

Atom-Type-Based AI Topological Descriptors: Application in Structure-Boiling Point Correlations of Oxo Organic Compounds

Biye Ren*

Research Institute of Materials Science, South China University of Technology,
Guangzhou 510640, P. R. China

Received December 10, 2002

Structure-boiling point relationships are studied for a series of oxo organic compounds by means of multiple linear regression (MLR) analysis. Excellent MLR models based on the recently introduced Xu index and the atom-type-based AI indices are obtained for the two subsets containing respectively 77 ethers and 107 carbonyl compounds and a combined set of 184 oxo compounds. The best models are tested using the leave-one-out cross-validation and an external test set, respectively. The MLR model produces a correlation coefficient of $r = 0.9977$ and a standard error of $s = 3.99$ °C for the training set of 184 compounds, and $r_{cv} = 0.9974$ and $s_{cv} = 4.16$ °C for the cross-validation set, and $r_{pred} = 0.9949$ and $s_{pred} = 4.38$ °C for the prediction set of 21 compounds. For the two subsets containing respectively 77 ethers and 107 carbonyl compounds, the quality of the models is further improved. The standard errors are reduced to 3.30 and 3.02 °C, respectively. Furthermore, the results obtained from this study indicate that the boiling points of the studied oxo compound dominantly depend on molecular size and also depend on individual atom types, especially oxygen heteroatoms in molecules due to strong polar interactions between molecules. These excellent structure-boiling point models not only provide profound insights into the role of structural features in a molecule but also illustrate the usefulness of these indices in QSPR/QSAR modeling of complex compounds.

1. INTRODUCTION

The construction of quantitative structure–property/activity relationship (QSPR/QSAR) models has been received considerable attentions in the past decade years.¹ Numerous QSPR/QSAR models have reported based on various known physicochemical parameters and molecular descriptors such as geometric, electronic or electrostatic, polar, steric, topological indices.² Among the most-used descriptors are topological indices because they can be easily and directly calculated from the molecular structures without any experimental effort.^{3–6} Hence, use of topological indices in QSPR/QSAR provides simple and straightforward means for the molecular design and the estimation of properties and activities of a molecule.

In general, traditional 2D and 3D topological indices characterize a molecule as a whole, i.e., molecular size or shape, such as the well-known molecular connectivity index (χ),^{5,6} Balaban's index (J),⁷ Wiener's index (W),⁸ and the recently introduced Xu index.⁹ However, these conventional indices are being faced with some challenges because the QSPR/QSAR models based on a single or few conventional topological indices cannot directly reflect the role of individual atom types or groups in molecules and their relative importance to the property or activity of a molecule.

In recent years, attention has been focused on the use of the atom-type-based topological indices.¹⁰ Unlike the conventional topological indices, the atomic level topological indices code the structural environment of each atom type

in a molecule and further describe the structural information of a molecule at the atomic level, especially since they offer the possibility of understanding the role of individual atom types or groups in a molecule such as $-\text{OH}$, $-\text{COOH}$, $-\text{NH}_2$, $>\text{CO}$, etc. One of the most interesting indices of this type is the electrotopological state (E-state) index introduced by Kier and Hall.¹¹ The E-state indices have been successfully used in a variety of QSPR/QSAR studies.^{10–14} Very recently, a new type of atom-type-based topological index, AI, was proposed in our laboratory.¹⁵ The AI indices have been shown to be very effective in modeling six physical properties of alkanes,¹⁶ physical properties and biological activities of alcohols,¹⁷ three physical properties of a mixed set of alkanes and alcohols,¹⁸ gas chromatographic retention index of a combined set of aldehydes and ketones,¹⁹ and the normal boiling points (BP) of a mixed set of sulfides and thiols.²⁰ However, the potential of these AI indices in QSPR/QSAR modeling has yet to be further verified in different systems.

The boiling point of a compound is an important physical property for the design and optimization of the processing engineering in chemical and petrochemical industries. Therefore, it is required to develop quantitative structure-boiling point models for estimation of boiling points of untested compounds or selection of reliable experimental boiling points of compounds.²¹ Recently, Toropov and Toropova²² studied the structure-boiling point relationships for a data set of 200 carbonyl compounds (mono- and dialdehydes, mono- and diketones, keto aldehydes, and esters of mono-carboxylic acids). The best one-variable model by means of optimization of correlation weights of local graph invariants

* Corresponding author phone/fax: +86-20-8711-2886; e-mail: renbiye@163.net.

based on nearest neighboring codes was obtained and produced a standard error of $s = 6.12$ °C for the training set and $s = 6.0$ °C for the test set.

In the present study, to further illustrate the usefulness of the AI indices in QSPR studies, a mixed set of 184 oxygen-containing compounds, including 77 acyclic aliphatic ethers, 33 aldehydes, and 77 ketones with a wide range of non-hydrogen atoms, are chosen to construct the structure-boiling point models using the Xu and AI indices. In addition, we wish to see what structural features or groups are important to the boiling points of oxo compounds using these indices.

2. METHOD

For a molecular graph $G = \{V, E\}$ with n vertices, where $V(G)$ and $E(G)$ are the vertex and edge sets, respectively, vertices correspond to individual atoms in a graph G , and edges correspond to chemical bonds between them. The vertex-adjacency matrix, $\mathbf{A} = [a_{ij}]_{n \times n}$, is a square symmetric matrix. The elements a_{ij} of matrix \mathbf{A} are 1 if vertices i and j are adjacent and 0 otherwise, where n is the number of vertices. The distance matrix, $\mathbf{D} = [d_{ij}]_{n \times n}$, is also a square symmetric matrix. The entries d_{ij} of matrix \mathbf{D} are the length of the shortest path between the vertices i and j in a graph G . The sum over row i or column j of matrix \mathbf{A} yields local vertex-degree v_i ; analogously, the sum over row i or column j of matrix \mathbf{D} yields distance sums s_i . The Xu index can be expressed below⁹

$$\text{Xu} = n^{1/2} \log \left(\sum_{i=1}^n v_i s_i^2 \middle| \sum_{i=1}^n v_i s_i \right) \quad (1)$$

where the sum is over all i vertices in a molecular graph.

For any atom i that belongs to j th atom type in a graph G , the corresponding topological index value is expressed as¹⁵

$$\text{AI}_i(j) = 1 + \phi_i(j) \quad (2)$$

with

$$\phi_i(j) = v_i(j) s_i^2(j) \middle| \sum_{i=1}^n v_i s_i \quad (3)$$

where the parameter ϕ is considered as a perturbing term reflecting the effects of the structural environment of i th atom on its topological index value, and the sum is over all i vertices in a graph G .

According to the above definition, for j th atom-type in a graph G , the corresponding atom type topological index, $\text{AI}(j)$, is the sum of all $\text{AI}_i(j)$ values of the same atom type

$$\begin{aligned} \text{AI}(j) &= \sum_{i=1}^l \text{AI}_i(j) = l + \sum_{i=1}^l \phi_i(j) = \\ &= l + \sum_{i=1}^l v_i(j) s_i^2(j) \middle| \sum_{i=1}^n v_i s_i \quad (4) \end{aligned}$$

where l is the count of the atoms or groups of the same type. Clearly, the AI index depends on both the account of the same atom-types and their microenvironment in molecule.

To modify these indices for complex compounds, a new type of vertex degree v^m was proposed to differentiate

heteroatoms and carbon atoms with multiple bonds in a graph G^{17}

$$v^m = \delta + k \quad (5)$$

$$k = 1/[(2/N)^2 \delta^v + 1] \quad (6)$$

where δ is the number of connections (edges) of that atom; parameter k is a perturbing term reflecting the effects of heteroatoms and multiple bonds; N is the principal quantum number of the valence shell; and δ^v is the valence connectivity of Kier-Hall.^{5,6} For heteroatoms in a graph G , the δ^v value is expressed as

$$\delta^v = (Z^v - h)/(Z - Z^v - 1) \quad (7)$$

The δ^v value of carbon atoms with multiple bonds in a graph G is expressed as

$$\delta^v = Z^v - h \quad (8)$$

where h is the number of hydrogen atoms connected to the heteroatom. Z and Z^v are the atomic number and the number of valence electrons for heteroatoms, respectively.

Consequently, for any molecular graph, both the Xu and AI indices can be expressed with the same formula defined above (eqs 1 and 4) by using proposed vertex degree v^m instead of the local vertex degree v_i . This approach has been shown to be very effective in modifying these indices for structure–property/activity correlations of organic compounds.^{15–20}

3. DATA SET AND REGRESSION ANALYSIS

Data Set. The data set consists of 205 oxo compounds including 86 ethers, 33 aldehydes, and 86 ketones with up to 17 non-hydrogen atoms (Table 1). Most of the experimental BP values for 86 ethers are taken from refs 23–27, and some are from ref 28 and Beilstein's *Handbuch der Organischen Chemie*. The majority of experimental BP data for aldehydes and ketones are collected from refs 23–27, and a few compounds are from ref 29. In the majority of the cases, the data among these sources agree fairly well, but there are some discrepancies between different sources. For compounds with a boiling point range of 1–4 °C we quote their arithmetic mean BP . However, the main problem with exploiting QSPR/QSAR models is the need for reliable experimental BP values. Hence, when a compound has different experimental boiling points reported in the literature, we use chemical intuition to make a choice between the disagreed values.

The data set is analyzed from two different approaches to evaluate the applicability of these indices in complex systems. First, the whole data set is divided into two subsets: one contains only 86 ethers and the other 119 carbonyl compounds. Approximately 10% of compounds are randomly selected from the original subset as a test set, while the other 90% remained as the training set, and then individual correlations for the two training subsets and the whole training set are carried out.

Regression Analysis. Multiple linear regression (MLR) method using the modified Xu (represented as X_u^m) and all AI indices present in molecules as variables is performed to

generate linear structure-boiling point models. The final models have the following general form (eq 9)

$$BP = a_0 + a_1 X_u^m + b_1 AI(1) + \dots + b_j AI(j) \quad (9)$$

where a_0 is a constant, a_1 is the contribution coefficient of the X_u^m index, and b_j is the contribution coefficient of j th atom type (group) AI index, $AI(j)$. Each coefficient describes the sensitivity of BP to each of the individual indices, so the coefficients of these parameters would measure the relative importance of each index. The significance of each variable is evaluated by monitoring the t -values and Fisher ratios (F) so as to choose the powerful descriptors.^{30–32} The quality of the model is considered as statistically satisfactory on the basis of the correlation coefficient (r), standard deviation (s), and F values when all the variables in the final model are significant at the 95% confidence level.

Model Validity. Two strategies are used to test the validity of the selected MLR model. First, the model for the training set is tested by the cross-validation using the leave-one-out method. In this way, each time one molecule is removed from the training set a model is developed with the remaining $N-1$ molecules. At each step, the BP value of the removed molecule is predicted by the model obtained. This process is repeated until each molecule is predicted once. In principle, the predictive ability of the model is quantified in terms of the corresponding leave-one-out cross-validated parameters, r_{cv} and s_{cv} values, which are defined as³³

$$r_{cv}^2 = 1.0 - PRESS/SSY \quad (10)$$

$$s_{cv} = \sqrt{PRESS/(N - M - 1)} \quad (11)$$

where $PRESS$ is the predicted residual sum of square and SSY is the sum of the squares of the experimental values. N is the number of samples used for model building. M is the number of descriptors involved in the model.

In addition, the actual prediction ability of the model is validated using an external prediction set not included in the training set. The performance of the model (its predictive ability) can be given by the predictive correlation coefficient (r_{pred}) (or the $PRESS/SSY$ ratio) and the standard error (s_{pred}) of prediction, which is defined as³⁴

$$s_{pred} = \sqrt{PRESS/N} \quad (12)$$

For a reliable model, the $PRESS/SSY$ ratio should be < 0.6 or the r_{cv}^2 or r_{pred}^2 values should be > 0.6 . The model is considered to be excellent if the $PRESS/SSY$ ratio is < 0.1 .

4. RESULTS AND DISCUSSION

In general, there are two different directions in which the MLR analysis is usually used in QSPR/QSAR studies.^{35,36} The first approach is based on a large number of structurally diverse compounds; the other way only deals with a smaller group of structurally related compounds. Both approaches have their merits and faults, and they serve somewhat different purposes. However, the best structure–property/activity correlations can be constructed for a rather narrow series of homologous compounds. In addition, it may be possible to manifest the role of all possible atom types or

groups in molecules and their relative importance to the properties or activities studied from the QSPR/QSAR models obtained for a small homologous set of compounds.³⁷

In this study, first the separate correlations for the two data subsets are analyzed; second, the study is further extended to the whole data set of oxygen-containing compounds.

Ether Subset. The subset of 86 ethers is randomly divided into a training set and a prediction set. The 77 ethers as the training set are used to generate the boiling point models in terms of the X_u^m index and all five AI indices present in molecules. Two suitable models are obtained (Table 2). As an illustration, the best five-parameter model (eq 13), along with the statistical parameters and the leave-one-out cross-validated statistics, is shown as follows:

$$\begin{aligned} BP = & -66.4027 (\pm 2.4346) + 49.2263 (\pm 1.2710) \\ & X_u^m + 3.9722 (\pm 0.6158) AI(-O-) - \\ & 0.6742 (\pm 0.09244) AI(>CH_2) - \\ & 2.1384 (\pm 0.2326) AI(>CH-) - \\ & 2.4418 (\pm 0.3507) AI(>C<) \quad (13) \end{aligned}$$

$$\begin{aligned} r = 0.9983; r_{cv} = 0.9981; s = 3.30; s_{cv} = 3.51; \\ F = 4260; P < 0.0001; N = 77 \end{aligned}$$

The t -values are -27.27 , 38.73 , 6.45 , -7.29 , -9.19 , and -6.96 , respectively. All variables are statistically significant at the level of $p < 0.0001$. Cross-correlation analysis shows that the variables in the final model are not highly correlated with each other. This model explains more than 99% of the variance in the experimental BP values for 77 aliphatic ethers. The statistics for the training set are $r = 0.9983$ ($r^2 = 0.9971$) and $s = 3.30$ °C, and for the cross-validation set $r_{cv} = 0.9981$ ($r_{cv}^2 = 0.9962$) and $s_{cv} = 3.51$ °C, respectively. The fact that the cross-validated r_{cv} and s_{cv} values are very close to the statistics of the model for the training set demonstrates the final model to be statistically significant. As a second strategy, the best model (eq 13) is used to predict the BP values of nine ethers as the test set. The predictive r_{pred} is 0.9930 ($r_{pred}^2 = 0.9861$) and s_{pred} is 4.89 °C for the test set, indicating a good predictive power of the model. There is only one outlier (no. 36, 2,2-dimethylpropyl ethyl ether) with a slightly larger predictive error of -11.9 °C. The reason for this is not clear, because it is obviously a very difficult work to identify data source publications and assess the reliability of data. In a previous paper,²⁸ 2,2-dimethylpropyl ethyl ether was also observed to show a large residual of -10.1 °C in a structure-boiling point correlation. However, if we discard this compound from the test set, for the remaining eight ethers the predictive r_{pred} and s_{pred} values are changed to 0.9980 and 2.65 °C, respectively.

The calculated BP values from the best model and residuals for 77 compounds are shown in Table 1, and the predicted BP values for the nine compounds are also given. A comparison of calculated or predicted BP values and experimental data for 86 compounds is shown in Figure 1. One can observe that calculated or predicted BP values agree well with experimental data, and the plot shows no obviously observable pattern. Therefore, the final model (eq 13) represents an excellent QSPR model judging from the statistics and the plot in Figure 1. This is in accordance with

Table 1. Experimental and Calculated or Predicted *BP* Values from the Subset and the Whole Data Set Correlations

									BP (°C)				
no.	compounds	X _u ^m	Al(O) ^b	Al(C=) ^c	Al(−CH ₃)	Al(>CH ₂)	Al(>CH−)	Al(>C<)	exp	calcd ^d	res ^d	calcd ^e	res ^e
Ethers													
1	dimethyl ether	0.7139	1.8334	0	3.7500	0	0	0	−24.9	−24.0	−0.9	−14.9	−10.0
2	ethyl methyl ether	1.3697	2.2001	0	4.5199	2.1200	0	0	7.4	8.3	−0.9	15.1	−7.7
3	methyl propyl ether	1.9931	2.7214	0	5.2786	4.7868	0	0	38.6	39.3	−0.7	44.0	−5.4
4	diethyl ether	1.9916	2.2677	0	5.2863	5.2206	0	0	34.6	37.1	−2.5	42.0	−7.4
5	isopropyl methyl ether	1.8662	2.4595	0	6.9540	0	2.4189	0	30.8	30.1	0.7	35.0	−4.2
6	butyl methyl ether	2.5917	3.3241	0	6.0332	8.0729	0	0	70.3	68.9	1.4	72.0	−1.7
7	ethyl propyl ether	2.5896	2.5598	0	6.0436	8.8048	0	0	63.9	65.3	−1.4	68.6	−4.7
8	ethyl isopropyl ether	2.4725	2.3834	0	7.8818	3.0173	2.9366	0	54.1	56.5	−2.4	59.8	−5.7
9	isobutyl methyl ether	2.4749	3.1553	0	7.8677	2.2873	2.9310	0	58.6	60.2	−1.6	63.3	−4.7
10	sec-butyl methyl ether ^a	2.4368	2.8217	0	7.8170	2.7001	2.5427	0	59.0	57.5	1.5	60.8	−1.8
11	tert-butyl methyl ether	2.3040	2.6496	0	9.3401	0	0	2.7319	52.0	50.9	1.1	51.8	0.2
12	methyl pentyl ether	3.1693	3.9769	0	6.7860	12.0092	0	0	99.0	97.3	1.7	98.9	0.1
13	ethyl butyl ether	3.1671	2.9698	0	6.7972	12.9743	0	0	92.2	92.6	−0.4	94.4	−2.2
14	dipropyl ether	3.1665	2.6797	0	6.8009	13.2534	0	0	89.6	91.2	−1.6	93.2	−3.6
15	isopropyl propyl ether	3.0620	2.5474	0	8.8363	6.4041	3.5780	0	83.0	82.5	0.5	84.5	−1.5
16	ethyl isobutyl ether	3.0626	2.8400	0	8.8313	6.1260	3.5758	0	82.0	83.9	−1.9	85.8	−3.8
17	isopentyl methyl ether	3.0652	3.8676	0	8.8164	5.1520	3.5692	0	91.0	88.7	2.3	90.4	0.6
18	methyl 1-methylbutyl ether	3.0159	3.2655	0	8.6925	5.9660	2.8766	0	91.5	84.8	6.7	86.6	4.9
19	ethyl sec-butyl ether	3.0144	2.6250	0	8.7027	6.5747	2.8800	0	81.2	81.8	−0.6	83.8	−2.6
20	methyl 2-methylbutyl ether ^a	3.0169	3.6251	0	8.6874	5.6250	2.8750	0	93.0	86.6	6.4	88.2	4.8
21	diisopropyl ether	2.9447	2.4152	0	10.7622	0	6.7943	0	69.0	73.6	−4.6	75.7	−6.7
22	methyl tert-pentyl ether	2.8413	2.9126	0	10.2697	2.7850	0	2.8883	86.3	76.1	10.2	75.4	10.9
23	1,2-dimethylpropyl methyl ether	2.8958	3.1472	0	10.5059	0	5.7783	0	81.3	76.3	5.0	78.1	3.2
24	2,2-dimethylpropyl methyl ether	2.9021	3.5175	0	10.3395	2.3902	0	3.2522	80.5	80.9	−0.4	79.7	0.8
25	1-ethylpropyl methyl ether	2.9601	3.0340	0	8.7146	5.7966	2.6017	0	88.5	81.9	6.6	83.8	4.7
26	ethyl tert-butyl ether	2.8992	2.4941	0	10.3585	3.3566	0	3.2589	73.1	76.0	−2.9	75.1	−2.0
27	hexyl methyl ether	3.7288	4.6632	0	7.5378	16.6124	0	0	125.0	124.5	0.5	125.0	0
28	ethyl propyl ether	3.7265	3.4572	0	7.5490	17.7683	0	0	118.0	118.8	−0.8	119.6	−1.6
29	butyl propyl ether	3.7257	2.9434	0	7.5546	18.2631	0	0	117.1	116.4	0.7	117.3	−0.2
30	butyl isopropyl ether ^a	3.6329	2.8393	0	9.8168	10.2860	4.3073	0	107.0	107.6	−0.6	108.5	−1.5
31	isobutyl propyl ether	3.6329	2.8393	0	9.8168	10.2860	4.3073	0	106.0	107.6	−1.6	108.5	−2.5
32	ethyl isopentyl ether	3.6338	3.3600	0	9.8094	9.7904	4.3073	0	111.0	110.0	1.0	110.8	0.2
33	tert-butyl propyl ether	3.4836	2.5756	0	11.4306	7.0988	0	3.9409	100.0	100.9	−0.9	98.2	1.8
34	tert-butyl isopropyl ether	3.4845	3.0951	0	11.4214	6.6058	0	3.9373	87.6	91.8	−4.2	89.3	−1.7
35	ethyl 1-methylbutyl ether	3.3645	2.4715	0	13.4495	0	3.8038	3.7466	105.5	103.8	1.7	104.7	0.8
36	2,2-dimethylpropyl ethyl ether ^a	3.4953	2.8464	0	9.1690	10.2527	2.9415	0	91.5	103.4	−11.9	100.6	−9.1
37	ethyl 2-methylbutyl ether	3.4983	3.0448	0	9.1576	9.8583	3.1917	0	109.5	104.4	5.1	105.3	4.2
38	ethyl tert-pentyl ether	3.4059	2.6935	0	11.2384	6.8069	0	3.2632	101.0	99.4	1.6	97.2	3.8
39	1,2-dimethylpropyl ethyl ether	3.5215	3.2831	0	11.0748	4.3293	6.8941	0	99.3	102.3	−3.0	103.1	−3.8
40	ethyl 1-ethylpropyl ether ^a	3.5077	2.7950	0	9.5982	10.1452	2.8461	0	104.5	104.4	0.1	105.2	−0.7
41	dibutyl ether	4.2689	3.0864	0	8.3087	24.0889	0	0	140.3	139.8	0.5	140.0	0.3
42	isopentyl propyl ether	4.1865	3.2158	0	10.8194	14.7222	5.1023	0	125.0	131.6	−6.6	131.7	−6.7
43	butyl isobutyl ether	4.1862	3.0005	0	10.8223	14.9297	5.1038	0	132.0	130.6	1.4	130.7	1.3
44	butyl sec-butyl ether	4.1372	2.8737	0	10.5370	15.6036	3.9225	0	130.5	129.8	0.7	129.8	0.7
45	butyl tert-butyl ether	4.0510	2.7877	0	12.5558	11.3191	0	4.7410	125.0	124.9	0.1	120.6	4.4
46	sec-butyl isobutyl ether	4.0429	2.7877	0	12.9282	6.9245	8.6247	0	122.0	120.6	1.4	120.5	1.5
47	di-sec-butyl ether	3.9884	2.6610	0	12.5804	7.8616	7.25	0	121.0	119.7	1.3	119.6	1.4
48	1,3-dimethylpentyl methyl ether	3.9910	4.1322	0	12.5582	6.4693	7.2364	0	121.0	126.6	−5.6	126.1	−5.1
49	1,3-dimethylbutyl ethyl ether	4.0216	3.2840	0	12.7034	7.1075	7.8102	0	121.5	123.1	−1.6	122.9	−1.4
50	diisobutyl ether ^a	4.0956	2.9150	0	13.2781	5.9826	10.0094	0	122.2	121.3	0.9	121.4	0.8
51	isobutyl tert-butyl ether	3.9474	2.7022	0	14.8856	2.7934	4.7234	4.5869	111.5	115.5	−4.0	111.3	0.2
52	di-tert-butyl ether	3.7781	2.4895	0	16.2978	0	0	8.3829	107.3	109.0	−1.7	101.4	5.9
53	isoproyl tert-pentyl ether	3.8598	2.6172	0	14.4182	3.4216	4.2602	3.6296	114.5	113.7	0.8	110.3	4.2
54	hexyl ethyl ether	4.2701	4.0002	0	8.2999	23.2082	0	0	143.0	144.0	−1.0	144.0	−1.0
55	heptyl methyl ether	4.2724	5.3729	0	8.2890	21.8924	0	0	151.0	150.5	0.5	150.1	0.9
56	1-ethylpropyl propyl ether	4.0563	2.5790	0	10.4820	15.0094	3.2169	0	128.5	126.5	2.0	126.4	2.1
57	ethyl 1-methylpentyl ether	4.1159	3.3623	0	10.4106	15.3873	3.3894	0	125.0	131.9	−6.9	131.8	−6.8
58	ethyl heptyl ether	4.7996	4.5851	0	9.0504	29.3072	0	0	166.6	168.3	−1.7	167.7	−1.1
59	butyl isopentyl ether	4.7240	3.2631	0	11.8444	20.1220	5.9502	0	157.0	152.8	4.2	152.4	4.6
60	tert-butyl isopentyl ether ^a	4.5109	3.0150	0	16.3709	6.1364	5.6631	5.5117	139.0	137.9	1.1	132.3	6.7
61	butyl pentyl ether	4.7981	3.3351	0	9.0610	30.5122	0	0	163.0	162.5	0.5	162.2	0.8
62	1,5-methylhexyl methyl ether	4.5963	4.8706	0	14.1547	9.1972	10.188	0	153.5	151.2	2.3	150.1	3.4
63	1,5-methylpentyl ethyl ether	4.5689	3.7583	0	13.9099	10.9534	9.2285	0	140.0	146.3	−6.3	145.4	−5.4
64	isoproyl isopentyl ether	4.6438	3.1915	0	14.5860	8.8782	11.7547	0	139.0	143.7	−4.7	142.9	−3.9
65	methyl 1-methylheptyl ether	4.6796	4.9120	0	11.4702	19.2430	4.5617	0	162.0	160.7	1.3	159.7	2.3
66	methyl octyl ether	4.8018	6.0996	0	9.0399	27.8555	0	0	173.0	175.4	−2.4	174.4	−1.4
67	2-ethylhexyl methyl ether	4.5596	4.7693	0	11.3217	18.3055	3.2650	0	159.5	157.7	1.8	156.4	3.1
68	3,5-dimethylhexyl methyl ether	4.5655	5.5473	0	13.7311	9.8134	8.6891	0	155.5	155.2	0.3	153.7	1.8
69	ethyl 1,1,3-trimethylbutyl ether	4.3863	3.2448	0	15.4441	7.1781	4.7696	3.7458	141.0	138.2	2.8	133.7	7.3
70	ethyl octyl ether ^a	5.3162	5.2027	0	9.8006	36.0746	0	0	189.2	191.6	−2.4	190.7	−1.5
71	ethyl 1,1,3,3-tetramethylbutyl ether	4.7864	3.5102	0	18.4256	7.4913	0	9.3167	156.5	155.4	1.1	145.1	11.4
72	bis (1-ethylpropyl) ether	4.9041	2.7848	0	14.4690	16.3893	7.4616	0	161.5	159.1	2.4	157.4	4.1
73	bis (1-methylbutyl) ether	5.0247	2.9611	0	14.3967	18.4193	7.9578	0	162.0	163.3	−1.3	161.9	0.1
74	butyl 1-methylpentyl ether	5.1658	3.2515	0	12.1171	27.8113	3.9228	0	170.0	173.7	−3.7	172.7	−2.7
75	diisopentyl ether	5.1761	3.3663	0	15.9211	14.5070	13.5689	0	172.0	163.0	9.0	162.2	9.8

Table 1 (Continued)

no.	compounds	X _u ^m	AI(O) ^b	AI(C=) ^c	AI(−CH ₃)	AI(>CH ₂)	AI(>CH−)	AI(>C<)	BP (°C)					
									exp	calcd ^d	res ^d	calcd ^e	res ^e	
Ethers														
76	dipentyl ether	5.3145	3.4909	0	9.8136	37.7253	0	0	186.8	183.6	3.2	183.2	3.6	
77	isopropyl heptyl ether	5.2484	4.1421	0	12.8745	25.3855	6.8326	0	173.0	176.7	−3.7	175.9	−2.9	
78	heptyl propyl ether	5.3151	4.1971	0	9.8078	37.0426	0	0	187.0	186.9	0.1	186.3	0.7	
79	isopentyl pentyl ether	5.2477	3.4284	0	12.8821	26.0688	6.8371	0	174.0	173.3	0.7	172.7	1.3	
80	methyl 1-methyloctyl ether ^a	5.2061	5.5297	0	12.4417	24.8861	5.2692	0	188.5	183.8	4.7	182.4	6.1	
81	octyl propyl ether	5.8202	4.7134	0	10.5579	44.6069	0	0	204.0	208.8	−4.8	208.0	−4.0	
82	isopropyl octyl ether	5.7597	4.6696	0	13.9231	31.6576	7.7506	0	198.5	197.8	0.7	196.9	1.6	
83	dihexyl ether	6.3136	3.8945	0	11.3170	54.1621	0	0	225.7	223.3	2.4	223.2	2.5	
84	diheptyl ether	7.2736	4.2976	0	12.8194	73.3992	0	0	258.5	259.2	−0.7	260.4	−1.9	
85	dioctyl ether	8.2001	4.7003	0	14.3212	95.4363	0	0	291.7	291.6	0.1	295.0	−3.3	
86	bis (2-ethylhexyl) ether	7.7680	3.8476	0	20.4474	58.1151	11.1525	0	269.8	268.2	1.6	267.6	2.2	
Aldehydes														
87	acetaldehyde	0.7153	1.9413	1.8235	1.8235	0	0	0	20.8	19.1	1.7	7.6	13.2	
88	propionaldehyde	1.3735	2.3781	2.2057	2.2057	1.5359	0	0	48.8	50.6	−1.8	43.7	5.1	
89	butyl aldehyde	1.9982	2.8091	2.7450	2.5828	4.6907	0	0	75.7	78.5	−2.8	74.4	1.3	
90	2-methyl propanal	1.8720	2.6899	2.4785	4.3363	0	2.3689	0	64.4	71.1	−6.7	67.7	−3.3	
91	pentaldehyde	2.5975	3.2383	3.3696	2.9583	7.9262	0	0	103	104.5	−1.5	102.2	0.8	
92	2-methyl butanal	2.4422	2.9689	2.8576	4.9559	2.6512	2.4983	0	92.5	95.8	−3.3	94.0	−1.5	
93	3-methyl butanal	2.4822	3.1234	3.1327	4.7298	2.2132	2.8199	0	92.5	97.1	−4.6	95.2	−2.7	
94	2,2-dimethylpropanal	2.1162	2.9202	2.6800	6.4999	0	0	2.6800	77.5	79.0	−1.5	79.0	−1.5	
95	hexanal	3.1755	3.6670	4.0476	3.3333	11.8094	0	0	128.0	128.9	−0.9	127.7	0.3	
96	2-methylpentanal ^a	3.0208	3.2605	3.3211	5.5770	5.8700	2.8312	0	117.0	119.6	−2.6	118.8	−1.8	
97	3-methylpentanal	3.0241	3.5127	3.6855	5.3431	5.5319	2.8269	0	118.0	120.6	−2.6	120.4	−2.4	
98	4-methylpentanal	3.0730	3.6487	3.9329	5.3486	5.0705	3.5028	0	121.0	121.6	−0.6	121.4	−0.4	
99	2-ethylbutanal	2.9646	3.1252	3.0845	5.7187	5.7058	2.5634	0	117.0	117.7	−0.7	116.9	0.1	
100	heptanal	3.7353	4.0953	4.7616	3.7081	16.3574	0	0	152.8	151.8	1.0	151.4	1.4	
101	2-methylhexanal	3.5894	3.5724	3.8484	6.2042	9.6580	3.2975	0	141.0	142.3	−1.3	142.2	−1.2	
102	3-methylhexanal	3.5771	3.8312	4.2191	5.9077	9.4282	3.0093	0	143.0	143.0	0	143.7	−0.7	
103	4-methylhexanal	3.5942	4.0098	4.5058	5.8018	9.0455	3.2904	0	144.0	143.7	0.3	144.5	−0.5	
104	5-methylhexanal	3.6443	4.0963	4.6871	5.9315	8.4922	4.2217	0	143.5	144.2	−0.7	144.7	−1.2	
105	2,2-dimethylpentanal	3.4118	3.3043	3.3143	8.0605	6.1874	0	3.2125	126.5 ^f	131.8	−5.3	133.6	−7.1	
106	2,3-dimethylpentanal ^a	3.4286	3.4791	3.5813	8.0729	3.2945	5.8719	0	141.0 ^f	134.1	6.9	135.1	5.9	
107	octanal	4.2789	4.5236	5.5011	4.0828	21.5805	0	0	171.0	173.6	−2.6	173.7	−2.7	
108	2-ethylhexanal	4.0608	3.5367	3.8089	7.1181	13.9362	3.1741	0	160.0	161.1	−1.1	161.4	−1.4	
109	2-propylpentanal	4.0330	3.4162	3.6097	7.2195	13.8938	2.9573	0	160.0	160.2	−0.2	160.4	−0.4	
110	nonanal	4.8084	4.9519	6.2592	4.4575	27.4854	0	0	191.0	194.2	−3.2	194.7	−3.7	
111	3,5,5-trimethylhexanal	4.4417	4.5042	5.2376	11.2826	5.9862	3.4253	4.9504	170.5	166.7	3.8	171.8	−1.3	
112	decanal	5.3249	5.3802	7.0314	4.8322	34.0767	0	0	208.5	213.9	−5.4	214.5	−6.0	
113	undecanal	5.8298	5.8085	7.8145	5.2069	41.3573	0	0	233.0	232.7	0.3	233.4	−0.4	
114	2-methyldecanal	5.7243	4.9710	6.3537	8.8120	30.8275	5.9572	0	229.0 ^f	221.7	7.3	223.0	6.0	
115	dodecanal	6.3241	6.2368	8.6062	5.5817	49.3297	0	0	254.0 ^f	250.6	3.4	251.3	2.7	
116	2-methylundecanal ^a	6.2265	5.3452	7.0438	9.4866	37.6974	6.7515	0	246.0 ^f	239.0	7.0	240.5	5.5	
117	tridecanal	6.8086	6.6652	9.4047	5.9564	57.9953	0	0	267.0	267.8	−0.8	268.4	−1.4	
118	tetradecanal	7.2841	7.0936	10.2088	6.3312	67.3555	0	0	287.0	284.2	2.8	284.7	2.3	
119	pentadecanal	7.7512	7.5220	11.0176	6.7060	77.4112	0	0	304.9	299.9	5.0	300.3	4.6	
Ketones														
120	acetone	1.2551	2.1288	2.1377	3.9751	0	0	0	56.2	47.4	8.8	42.2	14.0	
121	2-butanone	1.8667	2.3511	2.4775	4.6780	2.3298	0	0	79.6	74.4	5.2	71.2	8.4	
122	2-pentanone	2.4746	2.6245	3.0214	5.3558	5.2374	0	0	102.0	100.4	1.6	98.4	3.6	
123	3-pentanone	2.4355	2.4262	2.6169	5.4855	5.3411	0	0	101.7	98.3	3.4	96.0	5.7	
124	3-methyl-2-butanone	2.3432	2.5544	2.7623	7.0799	0	2.6522	0	93.5	92.7	0.8	91.5	2.0	
125	2-hexanone	3.0648	2.9336	3.6974	6.0332	8.7367	0	0	127.0	125.2	1.8	123.9	3.1	
126	3-hexanone ^a	3.0141	2.5861	2.9735	6.2246	8.9946	0	0	123.5	122.2	1.3	120.6	2.9	
127	3-methyl-2-pentanone	2.8967	2.7597	3.1896	7.8582	2.9706	2.6627	0	118.0	116.0	2.0	115.7	2.3	
128	4-methyl-2-pentanone	2.9486	2.8941	3.5063	7.9713	2.2946	3.3497	0	117.0	117.4	−0.4	117.0	0	
129	2-methyl-3-pentanone	2.8944	2.5365	2.7778	8.0685	2.9753	3.0576	0	115.5	114.5	1.0	113.5	2.0	
130	3,3-dimethyl-2-butanone	2.7668	2.7259	2.9969	9.4159	0	0	2.9722	106.0	107.3	−1.3	109.0	−3.0	
131	2-heptanone	3.6362	3.2672	4.4677	6.7171	12.8493	0	0	151.4	148.5	2.9	147.7	3.7	
132	3-heptanone	3.5830	2.8024	3.4726	6.9357	13.2723	0	0	147.0	145.4	1.6	144.3	2.7	
133	4-heptanone	3.5667	2.6550	3.1691	7.0137	13.4017	0	0	144.0	144.4	−0.4	143.0	1.0	
134	3-methyl-2-hexanone	3.5217	2.8764	3.5741	8.3353	7.3846	3.4132	0	143.5	139.5	4.0	138.9	4.6	
135	4-methyl-2-hexanone	3.4802	3.1541	4.0769	8.6880	5.9145	3.1666	0	139.0	139.7	−0.7	140.0	−1.0	
136	5-methyl-2-hexanone ^a	3.5343	3.2457	4.3254	8.8941	5.1825	4.1176	0	144.0	140.5	3.5	140.5	3.5	
137	2-methyl-3-hexanone	3.5200	2.5217	3.5771	8.6538	7.3909	3.4160	0	135.0	134.2	0.8	126.8	8.2	
138	4-methyl-3-hexanone	3.4208	2.6759	3.0853	8.8468	6.6340	2.9550	0	134.5	136.4	−1.9	136.0	−1.5	
139	5-methyl-3-hexanone	3.4768	2.7685	3.3179	9.0476	5.9260	3.8931	0	135.0	137.3	−2.3	136.8	−1.8	
140	2,2-dimethyl-3-pentanone	3.3059	2.6464	2.9299	10.5194	3.2429	0	3.4124	125.6	127.4	−1.8	128.9	−3.3	
141	2,4-dimethyl-3-pentanone	3.3444	2.5833	2.8559	10.9213	0	6.8603	0	125.4	128.8	−3.4	128.7	−3.3	
142	4,4-dimethyl-2-pentanone	3.3700	3.1398	3.9355	10.4214	2.3479	0	3.6959	126.4	130.6	−4.2	133.1	−6.7	
143	2-octanone	4.1901	3.6183	5.3086	7.4086	17.5939	0	0	172.5	170.7	1.8	170.0	2.5	
144	3-octanone	4.1378	3.0576	4.0763	7.6375	18.1765	0	0	167.5	167.6	−0.1	166.9	0.6	
145	4-octanone	4.1144	2.7947	3.5210	7.7601	18.4448	0	0	165.5	166.1	−0.6	165.1	0.4	
146	2-methyl-4-heptanone ^a	4.0203	2.7665	3.3933	10.0671	10.1635	4.5058	0	154.0	157.8	−3.8	157.3	−3.3	
147	3-methyl-4-heptanone	3.9654	2.6962	3.2076	9.7853	10.9351	3.3548	0	153.0	157.2	−4.2	156.9	−3.9	

Table 1 (Continued)

									BP (°C)				
no.	compounds	X _u ^m	AI(O) ^b	AI(C=) ^c	AI(−CH ₃)	AI(>CH ₂)	AI(>CH−)	AI(>C<)	exp	calcd ^d	res ^d	calcd ^e	res ^e
Ketones													
148	3-methyl-2-heptanone	4.0265	3.2493	4.3629	9.4887	10.6503	3.2357	0	164.0	161.1	2.9	161.5	2.5
149	4-methyl-2-heptanone	4.0251	3.4232	4.7224	9.4532	10.1308	3.2335	0	160.5	161.6	−1.1	162.1	−1.6
150	5-methyl-2-heptanone	4.0479	3.5443	5.0068	9.5558	9.4849	3.7598	0	166.5	161.9	4.6	162.4	4.1
151	6-methyl-2-heptanone	4.1005	3.6083	5.2024	9.8458	8.6763	4.9398	0	167.0	162.2	4.8	162.3	4.7
152	2-methyl-3-heptanone	4.0231	2.7654	3.3919	9.9374	10.6732	4.1621	0	158.0	158.3	−0.3	157.8	0.2
153	5-methyl-3-heptanone	3.9896	2.9663	3.7620	9.7437	10.3111	3.5894	0	161.0	158.6	2.4	158.7	2.3
154	6-methyl-3-heptanone	4.0441	3.0342	3.9524	10.0278	9.5039	4.7674	0	163.2	159.0	4.2	158.9	4.3
155	3,3-dimethyl-2-hexanone	3.8389	3.0532	3.7818	11.1751	6.7336	0	3.2955	151.5	150.3	1.2	152.9	−1.4
156	3,4-dimethyl-2-hexanone ^a	3.8637	3.1740	4.0538	11.3638	3.6417	6.1706	0	158.0	152.5	5.5	153.9	4.1
157	3,5-dimethyl-2-hexanone	3.9231	3.2382	4.2484	11.7105	2.6041	7.5080	0	154.0	152.8	1.2	154.0	0
158	4,4-dimethyl-2-hexanone	3.8653	3.3563	4.4198	11.2005	5.8658	0	3.5860	154.4	151.3	3.1	154.3	0.1
159	5,5-dimethyl-2-hexanone	3.9536	3.5415	4.9041	11.4840	5.3246	0	4.5259	146.0	152.0	−6.0	154.9	−8.9
160	2,2-dimethyl-3-hexanone	3.8668	2.6702	3.0782	11.5977	7.1225	0	3.9821	146.0	147.8	−1.8	149.6	−3.6
161	2,4-dimethyl-3-hexanone ^a	3.8597	2.6702	3.0782	11.8362	3.6507	7.1093	0	145.0	149.3	−4.3	149.6	−4.6
162	2,5-dimethyl-3-hexanone	3.9192	2.7389	3.2631	12.1786	2.6093	8.4593	0	147.5	149.8	−2.3	149.9	−2.4
163	4,4-dimethyl-3-hexanone	3.7993	2.7637	3.1945	11.3512	6.7935	0	3.3653	148.0	147.7	0.3	149.8	−1.8
164	4,5-dimethyl-3-hexanone	3.8609	2.8309	3.3830	11.6850	3.6477	6.8156	0	152.0	150.4	1.6	151.1	0.9
165	5,5-dimethyl-3-hexanone	3.8896	2.9474	3.6384	11.6363	6.2900	0	4.2980	146.5	148.7	−2.2	151.0	−4.5
166	3-ethyl-3-methyl-2-pentanone ^a	3.7647	2.9889	3.5886	11.1797	3.0780	0	3.0708	153.8	149.6	4.2	152.3	1.5
167	3-ethyl-4-methyl-2-pentanone	3.8277	3.0532	3.7818	11.5113	3.2004	6.3295	0	154.0	150.8	3.2	152.2	1.8
168	2,2,4-trimethyl-3-pentanone	3.7492	2.6478	2.9491	13.5814	0	3.7680	3.8256	135.1	139.9	−4.8	142.2	−7.1
169	2-nonanone	4.7284	3.9823	6.2039	8.1073	22.9849	0	0	195.0	191.8	3.2	191.1	3.9
170	3-nonanone	4.6782	3.3409	4.7614	8.3379	23.7161	0	0	190.0	188.8	1.2	188.3	1.7
171	4-nonanone	4.6516	2.9833	3.9860	8.4856	24.1226	0	0	187.5	187.1	0.4	186.4	1.1
172	5-nonanone	4.6434	2.8689	3.7437	8.5364	24.2523	0	0	188.4	186.5	1.9	185.7	2.7
173	4-methyl-2-octanone	4.5657	3.7072	5.4322	10.2532	14.9330	3.4587	0	184.0	182.6	1.4	183.2	0.8
174	5-methyl-3-octanone	4.5169	3.1855	4.2905	10.4869	15.3026	3.5494	0	179.0	179.6	−0.6	180.0	−1.0
175	7-methyl-3-octanone	4.5952	3.3252	4.6617	11.0179	13.7104	5.6725	0	182.5	179.6	2.9	179.7	2.8
176	3-methyl-4-octanone ^a	4.5085	2.7898	3.4813	10.7047	15.8395	3.8417	0	174.0	177.7	−3.7	177.6	−3.6
177	2,5-dimethyl-3-heptanone	4.4238	2.8866	3.6156	13.0295	6.9359	8.5888	0	166.0	169.8	−3.8	170.4	−4.4
178	2,6-dimethyl-3-heptanone	4.4800	2.9396	3.7705	13.4371	5.8268	10.1298	0	171.5	169.8	1.7	170.2	1.3
179	3,3-dimethyl-2-heptanone	4.3936	3.2546	4.3100	12.1154	11.1201	0	3.7114	173.5	171.0	2.5	173.8	−0.3
180	4,6-dimethyl-2-heptanone	4.4757	3.7076	5.3505	12.6923	5.8371	7.4344	0	173.0	175.8	−2.8	177.0	−4.0
181	4,6-dimethyl-3-heptanone	4.4154	3.0459	3.9223	12.8584	6.9132	8.1643	0	171.5	170.6	0.9	171.6	−0.1
182	5,5-dimethyl-3-heptanone	4.3684	3.1235	4.0331	12.3696	10.4691	0	4.0711	166.5	168.8	−2.3	171.6	−5.1
183	7-methyl-4-octanone	4.5676	2.9604	3.8780	11.0983	14.4253	5.4601	0	178.0	178.0	0	177.9	0.1
184	2,3-dimethyl-4-heptanone	4.3960	2.7998	3.4150	12.7860	7.9070	7.6462	0	167.5	169.9	−2.4	170.4	−2.9
185	2,6-dimethyl-4-heptanone	4.4698	2.8207	3.5241	13.5180	5.8544	10.1938	0	169.4	169.0	0.4	169.2	0.2
186	3,5-dimethyl-4-heptanone ^a	4.3555	2.7132	3.2179	12.7199	7.9956	7.1946	0	162.0	168.8	−6.8	169.4	−7.4
187	2,2,4,4-tetramethyl-3-pentanone	4.1483	2.6782	2.9984	16.4083	0	0	8.41710	152.0	149.2	2.8	153.9	−1.9
188	2-decanone	5.2525	4.3561	7.1421	8.8123	29.0331	0	0	210.0	211.8	−1.8	210.9	−0.9
189	3-decanone	5.2050	3.6456	5.5119	9.0401	29.9008	0	0	211.0	209.1	1.9	208.7	2.3
190	4-decanone	5.1771	3.2073	4.5390	9.2014	30.4394	0	0	206.5	207.3	−0.8	206.9	−0.4
191	5-decanone	5.1645	2.9997	4.0894	9.2850	30.6959	0	0	204.0	206.4	−2.4	205.9	−1.9
192	2-methyl-5-nonone	5.0895	2.9884	3.9946	12.1321	19.9432	6.1927	0	203.5	196.9	6.6	197.1	6.4
193	2-undecanone	5.7639	4.7377	8.1147	9.5225	35.7469	0	0	231.5	230.8	0.7	229.6	1.9
194	3-undecanone	5.7191	3.9668	6.3119	9.7454	36.7390	0	0	227.0	228.4	−1.4	228.0	−1.0
195	5-undecanone	5.6759	3.1717	4.5295	10.018	37.7801	0	0	227.0	225.5	1.5	225.3	1.7
196	6-undecanone ^a	5.6712	3.0793	4.3279	10.0536	37.9027	0	0	226.0	225.2	0.8	224.9	1.1
197	2,2,6,6-tetramethyl-4-heptanone	5.2769	2.9393	3.7701	19.9694	6.0388	0	12.6481	185.0	181.0	4.0	187.6	−2.6
198	2-dodecanone	6.2638	5.1255	9.1154	10.2373	43.1325	0	0	246.5	249.0	−2.5	247.4	−0.9
199	2-methyl-3-undecanone	6.1387	3.7164	5.7757	13.7244	33.1071	7.0940	0	233.5	235.1	−1.6	235.6	−2.1
200	2-tridecanone	6.7532	5.5182	10.1392	10.9559	51.1948	0	0	263.0	266.3	−3.3	264.2	−1.2
201	7-tridecanone	6.6567	3.2891	4.9172	11.5666	54.3546	0	0	261.0	260.7	0.3	261.0	0
202	2-methyl-3-tridecanone	7.1221	4.3208	7.3175	15.7224	47.8501	8.7993	0	267.0	268.0	−1.0	268.5	−1.5
203	7-ethyl-2-methyl-4-undecanone	6.8526	3.5718	5.4181	18.5202	33.2103	13.072	0	252.5	255.7	−3.2	257.7	−5.2
204	2-pentadecanone	7.7038	6.3155	12.242	12.4025	69.3638	0	0	294.0	298.7	−4.7	295.5	−1.5
205	8-pentadecanone	7.6057	3.4991	5.5059	13.0767	73.6078	0	0	291.0	293.4	−2.4	294.5	−3.5

^a The prediction set not involved in the construction of the final models. ^b The symbol AI(O) stands for the AI(−O−) index for ethers. ^c The symbol AI(C=) stands for the AI(−CH=) for aldehydes and AI(>C=) indices for ketones. ^d Calculated from the two subset correlations. ^e Calculated from the whole data set correlation. ^f From ref 29.

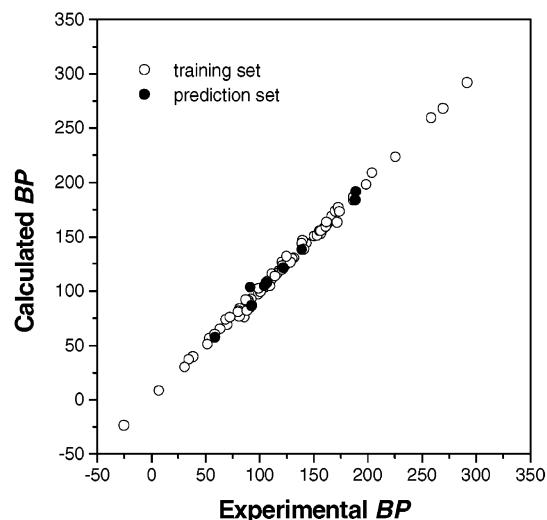
the criterion on high quality QSPR models proposed by Mihalić and Trinajstić³⁸ (e.g., $r > 0.995$ and $s < 5.0$ °C for BP). Of course, the linear regression using a single X_u^m index produces a slightly poor correlation for 77 compounds and gives $r = 0.9983$ and $s = 6.47$ °C. Obviously, the standard error in the best model (eq 13) is significantly reduced by 49% compared to the linear model with the X_u^m index. The results support our anticipation (eq 9) and also

show the usefulness of these indices in structure-boiling point correlations of ethers.

Recently, Balaban et al.²⁸ reported two best structure-boiling point models in terms of three topological indices for the same series of 72 ethers. In their study, the first model based on the first-order connectivity $^1\chi$ index and Balaban's modified J_{het} shape index produced $r = 0.993$ ($r^2 = 0.978$) and $s = 6.49$ °C. The second model with the electrotopo-

Table 2. MLR Models Obtained for the Ether Subset Based on the X_u^m and AI Indices

indices	model 1			model 2			corresponding atom types
	coefficients	errors	<i>t</i> -values	coefficients	errors	<i>t</i> -values	
constant	-58.9051	2.3484	-25.08	-66.4027	2.4348	-27.27	
X_u^m	50.5810	1.6651	30.38	49.2263	1.2710	38.73	
AI(-O-)	4.1098	0.6656	6.178	3.9722	0.6158	6.450	-O-
AI(-CH ₃)	-1.9572	0.3394	-5.766				-CH ₃
AI(>CH ₂)	-0.5743	0.09328	-6.157	-0.6742	0.09244	-7.293	>CH ₂
AI(>CH-)	-0.8520	0.1623	-5.248	-2.1384	0.2326	-9.193	>CH-
AI(>C<)				-2.4418	0.3507	-6.963	>C<
	$r^2 = 0.9962$			$r^2 = 0.9967$			
	$s = 3.53$ °C			$s = 3.30$ °C			
	$F = 3715$			$F = 4260$			
	$N = 77$			$N = 77$			

**Figure 1.** A plot of calculated or predicted *BP* values versus experimental data for the ether subset.

logical state of oxygen atom (S_o), number of methyl groups (N_{Me}), and Kier-Hall's valence index ($^0\chi^v$) produced an improved correlation ($r^2 = 0.9872$ and $s = 4.89$ °C). More recently, Ivanciuc et al.³⁹ further analyzed the relationships between boiling points and chemical structures using the CODESSA (Comprehensive Descriptors for Structural and Statistical Analysis) program for the same series of 73 ethers. Some excellent structure-boiling point correlations were constructed. The best four-, five-, and six-variable models gave $r = 0.996, 0.997, 0.997$ and $s = 4.0, 3.9,$ and 3.8 °C, respectively. The best model (eq 13) obtained in this study is obviously comparable to those obtained by Ivanciuc et al.,³⁹ although there are several discrepancies between the tabulated experimental *BP* data.

Furthermore, the relative (ψ_r) or fraction contributions (ψ_f) of the topological indices to *BP* is analyzed to manifest the role of the corresponding structural features in molecules⁴⁰

$$\psi_r(i) = a_i \bar{T}I_i \quad (14)$$

$$\psi_f(i) = r^2 |\psi_r(i)| / \sum_i |\psi_r(i)| \times 100\% \quad (15)$$

where a_i and $\bar{T}I_i$ are the coefficient and the average value of the i th topological index in the model. $\psi_r(i)$ and $\psi_f(i)$ are the relative and fraction contributions of the i th topological indices in the model. The sum is over all indices involved in the model.

The results obtained from eq 13 show that the X_u^m index has a larger ψ_f value (85.5%) than the other AI indices, and the ψ_f values for the four AI indices decrease in the order of AI(-O-) (5.8%) > AI(>CH₂) (4.5%) > AI(>CH-) (3.0%) > AI(>C<) (0.8%). The results indicate that the boiling points of ethers are dominantly determined by molecular size (because the X_u index characterizes molecular size⁹) and also depend on various parts of the molecule, particularly the -O- group. This may be readily understood because the boiling point is a physical property that strongly depends on the intermolecular interactions and is very much influenced by the molecular weight. Actually, the intermolecular interactions in complex compounds, as pointed out by Pitzer et al.⁴¹ in 1955, are not only interactions between molecular center but also a sum of interactions between various parts of the molecules, hence, both the molecular size and individual atom types or groups related to different fundamental interactions would provide the separate contributions to the *BP* value of a molecule.

In the case of ethers, polar interactions between oxygen atoms (-O- group) may be the second important factor. As expected, inclusion of the AI(-O-) index into the equation obviously improves the quality of the two-parameter model and produces $r = 0.9963$ and $s = 4.80$ °C. It is clear that the significant improvement should be attributed to the influence of polar interactions between oxygen atoms, because it is expected that the oxygen atoms would participate in the polar interactions between molecules. As a result, the AI(-O-) index makes a slightly large positive contribution ($\psi_r = 13.3$). On the other hand, as may be seen from Table 1, the position of oxygen atoms has an important effect on the *BP* values. In general, the closer to the middle of the chain the oxygen atoms are, the lower *BP* values the straight chain ethers have. The AI(-O-) index seems to simply reflect the strength of polar interactions between oxygen atoms in ethers with different chemical structures. In contrast, the three other atomic types (groups) make small negative contributions ($\psi_r = -10.3, -6.9,$ and -1.8 , respectively) to *BP*. Because AI(>CH₂), AI(>CH-), and AI(>C<) indices are obviously related to branching of a molecule,¹⁵⁻¹⁸ while branching prevents close contact with neighboring molecules and reduces the strength of intermolecular interactions due to their steric hindrance,⁴¹ *BP* displays negative dependence on the AI indices of three alkyl fragments.

Aldehyde and Ketone Subset. Recently, Li et al.⁴² studied the relationships between *BP* and chemical structures using 12 distance-edge vectors (μ) for a series of 72 carbonyl

Table 3. MLR Models Obtained for 77 Carbonyl Compounds Based on the X_u^m and AI Indices

indices	model 1			model 2			corresponding atom types
	coefficients	errors	<i>t</i> -values	coefficients	errors	<i>t</i> -values	
constant	-33.4042	5.4253	-6.157	-25.6024	5.6852	-4.503	
X_u^m	43.2732	0.6402	67.59	43.9577	1.1602	37.89	
AI(=O)	32.0217	4.0259	7.954	14.6442	5.2799	2.774	=O
AI(-CH=)	-20.5381	2.4306	-8.450	-8.3124	3.1365	-2.650	-CH=(O)
AI(>C=)	-14.0572	1.7845	-7.877	-6.2511	2.2958	-2.723	>C=(O)
AI(-CH ₃)	-2.4456	0.1868	-13.09				-CH ₃
AI(>CH ₂)				-0.4365	0.1058	-4.124	>CH ₂
AI(>CH-)				-1.8273	0.2046	-8.930	>CH-
AI(>C<)				-3.3337	0.2686	-12.41	>C<
	$r^2 = 0.9965$			$r^2 = 0.9971$			
	$s = 3.31\text{ }^\circ\text{C}$			$s = 3.02\text{ }^\circ\text{C}$			
	$F = 5700$			$F = 4874$			
	$N = 107$			$N = 107$			

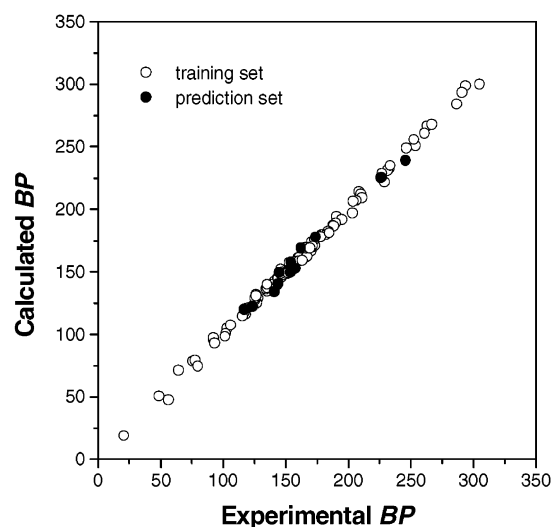
compounds containing 28 aldehydes and 44 ketones. However, the tabulated *BP* values in their study may include some estimated data. Amboni et al.⁴³ also developed a simple linear *BP* model for 35 oxo compounds using the semiempirical topological index (I_{ET}). The I_{ET} index leads to $r = 0.997$ and $s = 3.35\text{ }^\circ\text{C}$.

In the present study, the data set is extended to 119 carbonyl compounds, which are randomly divided into a training set and a test set. The models for the training set of 107 compounds are generated in terms of the X_u^m index and all AI indices present in molecules according to the same procedure. Two suitable models are obtained (Table 3). The best model is shown in eq 16:

$$BP = -25.6024 (\pm 5.6852) + 43.9577 (\pm 1.1602) X_u^m + 14.6442 (\pm 5.2799) \text{AI(=O)} - 8.3124 (\pm 3.1365) \text{AI(>CH=)} - 6.2511 (\pm 2.2958) \text{AI(>C=)} - 0.4365 (\pm 0.1058) \text{AI(>CH}_2\text{)} - 1.8273 (\pm 0.2046) \text{AI(>CH-)} - 3.3337 (\pm 0.2686) \text{AI(>C<)} \quad (16)$$

$$r = 0.9985; r_{cv} = 0.9982; s = 3.02; s_{cv} = 3.28; F = 4874; P < 0.0001; N = 107$$

The *t*-values indicate that each parameter is highly significant ($p < 0.0001$). The variables in the final model are not highly correlated with each other by cross-correlation analysis. This model explains more than 99% of the variance in the experimental *BP* values for 107 carbonyl compounds. The improvement in the statistical quality is about 46% compared to the simple model with the X_u^m index ($r = 0.9947$ and $s = 5.63\text{ }^\circ\text{C}$). The results indicate again that a combination of the X_u^m and AI indices can provide excellent structure-boiling point correlations. The statistics for the training set are $r = 0.9985$ and $s = 3.02\text{ }^\circ\text{C}$ and for the cross-validation set $r_{cv} = 0.9982$ and $s_{cv} = 3.25\text{ }^\circ\text{C}$. The cross-validated s_{cv} is only slightly larger than the standard error for the training set, indicating that the final model is statistically significant. In addition, the *BP* values of 12 compounds as the test set are predicted by using the above model (eq 16). The *PRESS*/*SSY* ratio for the test is 0.016, which is smaller than 0.1. The model gives $r_{pred} = 0.9921$ and $s_{pred} = 4.59\text{ }^\circ\text{C}$ for the test set, indicating an excellent predictive quality of the model.

**Figure 2.** A plot of calculated or predicted *BP* values versus experimental data for the aldehyde and ketone subset.

The calculated *BP* values and residuals for 107 compounds, along with the predicted *BP* values for the nine compounds, are shown in Table 1. A plot of calculated or predicted *BP* values versus experimental data for 119 compounds is shown in Figure 2. One can observe that calculated or predicted *BP* values agree well with experimental data, and the plot shows no obviously observable pattern. Therefore, the final model (eq 16) is considered to be excellent judging from the statistics and the plot in Figure 2. Obviously, the best model for the aldehyde and ketone subset in this study is better than that obtained by Li et al.⁴² for 72 compounds and is comparable to that obtained by Amboni et al.⁴³ for 35 ones, although there are several discrepancies between the tabulated experimental *BP* data.

It is shown that the X_u^m index has a larger ψ_f value (65.6%) than the other AI indices, and the ψ_f values for four AI indices decrease in the order of AI(=O) (18.1%) > AI(>C=) (7.0%) > AI(-CH=) (4.1%) > AI(>CH₂) (2.5%) > AI(>CH-) (1.5%) > AI(>C<) (0.8%). The results again indicate that the boiling points of aldehydes and ketones dominantly depend on molecular size and also depend on various parts of molecules, particularly the oxygen heteroatom of the carbonyl groups (>CO or -CHO). The slightly large ψ_f value for the AI(=O) index may be attributed to the influence of oxygen atoms of carbonyl groups, which would create polar interactions between molecules and thus

Table 4. MLR Models Obtained for the Entire Data Set Based on the X_u^m and AI Indices

indices	model 1			model 2			corresponding atom types
	coefficients	errors	<i>t</i> -values	coefficients	errors	<i>t</i> -values	
constant	-42.9396	2.0743	-20.70	-54.1919	2.1938	-24.70	
X_u^m	42.4663	0.4969	85.46	44.7638	0.9714	46.08	
AI(-O-)	4.7349	0.6536	7.244	3.9919	0.6283	6.353	-O-
AI(=O)	35.6441	1.5496	23.00	33.8787	1.5288	22.16	=O
AI(-CH=)	-21.9053	1.0084	-21.72	-19.7491	0.9747	-20.26	-CH=(O)
AI(>C=)	-15.4474	0.8016	-19.27	-14.9103	0.7660	-19.46	>C=(O)
AI(-CH ₃)	-1.7089	0.1541	-11.09				-CH ₃
AI(>CH ₂)				-0.3835	0.07391	-5.189	>CH ₂
AI(>CH-)				-1.7079	0.1820	-9.384	>CH-
AI(>C<)				-2.8046	0.2627	-10.68	>C<
	$r^2 = 0.9947$			$r^2 = 0.9955$			
	$s = 4.30\text{ }^\circ\text{C}$			$s = 3.99\text{ }^\circ\text{C}$			
	$F = 5551$			$F = 4834$			
	$N = 107$			$N = 107$			

make both aldehydes and ketones have higher boiling points than the corresponding apolar or weakly polar compounds with the same molecular sketch. As a result, the AI(=O) index shows a positive contribution value of $\psi_r = 50.5$. This is in accordance with the polar character of carbonyl compounds. In addition, the position of the carbonyl groups has also the influence on the BP values. As may be seen from Table 1, for straight-chain compounds with equal numbers of non-hydrogen atoms, both the AI(=O) indices and the BP values decrease when the carbonyl groups (>CO or >CH=) move from the peripheral toward the middle of the chain. This fact implies that the C=O bond of carbonyl groups in molecules behaves like a pseudobranch (although the C=O bond of aldehydes is present in the main chain). Hence, the AI(>C=) and AI(-CH=) indices, along with the AI(>CH₂), AI(>CH-), and AI(>C<) indices, provide negative contributions to BP.

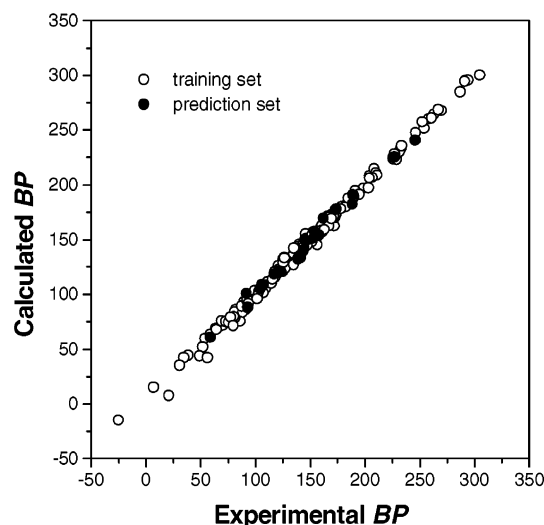
Combined Set. To further verify the potential of these indices in more complex systems, as an extension of the above studies, MLR analyses are performed on the entire data set of 184 compounds, which are the same as used in the structure-boiling point correlations for the two training subsets. Two suitable models are obtained (Table 4). We give below the final model (eq 17):

$$\begin{aligned}
 BP = & -54.1919 (\pm 2.1938) + 44.7638 (\pm 0.9714) \\
 & X_u^m + 3.9919 (\pm 0.6283) \text{AI}(-\text{O}-) + \\
 & 33.8787 (\pm 1.5288) \text{AI}(=\text{O}) - \\
 & 19.7491 (\pm 0.9747) \text{AI}(>\text{CH}-) - \\
 & 14.9103 (\pm 0.7660) \text{AI}(>\text{C}=) - \\
 & 0.3835 (\pm 0.07391) \text{AI}(>\text{CH}_2) - \\
 & 1.7079 (\pm 0.1820) \text{AI}(>\text{CH}-) - \\
 & 2.8046 (\pm 0.2627) \text{AI}(>\text{CH}-) \quad (17)
 \end{aligned}$$

$$r = 0.9977; r_{cv} = 0.9974; s = 3.99; s_{cv} = 4.16;$$

$$F = 4834; P < 0.0001; N = 184$$

All parameters are highly significant according to the *t*-values ($p < 0.0001$). This model explains more than 99% of the variance in the experimental BP values for 184 compounds. One can see again that a combined use of the X_u^m and AI indices produces a good correlation and gives $r = 0.9977$ and $s = 3.99\text{ }^\circ\text{C}$ for the training set and $r_{cv} = 0.9974$ ($r_{cv}^2 = 0.9949$) and $s_{cv} = 4.16\text{ }^\circ\text{C}$ for the cross-

**Figure 3.** A plot of calculated or predicted BP values versus experimental data for the whole data set.

validation. This cross-validation test demonstrates the model to be statistically significant. In addition, the above model (eq 17) is used to predict the BP values of 21 compounds in the test set. The PRESS/SSY ratio for the test is 0.01, which is smaller than 0.1. The predictive r_{pred} is 0.9949 and s_{pred} is 4.38 $^\circ\text{C}$ for the test set, indicative of an outstanding predictive ability of the model.

It is noted that the standard error for the combined set is slightly increased to 3.99 $^\circ\text{C}$, which is only slightly larger than those of the two separate correlations. The slightly large standard error for the total data set may be due to some of the inhomogeneities of the database. Similar observation was reported by Balaban et al.⁴⁴ in the structure-boiling point correlations for a data set of 200 acyclic saturated and unsaturated carbonyl compounds (mono- and dialdehydes, mono- and diketones, keto aldehydes, and esters of mono-carboxylic acids) in terms of five topological indices. The separate correlations produced a standard error of $s = 6.93\text{ }^\circ\text{C}$ for the combined set of all 200 carbonyl compounds, and $s = 6.49\text{ }^\circ\text{C}$ for 127 mono- and dialdehydes and -ketones, and $s = 4.0\text{ }^\circ\text{C}$ for 73 esters.

The calculated BP values and residuals from eq 17 for 184 compounds, along with predicted BP values for the 21 compounds, are shown in Table 1. A comparison of calculated or predicted BP values and experimental data for

Table 5. Comparison of Calculated or Predicted *BP* Values from the Best Models Obtained in This Study and Tabulated Experimental Data from the Literature for Several Compounds

no.	compounds	exp. <i>BP</i> (°C)	pred. <i>BP</i> ^a (°C)	pred. <i>BP</i> ^b (°C)
1	<i>tert</i> -butyl methyl ether	55.2 ²⁷ / 52 ²⁶	50.9	51.8
2	ethyl 1-ethylpropyl ether	104.5 ²⁵ /90 ²⁸	104.4	105.2
3	octanal	171 ²³ /163.4 ²⁵	173.6	173.7
4	3,5,5-trimethylhexanal	208.5 ^{23,24} / 170.5 ²⁵	166.7	171.8
5	4-methyl-2-hexanone	142 ²³ /144.5 ²⁴ /146.5 ²⁵ , 139 ²⁵	139.7	140.0
6	3-octanone	165.5 ²⁵ , 170 ²⁵ / 167.5 ²⁴	167.6	166.9
7	3,3-dimethyl-2-hexanone	151.5 ²⁵ /147 ²⁴	150.3	152.9
8	5-nonone	188.4 ²³ /181.5 ²⁵	186.5	185.7
9	4,6-dimethyl-2-heptanone	173 ²⁵ , 163 ²⁵	175.8	177.0
10	4,6-dimethyl-3-heptanone	171.5 ²⁵ , 182.5 ²⁵	170.6	171.6
11	2,6-dimethyl-4-heptanone	165.5 ²⁵ / 169.4 ²⁴	169.0	169.2
12	3-decanone	203 ²⁴ / 211 ²⁵	209.1	208.7

^a From the subset correlations (eqs 13 and 16). ^b From the entire data set correlation (eq 17).

205 compounds is shown in Figure 3. One can observe that calculated or predicted *BP* values are in good agreement with experimental data, and the plot shows no obviously observable pattern except acetaldehyde and acetone which have larger absolute residuals ($> 3 \times s$). The final model (eq 17) is therefore considered as an excellent QSPR model judging from the statistics and the plot in Figure 3. The reason acetaldehyde and acetone have larger residuals is not clear. Of course, if the two compounds are excluded from the combined set, a better correlation using the same variables can be obtained ($r = 0.9980$ and $s = 3.67$ °C) and gives $r_{\text{pred}} = 0.9958$ and $s_{\text{pred}} = 3.92$ °C.

On the other hand, the ψ_i values of individual indices are calculated as follows: X_u^m (58.3%), $\text{AI}(=\text{O})$ (21.7%), $\text{AI}(-\text{O}-)$ (1.8%), $\text{AI}(>\text{C}=\text{C})$ (8.7%), $\text{AI}(-\text{CH}=\text{C})$ (4.1%), $\text{AI}(=\text{CH}_2)$ (1.9%), $\text{AI}(=\text{CH}-)$ (1.5%) $>$ $\text{AI}(=\text{C}-)$ (0.6%). The results manifest again that the boiling point of oxo compounds dominantly depends on molecular size and also depends on various parts of molecules, particularly heteroatoms. This is in agreement with our anticipation (eq 9) and also supports the statement on intermolecular interactions by Pitzer et al.⁴¹

For an easy comparison, Table 5 lists the calculated or predicted *BP* values from the best models and experimental *BP* data from different references, where it can be found that the experimental *BP* values are unfortunately not unique for a given compound. For example, the tabulated *BP* value for ethyl 1-ethylpropyl ether is 104.5 °C,²⁵ while Balaban et al.²⁸ in their study quoted the *BP* value of 90.0 °C, but it was found to be an outlier with a predicted *BP* of 102.3 °C. In this study, the predicted *BP* values are 104.4 °C and 105.2 °C from the ether subset (eq 13) and the entire data set (eq 17) correlations, respectively, which are very close to the reported *BP* = 104.5 °C²⁵ and the predicted *BP* = 102.3 °C by Balaban et al.²⁸ Hence, 104.5 °C is considered as the reliable experimental *BP* value of ethyl 1-ethylpropyl ether. In another instance, two *BP* values differing by 38 °C for 3,5,5-trimethylhexanal were reported, i.e., 208.5 °C^{23,24} and 170.5 °C.²⁵ It is obvious that the lower is more reliable by comparison with the analogous nonanal (191 °C) and decanal (208.5 °C), because one would not expect the boiling point of 3,5,5-trimethylhexanal to be higher than that of nonanal or to be the same as that of decanal. The calculated value of *BP* = 171.8 °C from eq 16 for 3,5,5-trimethylhexanal is in accordance with the reported value of *BP* = 170.5 °C. From this it is concluded that the three best models may be used

as a basis to estimate the boiling points of unknown compounds and/or assess the reliability of experimental boiling points of compounds.

Finally, it should be mentioned that the modified Xu index and atom-type AI indices show an expected excellent discrimination power of isomers for all investigated compounds in the database. Therefore, these indices promise to be effective descriptors for QSPR/QSAR modeling and particularly database characterization in the future.

5. CONCLUSION

A few excellent QSPR models have been successfully developed to calculate or predict the normal boiling points for a series of aliphatic oxo compounds from topological indices using multiple linear regression (MLR) method. The best model based on a combined use of the modified Xu and atom-type-based AI indices produces $r = 0.9977$ and $s = 3.99$ °C for the total data set of 184 compounds. For the two subsets containing respectively 77 ethers or 107 carbonyl compounds, separate correlations significantly improve the quality of the models and give $s = 3.30$ and 3.02 °C, respectively. The best models may be used as a basis to estimate the boiling points of unknown compounds and/or assess the reliability of boiling points of compounds. The results indicate that the boiling points of oxo compounds studied dominantly depend on molecular size and also depend on the individual atom types, especially the oxygen heteroatoms in molecules due to intermolecular polar interactions. These excellent structure-boiling point models not only provide profound insights into the role of structural features in a molecule but also illustrate the usefulness of these indices in QSPR/QSAR modeling of complex compounds.

REFERENCES AND NOTES

- (1) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR. Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995.
- (2) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR. Hydrophobic, Electronic and Steric Constants*; American Chemical Society: Washington, DC, 1995.
- (3) Randić, M. The Connectivity Index 25 Years After. *J. Mol. Graph. Mod.* **2001**, 20, 19–35.
- (4) Trinajstić, N. *Chemical Graph Theory*, 2nd ed.; CRC Press: Boca Raton, 1992.
- (5) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.

- (6) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Studies*; Research Studies Press: Letchworth, 1986.
- (7) Balaban, A. T. Highly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett.* **1982**, 89, 399–404.
- (8) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, 69, 17–20.
- (9) Ren, B. A New Topological Index for QSPR of Alkanes. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 139–143. Ren, B.; Chen, G.; Xu, Y. A Novel Topological Index for QSPR/QSAR Study of Organic Compounds. *Acta Chim. Sinica* (in Chinese) **1999**, 57, 563–571. Ren, B.; Xu, Y.; Chen, G. Estimation of Heat Capacity of Complex Organic Compounds by a Novel Topological Index. *J. Chem. Eng. China* (in Chinese) **1999**, 50, 280–286. Ren, B.; Luo, B.; Zhang, Y. QSPR Studies of Solubilities and Octanol/Water Partition Coefficients of Organic Compounds. *J. S. China Univ. Technol.* (in Chinese) **1999**, 27, 89–95.
- (10) Huuskonen, J. QSAR Modeling with the Electrotopological State: TBO Derivatives. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 425–429.
- (11) Hall, L. H.; Mohnney, B.; Kier, L. B. The Electrotopological State: Structural Information at the Atomic Level for Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1991**, 31, 76–82.
- (12) Maw, H. H.; Hall, L. H. E-State Modeling of Corticosteroids Binding Affinity Validation of Model for Small Data Set. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 1248–1254.
- (13) Tetko, I. V.; Tanchuk, V. Y.; Kasheva, T. N.; Villa, A. E. P. Estimation of Aqueous Solubility of Chemical Compounds Using E-State Indices. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 1488–1493.
- (14) Rose, K.; Hall, L. H.; Kier, L. B. Modeling Blood-Brain Barrier Partitioning Using the Electrotopological State. *J. Chem. Inf. Comput. Sci.* **2002**, 42, 651–666.
- (15) Ren, B. Novel Atom-Type AI Topological Indices for QSPR Studies of Alkanols. *Comput. Chem.* **2002**, 26, 223–235.
- (16) Ren, B. Application of Novel Atom-Type AI Topological Indices to QSPR Studies of Alkanes. *Comput. Chem.* **2002**, 26, 357–369.
- (17) Ren, B. Novel Atomic-Level-Based AI Topological Descriptors: Application to QSPR/QSAR Modeling. *J. Chem. Inf. Comput. Sci.* **2002**, 42, 858–868.
- (18) Ren, B. Application of Novel Atom-Type AI Topological Indices in the Structure–Property Correlations. *J. Mol. Struct. (Theochem)* **2002**, 586, 137–148.
- (19) Ren, B. Atom-Type-Based AI Topological Descriptors for Quantitative Structure–Retention Index Correlations of Aldehydes and Ketones. *Chemometrics Intel. Lab. Syst.* **2003**, 66, 29–39.
- (20) Ren, B. Atomic-level-Based AI Topological Descriptors for Structure–Property Correlations. *J. Chem. Inf. Comput. Sci.* **2003**, 43, 161–169.
- (21) Rücker, G.; Rücker, C. On Topological Indices, Boiling Points, and Cycloalkanes. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 788–802.
- (22) Toropov, A. A.; Toropova, A. P. Modeling of Acyclic Carbonyl Compounds Normal Boiling Points by Correlation Weighting of Nearest Neighboring Codes. *J. Mol. Struct. (Theochem)* **2002**, 581, 11–15.
- (23) Weast, R. *CRC Handbook of Chemistry and Physics*, 70th ed.; CRC Press: Boca Raton, FL, 1989–1990.
- (24) Lide, D. R.; Milne, G. W. A. *Handbook of Data on Common Organic Compounds*; CRC Press: Boca Raton, FL, 1992.
- (25) *Dictionary of Organic Chemistry*, 6th ed.; Chapman & Hall: London, 1996.
- (26) Dean, J. A. *Lange's Handbook of Chemistry*, 15th ed.; McGraw-Hill: Beijing, 1999.
- (27) Yaws, C. L. *Chemical Properties Handbook*; McGraw-Hill: Beijing, 1999.
- (28) Balaban, A. T.; Kier, L. B.; Joshi, N. Correlations between Chemical Structure and Normal Boiling Points of Acyclic Ethers, Peroxides, Acetals, and Their sulfur Analogues. *J. Chem. Inf. Comput. Sci.* **1992**, 32, 237–244.
- (29) Huang, F.; Liu, X. *Aldehydes* In *Encyclopedia of Chemical Industry* (in Chinese); Chemical Industry Press: Beijing, 1997; Vol. 13. Huang, F.; Liu, X. *Ketones* In *Encyclopedia of Chemical Industry* (in Chinese); Chemical Industry Press: Beijing, 1997; Vol. 16.
- (30) Lučić, B.; Trinajstić, N. Multivariate Regression Outperforms Several Robust Architectures of Neural Networks in QSAR Modeling. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 121–132.
- (31) 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 Times and Response Factors. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 610–621.
- (32) Lučić, B.; Amić, D.; Trinajstić, N. Nonlinear Multivariate Regression Outperforms Several Concisely Designed Neural Networks in QSPR Modeling. *J. Chem. Inf. Comput. Sci.* **2000**, 40, 403–413.
- (33) Xu, L. *Chemometrical Method* (in Chinese); Scientific Press of China: Beijing, 1996.
- (34) Lučić, B.; Lukovits, I.; Nikolić, S.; Trinajstić, N. Distance-Related Indexes in the Quantitative Structure–Property Relationship Modeling. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 527–535.
- (35) Firpo, M.; Gavernet, L.; Castro, E. A.; Toropov, A. A. Maximum Topological Distances Based Indices as Molecular Descriptors for QSPR. Part 1. Application to Alkyl Benzenes Boiling Points. *J. Mol. Struct. (Theochem)* **2000**, 501/502, 419–425.
- (36) Castro, E. A.; Tueros, M.; Toropov, A. A. Maximum Topological Distances Based Indices as Molecular Descriptors for QSPR. 2—Application to Aromatic Hydrocarbons. *Comput. Chem.* **2000**, 24, 571–576.
- (37) Estrada, E.; Molina, E. Novel Local (fragment-based) Topological Molecular Descriptors for QSPR/QSAR and Molecular Design. *J. Mol. Grap. Mod.* **2001**, 20, 54–64.
- (38) Mihalić, Z.; Trinajstić, N. A Graph-Theoretical Approach to Structure–Property Relationships. *J. Chem. Educ.* **1992**, 69, 701–712.
- (39) Ivanciuc, O.; Ivanciuc, T.; Balaban, A. T. Quantitative Structure–Property Relationship Study of Normal Boiling Points for Halogen/Oxygen/Sulfur-Containing Organic Compounds Using the CODESSA Program. *Tetrahedron* **1998**, 54, 9129–9142.
- (40) Needham, D. E.; Wei, I.-C.; Seybold, P. G. Molecular Modeling of the Physical Properties of the Alkanes. *J. Am. Chem. Soc.* **1988**, 110, 4186–4194.
- (41) Pitzer, K. S.; Lippmann, D. Z.; Curl, R. F.; Huggins, C. M.; Petersen, D. E. The Volumetric and Thermodynamic Properties of Fluids. Compressibility Factor, Vapor Pressure and Entropy of Vaporization. *J. Am. Chem. Soc.* **1955**, 77, 3433–3400.
- (42) Lin, Z.; Xu, J.; Liu, S.; Zhen, X.; Li, Z. Study On Quantitative Structure–Property Relationship of Chain Hydrocarbons, Aldehydes and Alkanones by Molecular Distance-edge Vector. *Acta Phys.-Chim. Sinica* (in Chinese) **2000**, 16, 153–161.
- (43) Amboni, R. D. M. C.; Junkes, B. S.; Yunes, R. A.; Heinzen, V. E. F. Quantitative Structure–Property Relationship Study of Chromatographic Retention Indices and Normal Boiling Points for Oxo Compounds using Semiempirical Topological Method. *J. Mol. Struct. (THEOCHEM)* **2002**, 586, 71–80.
- (44) Balaban, A. T.; Mills, D.; Basak, S. C. Correlation between Structure and Normal Boiling Points of Acyclic Carbonyl Compounds. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 758–764.