distinction, nevertheless, may be of interest for interpretation of the model.

VARIABLE CONNECTIVITY INDEX

Variable connectivity index was introduced about 10 years ago[2] as an alternative route for characterization of heteroatoms for structure−property-activity studies to Kier and Hall valence connectivity indices.[3] The potential of this index in QSAR and QSPR has yet to be evaluated. The present paper is a continuation of this effort that resulted in half a dozen recent studies.[4−9] We decided, therefore, to see how well the variable connectivity index performs to ether toxicity data of Table 1. We are motivated to explore the potential of variable connectivity index under different circumstances. The challenge is apparent: The connectivity index is a bond additive index, while the property considered could be interpreted as an atom-additive property. Can a bond-additive index with all its flexibility describe an atom-additive property?

In the last column of Table 1 we have listed the values of the connectivity index $^1\chi$. A glance at the values of the $^1\chi$ for alkyl ether isomers shows that connectivity indices vary over a much wider range than the values of *N*, WID, or pC. This is also clearly visible from Figure 2 in which we have plotted the correlation of pC against the connectivity index. At the first sight the regression shown in Figure 2 appears satisfactory. It is accompanied by the statistical parameters $r = 0.9548$, $s = 0.130$, and $F = 93$. Many QSAR regressions reported in the literature are not even close to the quality of correlation shown in Figure 2. Pretending that one was not aware of the results of Szymanski, Müller, Knop, and Trinajstić,[1] the results of Figure 2 would end further exploration of correlation of the toxicity of Table 1. However, having the flexibility of the variable connectivity index at one's disposal, we were interested to see how much the results of Figure 2 can be improved when variable parameter *x* associated with vertex degree of carbon atoms is allowed to vary.

Table 2. List of Values of the Variable Connectivity Index for a Selection of Values of *x*, Assuming $y = -1$

	(+1, −1)	(10, −1)	50, −1	1000, −1	10000, −1
1	1.41421	.60302	.28006	.06321395	.01999900
2	1.69271	.67723	.29812	.06419668	.02009849
3	1.97120	.75143	.31619	.06517941	.02019797
4	1.91421	.74611	.31586	.06517843	.02019796
5	2.02604	.76056	.31735	.06519468	.02019847
6	2.30453	.83476	.33542	.06617741	.02029795
7	2.19271	.82031	.33392	.06616116	.02029745
8	2.35937	.84389	.33658	.06619269	.02029845
9	2.28024	.82031	.33622	.06619169	.02029844
10	2.25758	.82959	.33509	.06619219	.02029794
11	2.10300	.81052	.33328	.06615922	.02029743
12	2.69271	.92723	.35581	.06719069	.02039843
13	2.63786	.91809	.35465	.06717542	.02039793
14	2.55873	.91170	.35429	.06717442	.02039792
15	2.53608	.90379	.35316	.06715917	.02039743
16	2.38150	.88472	.35134	.06714195	.02039691
17	2.63786	.91809	.35465	.06717542	.02039793
18	2.52603	.90365	.35315	.06715917	.02039743
19	2.41421	.88920	.35166	.06714292	.02039692
20	2.97120	1.00143	.37388	.06817342	.02049791
21	2.73171	.96833	.37058	.06813995	.02049689

Initial exploratory tests revealed that the optimal value for the *y* variable (that characterizes oxygen atom) is around −1. In Table 2 we list the values of the variable connectivity index for a selection of values of *x*, assuming $y = -1$. As we see the variable *x* assumes widely different values. It is significant that the variability in $^1\chi$ among isomers decreases as *x* increases. Thus for very large values of *x* all the isomers assume almost the same value of $^1\chi^f$. In Figure 3 we plotted the variable index when $x = 1000$ and $y = -1$ against the size of the compound (*N*) which resulted in a straight line, confirming that in this case $^1\chi^f$ has reduced to descriptor which is only size dependent. That means that all carbon−carbon bonds, or alternatively all carbon atoms, make equal contribution to the magnitude of $^1\chi^f$. In other words, the variable connectivity index in the limit apparently behaves similarly to *N* (or WID), hence, and can be viewed as an atomic descriptor. Gradual transformation of a bond-additive

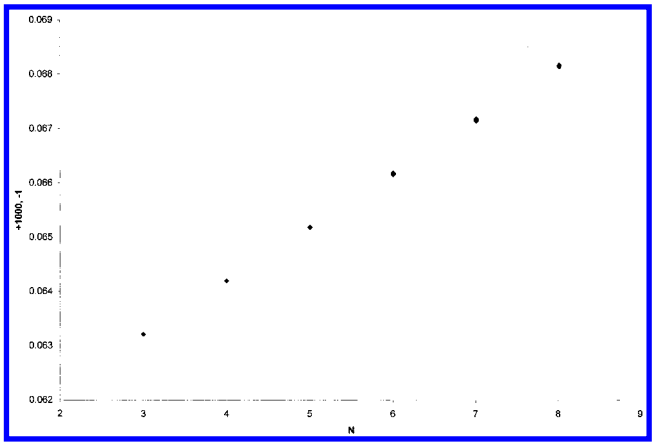

Figure 3. Correlation of the variable index ($x = 1000$, $y = -1$) against N.

behavior to atom-additive character is illustrated in Figure 4a−d, in which the correlation of pC was illustrated for the variable connectivity index for the values of $x = 10$, 100, 1000, and 10 000, assuming always $y = -1$. As we see from Figure 4a−d as x increases we can see the clustering of regression points around different N values. When $x = 1000$, the patterns of the clusters form vertically arranged points indicating that each group of ether isomers has almost the same descriptor value, which coincides with N.

DISCUSSION

In Table 3 we have summarized the statistic associated with the variation of x. We included also the statistical parameters when N and WID were used as descriptors. Observe that the Fisher ratio F has almost doubled from the initial value associated with ${}^1\chi$ (i.e., $x = 0$) and the optimal ${}^1\chi^f$ (when $x = 1000$). The similarity between Figure 1 and Figure 4d only confirms the already stated conclusion that, despite the different definition of each of the indices, they do not discriminate between the primary, the secondary, and the tertiary carbon atoms. In addition indices N and WID also do not differentiate carbon and oxygen which are differentiated by the variable connectivity index.

The discrimination among carbon and oxygen atoms may appear irrelevant, because it is the number of non-hydrogen atoms that play the dominant role in the regression, regardless of whether the atoms are carbon atoms or oxygen atom. However, if we analyze contributing terms to the variable connectivity index we can clearly see that not all atoms are equally important in this particular structure−activity correlation. In Table 4 we illustrate this on ethyl isopropyl ether, one of the compounds included in the correlation. The variable connectivity index for ethyl isopropyl ether is given by

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

Here the first two terms belong to CH_3-CH_2 and the two CH_3-CH bonds, respectively, while the last two terms belong to the oxygen bridge in ether, the bonds CH_2-O and $CH-O$, respectively. As x increases the contributions of all bonds decrease, but as we see from Table 4 this occurs much faster for CC bonds than for CO bonds. In the last column of the table we give the quotient of the contributions of the

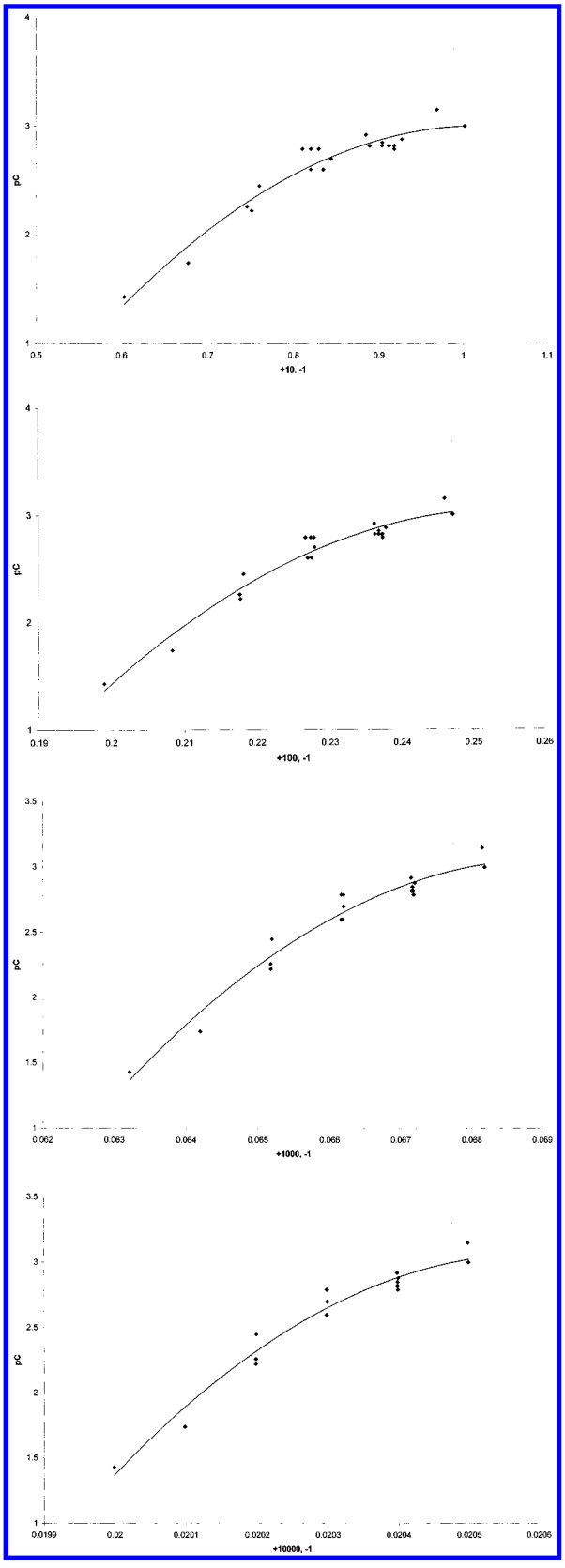

Figure 4. Correlation of ether toxicity (pC) illustrated for variable connectivity index for the values of $x = 10$, 100, 1000, and 10,000, assuming always $y = -1$.

two CO bonds relative to the three CC bonds, indicated as $\Sigma CO/\Sigma CC$. As we see from this column in the case $x = 0$ (the simple connectivity index), the CO bonds make a smaller contribution to ${}^1\chi$ than CC bonds. As x increases the role of

Table 3. Summary of the Statistic Associated with the Variation of x^a

(x, y)	r	s	F
0, 0	0.9548	0.1297	92.8
0, −1	0.9576	0.1256	99.5
1, −1	0.95996	0.1222	105.7
2, −1	0. 96303	0.1175	115.0
5, −1	0. 96556	0.1135	124.0
10, −1	0. 96938	0.1071	140.2
20, −1	0. 97394	0.0989	166.0
50, −1	0. 97522	0.0965	174.8
100, −1	0. 97531	0.0963	175.5
1000, −1	0. 97560	0.0958	177.7
2000, −1	0. 97560	0.0958	177.7
5000, −1	0. 97535	0.0962	175.8
10000, −1	0. 97522	0.0965	174.9
N	0.9751	0.0967	174.0
WID	0.9761	0.0949	181.2

[a] r = the coefficient of regression; s = the standard error; F = the Fisher ratio.

Table 4. Magnitudes of the Contributing Bond Terms to the Variable Connectivity Index for CC and CO Bonds as x Increases, Illustrated on Ethyl Isopropyl Ether

x	CH_3-CH_2	CH_3-CH	CH_2-O	$CH-O$	$\Sigma CO/\Sigma CC$
0	0.70711	0.57735	0.70711	0.57735	0.69
1	0.40824	0.35355	0.57735	0.50000	1.41
5	0.15430	0.14434	0.37796	0.35355	2.45
10	0.087039	0.083624	0.28868	0.27735	3.32
50	0.019418	0.019234	0.13868	0.20851	8.98
100	0.0098523	0.0098044	0.099015	0.098533	10.05
500	0.0019940	0.0019920	0.044632	0.044588	22.38
1000	0.00099850	0.00099800	0.031591	0.031575	31.64
5000	0.00019994	0.00019992	0.014139	0.014138	70.72
10000	0.000099985	0.000099980	0.0099990	0.0099985	100.01

CO bonds increases too. For the value of $x = 1000$, the contribution of CO bonds is more than 30 times that of CC bonds, increasing further to a factor of 100 when $x = 10\,000$, a value close to the limiting value of $^1\chi^f$ approaching N.

As we see the relative roles of CC and CO bonds as x varied from $x = 0$ to $x = 1000$ have *reversed.* Thus a "small" improvement (from the statistical point of view) in characterization of the molecules may result in dramatic change in interpretation of the results. This clearly points to a need to search for better molecular descriptors that will lead to better regressions, regardless of how good the regression is.

There is another important aspect of Table 4 that we should point out. If we compare the relative role of different CC bonds giving contributions to $^1\chi^f$ among themselves, shown in the first two columns of Table 4, we see that as x increases the distinction between nonequivalent CC bonds diminishes. The same is true of the two CO bonds, which make visibly different contributions to $^1\chi^f$ only for smaller values of x. But when x reaches the optimal value, the difference between nonequivalent bonds is well below the value that could influence the regression. Of course, as x increases further, these differences become less and less and eventually will disappear in the limit. The significance of this "equalization" of bond contributions is that all the bonds of the same class in a molecule (here CC bonds form one class and CO bonds form the other class) make an equal contribution to the calculated index. Hence, in acyclic systems, we can alternatively assign contributions giving rise to $^1\chi^f$ to individual atoms, rather than individual CC and CO bonds. The data in Table 4 confirm what we have expected from the start that pC as a property is rather atom additive than bond additive.

Table 5. Comparison of Computed PC Values Based on WID, N, and the Present Work[a]

	WID	N	(1000,−1)	res
1	1.38789	1.36204	1.36703	0.06297
2	1.86706	1.91720	1.89004	−0.15004
3	2.30621	2.35473	2.31665	−0.09665
4	2.31270	2.26343	2.31627	−0.05627
5	2.30621	2.43063	2.32252	0.12748
6	2.65293	2.72615	2.65125	−0.05125
7	2.65461	2.61068	2.64660	−0.04660
8	2.65293	2.77264	2.65560	0.04440
9	2.65539	2.69163	2.65532	0.13468
10	2.65461	2.69163	2.65546	0.13454
11	2.65843	2.50150	2.64604	0.14396
12	2.88878	2.94323	2.88927	−0.00927
13	2.88878	2.92614	2.88644	−0.06644
14	2.88950	2.88607	2.88625	−0.06625
15	2.88921	2.88607	2.88341	−0.03341
16	2.89021	2.77264	2.88016	0.03984
17	3.00825	2.95472	3.02222	−0.02222
18	3.00858	2.95226	3.01928	0.13072
19	2.88878	2.92614	2.88644	−0.09644
20	2.88921	2.87313	2.88341	−0.06341
21	2.88974	2.79736	2.88035	−0.06035

[a] The last column gives the residuals for here computed PC values.

This may be understood in terms of physical models. What it means is that for the toxicity considered here, the size of the molecule plays the dominant role, and the shape of a molecule is of less concern here. This conclusion is in line with the critical volume hypothesis of Mullins[10] who speculated that the biological activity is determined by the extent of expansion of the membrane caused by the intruding molecule. This conclusion is more specific than what would correlations with N or WID suggest, that the dominant factor for the toxicity of ether is the size of molecules. This apparently is only partially true because it is the oxygen atom that plays the dominant role here. However, the presence of other atoms show enough influence to account for the size effect.

In Table 5 we have summarized the computed pC values based on the present work and compared them to the computed pC values reported for N and WID. The last column in this table gives the individual values of the optimal variable connectivity index for the 21 ethers considered.

CONCLUDING REMARKS

Already from the correlation of pC with N it is apparent that molecular size plays the dominant role for the toxicity of ethers. This is understandable within the Meyer-Overton theory of narcosis[11−13] where it is assumed that critical volume plays in expansion of the cell membrane. However, neither N nor molecular weight MW as an alternative parameter is fully equivalent to molecular volume. In Figure 5 we show a plot of pC against the van der Waals volumes of ether considered. van der Waals volumes were calculated by an in-house computer program.[15] The regression based on van der Waals volumes is not of as high quality as was the regression based on N (Figure 1), which indicates that perhaps it is not merely molecular volume that here plays

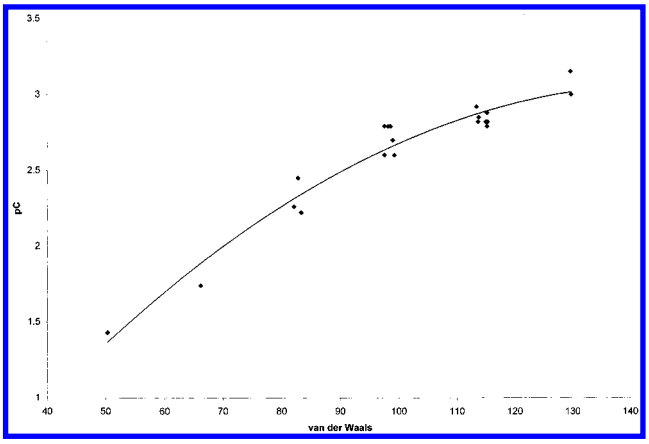

Figure 5. Plot of pC against the van der Waals volumes for ether considered.

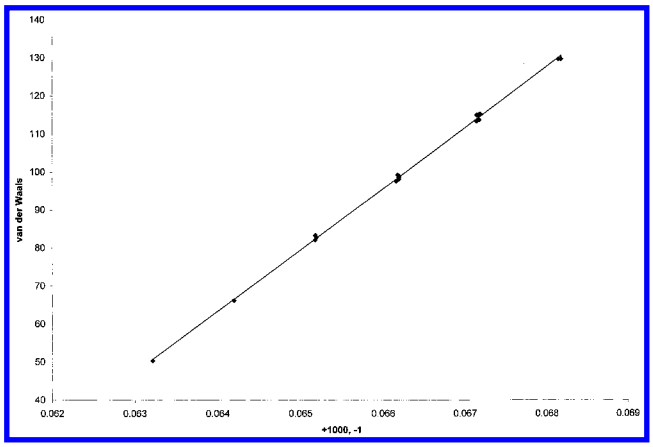

Figure 6. Plot of van der Waals volumes against the variable index ($x = 1000$, $y = -1$) for ether considered.

the dominant role but another aspect of molecular size yet to be better understood, that is described by N, WID, and ${}^1\chi^f$ ($x = 1000$; $y = -1$).

The versatility of the variable connectivity index can again be seen from a correlation between the calculated van der Waals volumes of ethers and ${}^1\chi^f$ (shown in Figure 6). As we see from Figure 6 ${}^1\chi^f$ gives very satisfactory regression for van der Waals volumes, and previously we have seen that it gave very satisfactory regression for pC. Yet the two quantities, pC and van der Waals radii, do not show mutual regression of a similar high quality.

ACKNOWLEDGMENT

We would like to thank Brian Gute for technical assistance in computation of van der Waals radii. This is contribution number 292 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.

REFERENCES AND NOTES

(1) Szymanski, V.; Müller, W. R.; Knop, J. V.; Trinajstić, N. *Int. J. Quantum Chem.* **1984**, *S20*, 173−183.
(2) Randić, M. Novel graph theoretical approach to heteroatoms in QSAR. *Chemometrics Intel. Lab. Syst.* **1991**, *10*, 213.
(3) Randić, M. On computation of optimal parameters for multivariate analysis of structure−property relationship. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 970−980.
(4) Kier, L. B.; Hall, H. L. Molecular connectivity VII. Specific treatment of heteroatoms. *J. Pharm. Sci.* **1976**, *65*, 1806−1809.
(5) Randić, M.; Basak, S. C. Variable Molecular Descriptors. In *Some Aspects of Mathematical Chemistry*; Sinha, D. K., Basak, S. C., Mohanty, R. K., Busamallick, I. N., Eds.; Visva-Bharati University Press: Santiniketan, India (in press).
(6) Randić, M.; Pompe, M. The variable connectivity index ${}^1\chi^f$ versus the traditional molecular descriptors: A comparative study of ${}^1\chi^f$ against descriptors of CODESSA. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 631−638.
(7) Randić, M.; Dobrowolski, J. Cz. Optimal molecular connectivity descriptors for nitrogen-containing molecules. *Int. J. Quantum Chem.* **1998**, *70*, 1209−1215.
(8) Randić, M. High quality structure−property regressions. Boiling point of smaller alkanes. *New J. Chem.* **2000**, *40*, 165−171.
(9) Randić, M.; Basak, S. C. On construction of high quality structure-property-activity regressions: The boiling points of Sulfides. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 899−905.
(10) 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.
(11) Mullins, L. J. Some physical mechanism in narcosis. *Chem. Rev.* **1954**, *54*, 289−323.
(12) Roth, S. H. Physical mechanisms of anesthesia. *Annu. Rev. Pharmacol. Toxicol.* **1979**, *19*, 159−178.
(13) Trudell, J. R. Biophysical concepts in molecular mechanisms of anesthesia. In *Molecular Mechanisms of Anesthesia*; Fink, B. R., Ed.; Raven: New York 1980.
(14) LaBella, F. S. The general anesthesia receptor. *Perspectives Biol. Med.* **1982**, *25*, 322−331.
(15) Grunewald, G. D. Natural Resources Research Institute, University of Minnesota at Duluth, Duluth, MN.

CI000114U