

## Atomic-Level-Based AI Topological Descriptors for Structure–Property Correlations

Biye Ren\*

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

Received August 19, 2002

Multiple linear regression (MLR) analysis is used to construct the structure-boiling point models for 71 sulfur-containing organic compounds in terms of the Xu index and atomic level AI indices. The potential of these descriptors is further verified by three high quality QSPR models obtained for two subsets of compounds and a combined set of all compounds. For these subsets, containing respectively 45 sulfides and 26 thiols, the best three-parameter models are obtained, and the best four-variable model is obtained for the whole data set of 71 compounds. The correlation coefficients  $r$  are larger than 0.997 in all three final models. The standard errors  $s$  are 3.14, 2.48, and 3.48 °C for the sulfide subset, the thiol subset, and the whole data set, respectively. Furthermore, the results indicate that the boiling points are dominated by the molecular size, but some atomic types in a molecule are important due to interactions between atomic groups of the molecules. Both the molecular size and atomic types related to different fundamental interactions provide the separate contributions to boiling points. Finally, the three final models are further validated to be statistically significant and reliable by the leave-one-out cross-validation method.

## 1. INTRODUCTION

In general, there are causal relationships between the chemical structures of molecules and their physical properties and biological activities. By constructing the quantitative structure–property/activity relationships (QSPR/QSAR) one would be able to predict the properties of any molecule and also obtain more profound insights into the different interactions present in any system from a theoretical viewpoint. A variety of QSPR/QSAR models has been studied using various model parameters including various known physicochemical parameters such as the multiple solvatochromic parameters,  $n$ -octanol/water partition coefficient ( $\log P$ ), and other molecular descriptors such as geometric, electronic or electrostatic, polar, steric, and graph-theoretical topological indices.<sup>1,2</sup> Among these descriptors, the recently developed three-dimensional (3D) descriptors are particularly interesting because they take into account the geometric conformation and the nature of the bonding of groups in a molecule compared with the two-dimensional (2D) descriptors.<sup>3–5</sup>

In past years, graph-theoretical topological indices have received considerable attention because they can be derived directly from the molecular structures without any experimental effort.<sup>6–10</sup> Hence, the topological index approach to QSPR/QSAR studies represents simple and straightforward means from a viewpoint of molecular design. However, the 3D and 2D topological indices are usually highly correlated with each other and contain similar structural information to some extent.<sup>11</sup> In some cases, the 2D topological indices may produce better results than the 3D indices.<sup>12</sup> Up to data, there are mainly two types of topological indices.

The first type including various conventional 2D and 3D topological indices characterizes a molecule as a whole, i.e., molecular size or shape, such as the well-defined molecular connectivity index ( $\chi$ ),<sup>6,7</sup> Hosoya's index ( $Z$ ),<sup>13</sup> Balaban's index ( $J$ ),<sup>14</sup> Bonchev's index ( $I_D$ ),<sup>15</sup> Schulz's index ( $MTI$ ),<sup>16</sup> Wiener's index ( $W$ ),<sup>17</sup> and the recently introduced Xu index.<sup>18–21</sup> Great developments have been achieved in this field, such as differentiation of heteroatoms and multiple bonds as well as the 3D information of molecules.<sup>22–26</sup> However, these conventional indices are being faced with some challenges because the QSPR/QSAR models, using a single or few conventional topological indices, cannot reflect the role of individual atomic types or groups in a molecule.

In recent years, focus has been turned to the second type of topological indices, i.e., the atomic-level-based topological indices. In contrast to the above-mentioned conventional indices, the atomic level topological indices characterize the structural environment of each atom type in a molecule and offer the possibility of understanding the role of individual atomic types or groups in a molecule, particularly special functional groups such as  $-\text{OH}$ ,  $-\text{COOH}$ ,  $-\text{NH}_2$ , etc., for polar or polar protic compounds. One of the most interesting indices of this type is the electrotopological state (E-state) index introduced by Kier and Hall.<sup>27</sup> The E-state indices have been successfully used in a variety of QSPR/QSAR studies.<sup>28–30</sup> However, the development of the atomic-level topological indices is not very advanced. This motivates us to find new atomic level topological indices to model different physical properties and activities.

Recently, a new type of atom-type-based AI topological index was proposed in our laboratory.<sup>31–34</sup> The AI index derived from the adjacency matrix and distance matrix of a graph has been shown to be very effective in modeling the

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

normal boiling points (*BP*), molar volumes (*MV*), molar refractions (*MR*), Pitzer's eccentric factors ( $\omega$ ), heats of vaporization (*HV*), and van der Waals' constants (*b*) for C1–C9 alkanes;<sup>31</sup> several representative physical properties (*BP*, *MV*, *MR*, molecular total surface areas, aqueous solubility, *n*-octanol/water partition coefficient) and biological activities (narcosis activity and toxicity) for several data sets of alcohols with a wide range of non-hydrogen atoms;<sup>32,33</sup> and three physical properties (*MV*, *MR*, and molecular total surface areas) for a mixed set of alkanes and alcohols.<sup>34</sup> Of course, the potential of these indices in QSPR/QSAR has to be further evaluated for different systems.

In the present study, we select a mixed data set of 71 compounds containing 45 sulfides and 26 thiols with a wide range of non-hydrogen atoms to construct the structure-boiling point relationships. The data set is analyzed from two approaches. First, the overall data set of 71 compounds are divided into two subsets containing respectively 45 sulfides and 26 thiols, which are analyzed individually; then, we further deal with the structure-boiling point correlations for the whole data set of 71 compounds; and finally, we try to illustrate the role of molecule size and individual atomic types (groups) and see what structural features are important to the boiling points.

## 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 any row or column  $i$  of matrix  $\mathbf{A}$  yields the local vertex-degree  $v_i$ . Analogously, the sum over any row or column  $i$  of matrix  $\mathbf{D}$  yields distance sums  $s_i$ . The Xu index was defined as<sup>18–21</sup>

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

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

For any atom  $i$  that belongs to  $j$ th atomic type in a graph  $G$ , the corresponding topological index value,  $AI_i(j)$ , was defined as<sup>31–34</sup>

$$AI_i(j) = 1 + \phi_i(j) \quad (2)$$

with

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

where the parameter  $\phi$  is considered as a perturbing term reflecting the effects of the structural environment of the  $i$ th atom on its  $AI_i(j)$  value.

According to the above definition, for any atom type  $j$  in a graph  $G$ , the corresponding AI index,  $AI(j)$ , is the

sum of all  $AI_i(j)$  values

$$AI(j) = \sum_{i=1}^m AI_i(j) = m + \sum_{i=1}^m \phi_i(j) = m + \sum_{i=1}^m v_i(j) s_i^2(j) / \sum_{i=1}^n v_i s_i \quad (4)$$

where  $m$  is the count of the atoms or groups of the same type. Clearly, the AI value is equal to the count of atomic groups of the same type plus total perturbation terms.

To modify these indices to complex compounds, a new type of vertex degree,  $v^m$ , was proposed to differentiate heteroatoms and carbon atoms with multiple bonds in a graph  $G$ <sup>32–34</sup>

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

with

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

where  $\delta$  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  $\delta^v$  is the valence connectivity of Kier–Hall.<sup>6,7</sup> For heteroatoms in a graph  $G$ , the  $\delta^v$  value is expressed as

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

The  $\delta^v$  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 and  $Z$  and  $Z^v$  are the atomic number and the number of valence electrons for the heteroatom, respectively.

Consequently, for any molecular graph, both the Xu and AI indices can be expressed with the same formula defined above (eq 1 and eq 4) by using the proposed vertex degree  $v^m$  instead of the local vertex degree  $v_i$ .

## 3. MULTIPLE LINEAR REGRESSION ANALYSIS

**Regression Analysis.** For the whole data set and two individual subsets the multiple linear regression (MLR) using the modified Xu (represented as  $X_u^m$ ) and all atom-type AI indices present in molecules is performed to develop the final structure-boiling point models. The final model is obtained in the form of 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 modified Xu index, and  $b_j$  is the contribution coefficient of  $j$ th atomic type (group). Each coefficient describes the sensitivity of a property or activity to each of the individual indices, so the coefficients of these parameters would measure the relative importance of each index. As indices are added and removed, the changes in the statistics from model to model can be monitored. Therefore, the significance of each index can be evaluated by monitoring the statistics

such as *t*-test values (*t*-values) and Fisher ratios (*F* values) so as to choose a high quality subset of indices.<sup>35</sup> The standard error (*s*) is used to evaluate the quality of the models.

**Model Validation.** The cross-validation is a practical and reliable method for testing the significance of a model. Hence, the cross-validation using a leave-one-out method is used to test the validity of all final models obtained. The predictive ability of the model is also quantified in terms of the corresponding leave-one-out cross-validated parameters,  $r_{cv}^2$  and  $s_{cv}$  values, which are defined as<sup>36</sup>

$$r_{cv}^2 = 1.0 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (10)$$

where  $y_i$  and  $\hat{y}_i$  are the experimental and predicted value, respectively.  $\bar{y}$  is the mean value of  $y_i$

$$s_{cv} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{N - M - 1}} \quad (11)$$

where *N* is the number of samples used for model building. *M* is the number of descriptors.

#### 4. RESULTS AND DISCUSSION

In general, there are two different directions in which multiple regression analysis is usually used in QSPR/QSAR studies.<sup>37,38</sup> The first approach is based on a large number of compounds with a wide range of structural types; the second way only deals with a smaller group of structurally related compounds. Both approaches have their merits and demerits, and they serve different purposes. In this study, we select a mixed data set of 71 structurally related compounds composed of 45 sulfides and 26 thiols. First, we will consider the individual correlations for two data subsets and then extend the study to the whole data set.

**Sulfide Subset.** Recently, Randic and Basak<sup>39</sup> dealt with the structure-boiling point correlations using the variable connectivity index  ${}^1\chi^f$  for sulfides and reported a good quadratic regression. Of course, the compounds used to generate the structure-boiling point correlations were limited to a smaller set of 21 sulfides. Zefirov et al.<sup>40</sup> further analyzed the relationships between boiling points and chemical structures using a logically preselected solvation index for the same data set and obtained a very good one-parameter regression. In this section we select a data subset of 45 sulfides to construct the structure-boiling point models. The majority of experimental *BP* values for 45 sulfides are available in ref 41, but some are taken from refs 42–45. In the majority of the cases, the data among these sources agree fairly well, but when a compound has a boiling point range of 1–4 °C we quote their arithmetic mean *BP*. When a compound has different boiling points reported in the literature, we use chemical intuition to make a choice between the divergent values. For example, there are two

tabulated boiling points for ethyl heptyl sulfide, i.e., 195 °C in ref 41 and 218 °C in ref 44, differing by 23 °C. Balaban et al.<sup>41</sup> in their database quoted 195 °C for ethyl heptyl sulfide, but they found that ethyl heptyl sulfide was an outlier in structure-boiling point correlations and suggested that its experimental accuracy was subject to doubt. In fact, one would not expect the boiling point of ethyl heptyl sulfide to be the same as that of heptyl methyl sulfide (195 °C). Hence, in our database we quote 218 °C as the experimental *BP* of ethyl heptyl sulfide because a *BP* of 218 °C might be more possible for ethyl heptyl sulfide compared with methyl octyl sulfide with the same number of non-hydrogen atoms (218 °C).

The boiling point model for 45 sulfides is then developed using the modified Xu index and all five AI indices present in molecules according to the above-mentioned procedure. As an illustration, we give below the best three-parameter model (eq 12):

$$BP = 9.8981 (\pm 2.0619) + 43.2605 (\pm 0.9908) X_u^m + \\ 2.3412 (\pm 0.6162) AI(-S-) - 1.7992 (\pm 0.2665) AI(-CH_3) \quad (12)$$

$$r = 0.9975; s = 3.14 \text{ }^\circ\text{C}; F = 2778; \\ P < 0.0001; \text{ and } N = 45$$

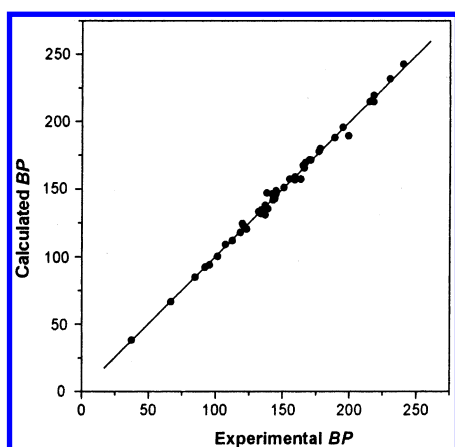
The *t*-values are 4.800, 43.66, 3.800, and –6.750, respectively. The standard error *s* is only 3.14 °C. This model explains more than 99% of the variance in the experimental *BP* values for 45 sulfides. The indices in the final model are not highly correlated with each other, and each coefficient is highly significant. The calculated values and residuals for 45 compounds, along with the topological indices used to generate the final model, are shown in Table 1. A comparison of calculated and experimental *BP* data is shown in Figure 1. One can observe from Table 1 and Figure 1 that the agreement between calculated and experimental data is excellent, and the plot shows no obviously observable pattern. According to the statement on the quality of QSPR models proposed by Mihalic and Trinajstić<sup>46</sup> (e.g.,  $r > 0.995$  and  $s < 5.0$  °C for boiling points), eq 12 represents an excellent structure-boiling point model judging from the statistical parameters and the plot in Figure 1.

We note that there are only two compounds (nos. 27 and 41) with slightly large absolute residuals ( $\geq 9.0$  °C). It is difficult to propose a reason for this and assess the reliability of data from different sources. However, if we discard the two compounds from the database, an obviously improved model using the same three indices for the remaining 43 sulfides can be obtained ( $r = 0.9986$  and  $s = 2.36$  °C).

In 1992, Balaban et al.<sup>41</sup> reported the two best structure-boiling point models in terms of topological indices for the same series of 44 sulfides with the exception of dimethyl sulfide. The first model based on the first-order connectivity  ${}^1\chi$  index and Balaban's modified  $J_{het}$  shape index produced a correlation coefficient of  $r = 0.994$  ( $r^2 = 0.989$ ) and a standard error of  $s = 4.28$  °C. The second model with the electrotopological state of sulfur atom ( $ES_s$ ), number of methyl groups ( $N_{Me}$ ), and Kier's first-order shape index ( $K\alpha_1$ ) produced practically the same correlation ( $r = 0.993$  and  $s = 4.3$  °C) as the first model. In 1998, Ivanciuc et al.<sup>47</sup>

**Table 1.** Calculated and Experimental *BP* Values for Sulfide Subset along with Topological Indices

no.	compounds	$X_u^m$	AI (–S–)	AI (–CH <sub>3</sub> )	<i>BP</i> (°C)		
					exp	calcd	res
1	dimethyl sulfide	0.6952	1.9603	3.5595	37.3	38.2	–0.9
2	ethyl methyl sulfide	1.3574	2.4263	4.3163	66.7	66.5	0.2
3	methyl propyl sulfide	1.9865	3.0762	5.0582	95.5	93.9	1.6
4	diethyl sulfide	1.9786	2.5436	5.0947	92.1	92.3	–0.2
5	isopropyl methyl sulfide	1.8583	2.7617	6.6909	84.8	84.7	0.1
6	ethyl isopropyl sulfide	2.4613	2.7025	7.6463	107.4	108.9	–1.5
7	butyl methyl sulfide	2.5894	3.8299	5.7981	123.2	120.4	2.8
8	isobutyl methyl sulfide	2.4734	3.6213	7.5786	112.5	111.7	0.8
9	ethyl propyl sulfide	2.5787	2.9194	5.8481	118.5	117.8	0.7
10	<i>t</i> -butyl methyl sulfide	2.2990	3.0114	9.0358	101.5	100.1	1.4
11	methyl pentyl sulfide	3.1702	4.6503	6.5385	145.0	146.2	–1.2
12	butyl ethyl sulfide	3.1585	3.4388	6.5932	144.2	142.7	1.5
13	dipropyl sulfide	3.1557	3.0864	6.6117	142.8	141.7	1.1
14	isopropyl propyl sulfide	3.0513	2.9216	8.6052	132.0	133.2	–1.2
15	ethyl isobutyl sulfide	3.0544	3.2769	8.5805	134.2	134.3	–0.1
16	isopentyl methyl sulfide	3.0677	4.5112	8.5079	137.0	137.9	–0.9
17	methyl 2-methylbutyl sulfide	3.0177	4.2176	8.3912	139.0	135.2	3.8
18	<i>s</i> -butyl ethyl sulfide	3.0047	3.0140	8.4661	133.6	131.7	1.9
19	<i>t</i> -butyl ethyl sulfide	2.8896	2.8509	10.0918	120.4	123.4	–3.0
20	diisopropyl sulfide	2.9342	2.7570	10.4929	120.0	124.4	–4.4
21	1-ethylpropyl methyl sulfide	2.9554	3.5056	8.4442	137.0	130.8	6.2
22	hexyl methyl sulfide	3.7321	5.5163	7.2801	171.0	171.2	–0.2
23	butyl propyl sulfide	3.7160	3.4266	7.3637	166.0	165.4	0.6
24	isobutyl propyl sulfide	3.6234	3.2959	9.5804	155.0	157.1	–2.1
25	isobutyl isopropyl sulfide	3.5209	3.1656	11.7195	145.0	148.5	–3.5
26	ethyl 2-methylbutyl sulfide	3.5772	3.7248	9.3456	159.0	156.6	2.4
27	<i>t</i> -butyl propyl sulfide	3.4735	2.9675	11.1758	138.0	147.0	–9.0
28	<i>s</i> -butyl isopropyl sulfide	3.4662	2.9675	11.4447	142.0	146.2	–4.2
29	ethyl isopentyl sulfide	3.6281	3.9323	9.5432	159.0	158.9	0.1
30	butyl isopropyl sulfide	3.6234	3.2959	9.5804	163.5	157.1	6.4
31	dibutyl sulfide	4.2595	3.6177	8.1216	188.9	188.0	0.9
32	diisobutyl sulfide	4.0868	3.4013	12.9976	170.0	171.3	–1.3
33	butyl isobutyl sulfide	4.1770	3.5093	10.5881	178.0	179.8	–1.8
34	di- <i>tert</i> -butyl sulfide	3.7690	2.8677	15.9865	151.0	150.9	0.1
35	di- <i>s</i> -butyl sulfide	3.9787	3.0843	12.3271	165.0	167.1	–2.1
36	butyl <i>s</i> -butyl sulfide	4.1276	3.3510	10.3139	177.0	177.7	–0.7
37	<i>s</i> -butyl isobutyl sulfide	4.0336	3.2425	12.6613	167.0	169.2	–2.2
38	heptyl methyl sulfide	4.2775	6.4150	8.0229	195.0	195.5	–0.5
39	ethyl heptyl sulfide	4.7969	5.4838	8.8193	218.0	214.4	3.6
40	methyl octyl sulfide	4.8082	7.3381	8.7666	218.0	219.3	–1.3
41	di- <i>t</i> -pentyl sulfide	4.7286	3.0936	17.9716	199.0	189.4	9.6
42	dipentyl sulfide	5.3059	4.1444	9.6280	230.0	231.8	–1.8
43	diisopentyl sulfide	5.1681	3.9858	15.6330	215.0	214.7	0.3
44	isobutyl 4-methylpentyl sulfide	5.1690	4.1995	15.6206	216.0	215.2	0.8
45	methyl nonyl sulfide	5.3258	8.2796	9.5112	240.0	242.6	–2.6

**Figure 1.** A plot of calculated versus experimental *BP* (°C) for the sulfide subset.

further analyzed the relationships between boiling points and chemical structures using the CODESSA (Comprehensive Descriptors for Structural and Statistical Analysis) program

for the same 45 sulfides. They obtained some excellent structure-boiling point correlations. The best four-, five-, and six-variable models gave  $r = 0.997, 0.998, 0.998$  and  $s = 3.4, 3.2,$  and  $3.0$  °C, respectively. Our results are obviously comparable to these best models reported by Ivanciuc et al.<sup>47</sup> for the same data set, although there are several discrepancies between the tabulated experimental *BP* data.

**Thiol Subset.** To our knowledge, very few studies on the structure-boiling point correlations for thiols were reported in the literature using graph theoretic topological indices. This motivates us to develop the structure-boiling point models for the thiol subset. The experimental *BP* data for 26 compounds are taken from refs 42–45. It should be mentioned that there are some discrepancies between these data sources. In this case, the chemical intuition is used to make a choice between the divergent values. Table 2 shows the compounds and the corresponding *BP* values, along with the topological indices used to generate the final model. This model for 26 thiols is developed using the modified Xu index and five AI indices according to the aforementioned proce-



**Table 2.** Calculated and Experimental *BP* Values for Thiol Subset along with Topological Indices

no.	compounds	$X_u^m$	AI ( $-\text{CH}_3$ )	AI ( $>\text{CH}_2$ )	BP ( $^{\circ}\text{C}$ )		
					exp	calcd	res
1	methyl mercaptan	0	1.3609	0	6.0	1.1	4.9
2	ethyl mercaptan	0.7402	1.7302	1.6497	36.5	38.1	-1.6
3	1-propyl mercaptan	1.4013	2.1034	3.9616	67.5	70.2	-2.7
4	2-propyl mercaptan	1.2777	3.7950	0	59.0	58.2	0.8
5	1-butyl mercaptan	2.0255	2.4769	6.9581	98.5	99.6	-1.1
6	2-methyl-1-propyl mercaptan	1.9034	4.1717	2.2216	88.5	88.4	0.1
7	1-methyl-2-propyl mercaptan	1.8853	4.4928	2.2379	84.5	85.7	-1.2
8	2-methyl-2-propyl mercaptan	1.7446	5.9759	0	66.0	72.2	-6.2
9	1-pentyl mercaptan	2.6233	2.8509	10.6466	126.6	127.0	-0.4
10	2-pentyl mercaptan	2.4914	5.1591	5.0476	112.9	112.7	0.2
11	3-pentyl mercaptan	2.4483	5.2872	5.1511	105.0	109.6	-4.6
12	2-methyl-1-butyl mercaptan	2.4698	4.7878	5.1147	117.0	113.4	3.6
13	2-methyl-2-butyl mercaptan	2.3184	6.8346	2.4268	99.1	96.9	2.2
14	3-methyl-1-butyl mercaptan	2.5146	4.6367	5.0029	118.0	116.8	1.2
15	1-hexyl mercaptan	3.1998	3.2251	15.0307	152.7	152.5	0.2
16	2-hexyl mercaptan	3.0802	5.8356	8.4555	142.0	138.3	3.7
17	1-heptyl mercaptan	3.7581	3.5996	20.1120	176.9	176.5	0.4
18	1-octyl mercaptan	4.3005	3.9741	25.8913	199.0	198.9	0.1
19	2-octyl mercaptan	4.2040	7.2076	17.1257	186.4	184.6	1.8
20	1-nonyl mercaptan	4.8287	4.3488	32.3693	219.8	219.9	-0.1
21	1-decyl mercaptan	5.3442	4.7235	38.5463	240.6	240.5	0.1
22	1-undecyl mercaptan	5.8482	5.0983	47.4226	257.4	258.0	-0.6
23	1-dodecyl mercaptan	6.3417	5.4731	55.9984	274.6	275.3	-0.7
24	1-tridecyl mercaptan	6.8254	5.8479	65.2737	290.8	291.3	-0.5
25	1-tetradecyl mercaptan	7.3002	6.2228	75.2486	306.2	306.2	0
26	1-pentadecyl mercaptan	7.7667	6.5976	85.9233	320.7	320.1	0.6

dure. As an illustration, we give below the best three-parameter model (eq 13):

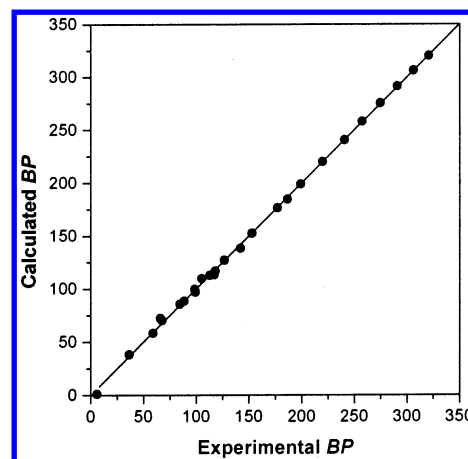
$$BP = 8.2334 (\pm 1.6974) + 54.6482 (\pm 1.0408) X_u^m - 5.2439 (\pm 0.4506) \text{AI} (-\text{CH}_3) - 0.9076 (\pm 0.07765) \text{AI} (>\text{CH}_2) \quad (13)$$

$$r = 0.9996; s = 2.48^{\circ}\text{C}; F = 10459; P < 0.0001; \text{ and } N = 71$$

The *t*-values are 4.850, 52.50, -11.64, and -11.69, respectively. One observes that the correlation coefficient *r* is larger than 0.999 and the standard error *s* is only 2.48  $^{\circ}\text{C}$ . The model explains more than 99.9% of the variance in experimental *BP* data for 26 thiols. The indices in eq 13 are not also highly intercorrelated, and each coefficient is clearly highly significant. The larger correlation coefficient and smaller standard error indicate that the quality of the final model is excellent according to the above-mentioned criteria. The calculated *BP* values and residuals for 26 thiols are shown in Table 2. A plot of calculated versus experimental data for *BP* is shown in Figure 2. One observes that the calculated values agree fairly well with experimental data, and the plot shows no obviously observable pattern.

There is only one compound with a slightly large residual (no. 8, 2-methyl-2-propyl mercaptan, -6.2  $^{\circ}\text{C}$ ). However, if we omit the compound, the quality of the model can be further improved. The standard error *s* is significantly reduced to only 2.03  $^{\circ}\text{C}$ , which approximates the magnitude of the experimental error (1–1.5  $^{\circ}\text{C}$ ).

**Whole Data Set.** In order to evaluate the applicability of these indices in more complex systems, as an extension of above studies, we further deal with the structure-boiling point correlations for a mixed data set of 45 sulfides and 26 thiols. The final four-parameter model for the whole data set is

**Figure 2.** A plot of calculated versus experimental *BP* ( $^{\circ}\text{C}$ ) for the thiol subset.

developed and is shown as follows (eq 14):

$$BP = 7.1237 (\pm 1.4422) + 51.8897 (\pm 1.1876) X_u^m + 1.9958 (\pm 0.2484) \text{AI} (-\text{S}-) - 3.7760 (\pm 0.2834) \text{AI} (-\text{CH}_3) - 0.7208 (\pm 0.09361) \text{AI} (>\text{CH}_2) \quad (14)$$

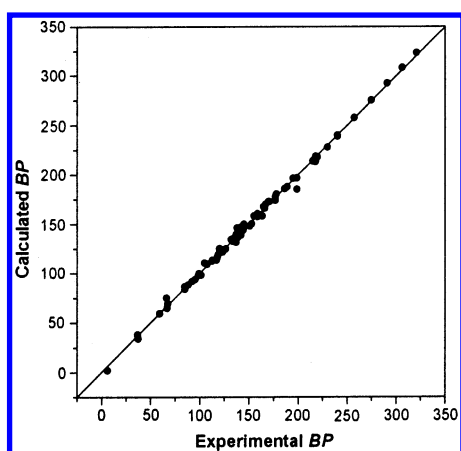
$$r = 0.9986; s = 3.40^{\circ}\text{C}; F = 5931; P < 0.0001; \text{ and } N = 71$$

The *t*-values are 4.939, 43.69, 8.036, -13.32, and -7.700, respectively. The standard error *s* is only 3.40  $^{\circ}\text{C}$ , which is only slightly higher than those of the separate subsets. This model explains more than 99% of the variance in the experimental *BP* values for this combined set of all compounds. The indices in the final model are not highly correlated with one another, and each coefficient is obviously highly significant. The calculated values and residuals for 71 compounds are shown in Table 3. A comparison of

**Table 3.** Calculated and Experimental *BP* Values for a Combined Set of Sulfides and Thiols

no.	compounds	<i>BP</i> (°C)					no.	compounds	<i>BP</i> (°C)				
		exp	calcd	res	pred <sub>cv</sub> <sup>a</sup>	res <sub>cv</sub> <sup>b</sup>			exp	calcd	res	pred <sub>cv</sub> <sup>a</sup>	res <sub>cv</sub> <sup>b</sup>
1	dimethyl sulfide	37.3	33.7	3.6	33.1	4.2	37	<i>s</i> -butyl isobutyl sulfide	167.0	170.2	-3.2	170.4	-3.4
2	ethyl methyl sulfide	66.7	64.6	2.1	64.5	2.2	38	heptyl methyl sulfide	195.0	196.3	-1.3	196.4	-1.4
3	methyl propyl sulfide	95.5	93.9	1.6	93.8	1.7	39	ethyl heptyl sulfide	218.0	213.1	4.9	212.7	5.3
4	diethyl sulfide	92.1	92.0	0.1	92.0	0.1	40	methyl octyl sulfide	218.0	218.7	-0.7	218.8	-0.8
5	isopropyl methyl sulfide	84.8	83.8	1.0	83.8	1.0	41	di- <i>t</i> -pentyl sulfide	199.0	185.1	13.9	181.6	17.4
6	ethyl isopropyl sulfide	107.4	109.2	-1.8	109.3	-1.9	42	dipentyl sulfide	230.0	227.7	2.3	227.5	2.5
7	butyl methyl sulfide	123.2	121.6	1.6	121.6	1.6	43	diisopentyl sulfide	215.0	213.9	1.1	213.8	1.2
8	isobutyl methyl sulfide	112.5	112.5	0	112.5	0	44	isobutyl 4-methylpentyl sulfide	216.0	214.6	1.4	214.4	1.6
9	ethyl propyl sulfide	118.5	118.5	0	118.5	0	45	methyl nonyl sulfide	240.0	239.9	0.1	239.9	0.1
10	<i>t</i> -butyl methyl sulfide	101.5	98.3	3.2	98.2	3.3	46	methyl mercaptan	6.0	2.0	4.0	0.8	5.2
11	methyl pentyl sulfide	145.0	147.8	-2.8	148.0	-3.0	47	ethyl mercaptan	36.5	37.8	-1.3	38.0	-1.5
12	butyl ethyl sulfide	144.2	143.9	0.3	143.9	0.3	48	1-propyl mercaptan	67.5	69.0	-1.5	69.1	-1.6
13	dipropyl sulfide	142.8	142.8	0	142.8	0	49	2-propyl mercaptan	59.0	59.1	-0.1	59.1	-0.1
14	isopropyl propyl sulfide	132.0	134.3	-2.3	134.3	-2.3	50	1-octyl mercaptan	98.5	97.8	0.7	97.8	0.7
15	ethyl isobutyl sulfide	134.2	135.5	-1.3	135.5	-1.3	51	2-methyl-1-propyl mercaptan	88.5	88.5	0	88.5	0
16	isopentyl methyl sulfide	137.0	139.6	-2.6	139.7	-2.7	52	1-methyl-2-propyl mercaptan	84.5	86.4	-1.9	86.4	-1.9
17	methyl 2-methylbutyl sulfide	139.0	136.5	2.5	136.4	2.6	53	2-methyl-2-propyl mercaptan	66.0	75.1	-9.1	75.6	-9.6
18	<i>s</i> -butyl ethyl sulfide	133.6	132.5	1.1	132.5	1.1	54	1-pentyl mercaptan	126.6	124.8	1.8	124.7	1.9
19	<i>t</i> -butyl ethyl sulfide	120.4	122.3	-1.9	122.4	-2.0	55	2-pentyl mercaptan	112.9	113.3	-0.4	113.3	-0.4
20	diisopropyl sulfide	120.0	125.2	-5.2	125.4	-5.4	56	3-pentyl mercaptan	105.0	110.5	-5.5	110.7	-5.7
21	1-ethylpropyl methyl sulfide	137.0	131.5	5.5	131.4	5.6	57	2-methyl-1-butyl mercaptan	117.0	113.5	3.5	113.3	3.7
22	hexyl methyl sulfide	171.0	172.7	-1.7	172.9	-1.9	58	2-methyl-2-butyl mercaptan	99.1	99.9	-0.8	99.9	-0.8
23	butyl propyl sulfide	166.0	166.1	-0.1	166.1	-0.1	59	3-methyl-1-butyl mercaptan	118.0	116.5	1.5	116.4	1.6
24	isobutyl propyl sulfide	155.0	158.3	-3.3	158.4	-3.4	60	1-hexyl mercaptan	152.7	150.2	2.5	149.9	2.8
25	isobutyl isopropyl sulfide	145.0	150.0	-5.0	150.3	-5.3	61	2-hexyl mercaptan	142.0	138.8	3.2	138.6	3.4
26	ethyl 2-methylbutyl sulfide	159.0	157.6	1.4	157.6	1.4	62	1-heptyl mercaptan	176.9	174.0	2.9	173.8	3.1
27	<i>t</i> -butyl propyl sulfide	138.0	146.1	-8.1	146.4	-8.4	63	1-octyl mercaptan	199.0	196.6	2.4	196.4	2.6
28	<i>s</i> -butyl isopropyl sulfide	142.0	147.2	-5.2	147.4	-5.4	64	2-octyl mercaptan	186.4	185.7	0.7	185.6	0.8
29	ethyl isopentyl sulfide	159.0	160.3	-1.3	160.4	-1.4	65	1-nonyl mercaptan	219.8	217.9	1.9	217.7	2.1
30	butyl isopropyl sulfide	163.5	158.3	5.2	158.2	5.3	66	1-decyl mercaptan	240.6	238.8	1.8	238.6	2.0
31	dibutyl sulfide	188.9	187.7	1.2	187.7	1.2	67	1-undecyl mercaptan	257.4	257.2	0.2	257.1	0.3
32	diisobutyl sulfide	170.0	172.7	-2.7	172.8	-2.8	68	1-dodecyl mercaptan	274.6	275.2	-0.6	275.2	-0.6
33	butyl isobutyl sulfide	178.0	180.4	-2.4	180.4	-2.4	69	1-tridecyl mercaptan	290.8	292.2	-1.4	292.4	-1.6
34	di- <i>tert</i> -butyl sulfide	151.0	148.0	3.0	147.5	3.5	70	1-tetradecyl mercaptan	306.2	308.2	-2.0	308.8	-2.6
35	di- <i>s</i> -butyl sulfide	165.0	167.6	-2.6	167.8	-2.8	71	1-pentadecyl mercaptan	320.7	323.3	-2.6	324.9	-4.2
36	butyl <i>s</i> -butyl sulfide	177.0	178.0	-1.0	178.1	-1.1							

<sup>a</sup> Predicted *BP* by leave-one-out cross-validated approach. <sup>b</sup> Predicted residuals from the cross-validation.

**Figure 3.** A plot of calculated versus experimental *BP* (°C) for the whole data set of sulfides and thiols.

calculated and experimental data for *BP* is shown in Figure 3. One can observe that calculated values agree well with experimental data and the plot shows no obviously observable pattern. According to the above-mentioned criteria, the statistical parameters indicate the excellent quality of the correlation.

There is only one compound (no. 41, i.e., di-*t*-pentyl sulfide) with a slightly large residual ( $>10$  °C), which is the same as that in the sulfide subset correlation (eq 12).

The reason for this is not clear. It is, of course, difficult to state that its boiling point might be in error. In fact, after discarding this compound from the database, we obtain an obviously improved model using the same four indices for the remaining 70 compounds. The correlation coefficient  $r$  is increased to 0.9990 and the standard error  $s$  is further reduced to only 2.84 °C. The significant improvement again indicates the size of the problem with the data for di-*t*-pentyl sulfide.

**Interpretation of Models.** In our previous studies,<sup>31–34</sup> the atom type AI indices were shown to be particularly effective in the system with a limited number of atomic types in molecules and also offer a new approach to understand the role of individual atomic types (groups) in a molecule. From the final models obtained using these indices we can obtain more meaningful insights into the role of chemical structures in molecules and intermolecular interactions.

We simply estimate the relative importance of the molecule size and atom type indices by their relative or fraction contributions to boiling points. The relative contribution of each index is estimated by multiplying the coefficients by the mean index values; fraction contributions are calculated by multiplying the absolute values by the coefficient of determination ( $r^2$ ), i.e., the square of the multiple correlation coefficient ( $r$ ), and dividing by the sum of the absolute values.<sup>48</sup> The results are shown in Table 4. First, let us focus

**Table 4.** Relative and Fraction Contributions for Individual Topological Indices

models	$X_u^m$		AI ( $-\text{CH}_3$ )		AI ( $-\text{S}-$ )		AI ( $>\text{CH}_2$ )	
	relative	fraction	relative	fraction	relative	fraction	relative	fraction
sulfide subset	148.2	85.0%	-16.63	9.5%	8.60	4.9%		
thiol subset	186.3	81.4%	-23.76	10.4%			-18.77	8.2%
overall data set	177.4	80.2%	-28.38	10.6%	4.65	2.1%	-10.05	4.5%

**Table 5.** Statistical Parameters of Models with One-, Two-, Three-, and Four Indices

data set	parameters	$r$	$s$ ( $^{\circ}\text{C}$ )	F	improvement (%) <sup>a</sup>
sulfide subset	$X_u^m$	0.9884	6.66		
	$X_u^m, \text{AI}(-\text{CH}_3)$	0.9967	3.61	3152	45.8
	$X_u^m, \text{AI}(-\text{CH}_3), \text{AI}(-\text{S}-)$	0.9975	3.14	2778	52.8
thiol subset	$X_u^m$	0.9968	7.22		
	$X_u^m, \text{AI}(-\text{CH}_3)$	0.9975	6.51	2265	9.8
	$X_u^m, \text{AI}(-\text{CH}_3), \text{AI}(>\text{CH}_2)$	0.9996	2.48	10459	65.6
	$X_u^m$	0.9941	6.90		
total set	$X_u^m, \text{AI}(-\text{CH}_3)$	0.9961	5.64	4304	18.3
	$X_u^m, \text{AI}(-\text{CH}_3), \text{AI}(-\text{S}-)$	0.9973	4.66	4218	32.5
	$X_u^m, \text{AI}(-\text{CH}_3), \text{AI}(-\text{S}-), \text{AI}(>\text{CH}_2)$	0.9986	3.40	5931	50.7
	$X_u^m$				

<sup>a</sup> The improvement in the quality of the models is relative to the simple linear models.

attention on the sulfide subset. We can see from Table 4 that the  $X_u^m$  index makes a major contribution (85.0%) to  $BP$  while two AI indices have smaller contributions. The major contribution from the  $X_u^m$  index indicates that the molecular size plays a dominant role in determining boiling points for sulfides, because the  $X_u$  index mainly characterizes the molecular size.<sup>18</sup> On the other hand, the contributions for AI ( $-\text{CH}_3$ ) and AI ( $-\text{S}-$ ) indices are -16.63 (9.5%) and 8.60 (4.9%), respectively. The smaller contributions from these AI indices indicate that the influence of atom types in a molecule is obvious. This may be reasonably interpreted as follows: As is well-known, the boiling point is a physical property that strongly depends on intermolecular interactions and is very much influenced by the molecular weight, while the intermolecular interactions in complex compounds, as pointed out by Pitzer et al.<sup>49</sup> in 1955, are not only interactions between molecular centers but also a sum of interactions between various parts of the molecules. According to this point of view, both the molecular size and individual atomic types or groups related to different fundamental interactions would make the separate contributions to the total physical property value of a molecule. Furthermore, the slightly large and negative contribution for the AI ( $-\text{CH}_3$ ) index may be attributed to the effects of branching of a molecule compared with the AI ( $-\text{S}-$ ) index, because the AI ( $-\text{CH}_3$ ) index is clearly related to the count of  $-\text{CH}_3$  groups in a molecule and may be interpreted physically as a parameter characterizing the degree of branching.<sup>31,32,34</sup> Clearly, branching would prevent molecules from contacting closely with neighboring molecules and cause the reduction of  $BP$  due to varying the intermolecular interaction. As a result, the AI ( $-\text{CH}_3$ ) index makes a slightly larger contribution than the AI ( $-\text{S}-$ ) index. The small and positive contribution for the AI ( $-\text{S}-$ ) index may be related to its larger atomic weight relative to carbon atoms. In addition, the other three AI indices are absent in the final correlation. It is probable that the three AI indices do not have basic structure-boiling point significance for sulfides. The results suggest that branching and heavy atoms in a molecule are also very important factors determining boiling points of sulfides.

The final model for the thiol subset is analyzed also. Analogously, the  $X_u^m$  index makes a major contribution (81.4%) to  $BP$ , indicating that the boiling points for thiols are also dominated by the molecular size. This is similar with the results obtained for the sulfide subset. On the other hand, the contributions for the AI ( $-\text{CH}_3$ ) and AI ( $-\text{CH}_2-$ ) indices are -23.76 (10.4%) and -18.77 (8.2%), respectively, indicating the influence of branching on  $BP$ . It is worthwhile to note that the AI ( $-\text{SH}$ ) index of  $-\text{SH}$  groups in the final model for thiol subset is absent. The reason for this is because the  $X_u^m$  index is highly intercorrelated with the AI ( $-\text{SH}$ ) index for thiols ( $r = 0.9856$ ).

In the case of the overall data set, the same trends are also observed. The contributions are 177.4 (80.4%), -28.38 (10.6%), 4.65 (2.1%), and -10.05 (4.5%) for  $X_u^m$ , AI ( $-\text{CH}_3$ ), AI ( $-\text{S}-$ ), and AI ( $-\text{CH}_2-$ ) indices, respectively. Clearly, the contributions of the  $X_u^m$  and AI ( $-\text{CH}_3$ ) indices for the entire data set are very close to that for the two individual subsets. Of course, the contributions for the AI ( $-\text{S}-$ ) and AI ( $-\text{CH}_2-$ ) indices are slightly altered.

The above results can be further illustrated by the stepwise regression analysis using the same indices as those in three final models. Table 5 shows the models with one, two, three, and four parameters obtained by the stepwise regression for two subsets and the combined set. Clearly, in all cases the models are significantly improved with the addition of atom type AI indices into the equations. For all three final models the improvement is more than 50% compared with the one-parameter model based on the  $X_u^m$  index alone. For example, the simple linear regression using the  $X_u^m$  index leads to  $r = 0.9941$  and  $s = 6.90$   $^{\circ}\text{C}$  for the whole data set, while the standard error in the final model is largely reduced to 3.40  $^{\circ}\text{C}$ . The reduction in the standard error is 50.7% relative to the simple model, indicating the importance of atom types in a molecule.

**Model Validation** All three models are further tested by the cross-validation using a leave-one-out method. As a quantitative evaluation of the cross-validated results, the statistical parameters ( $r$  and  $s$ ) of the final models and the



**Table 6.** Statistical Parameters of MLR and Leave-One-Out Cross-Validation for Three Final Models

data set	<i>r</i>	<i>s</i> (°C)	<i>r</i> <sub>cv</sub>	<i>s</i> <sub>cv</sub> (°C)
total data set	0.9986	3.40	0.9982	3.81
sulfide subset	0.9975	3.14	0.9969	3.54
thiol subset	0.9996	2.48	0.9994	3.18

corresponding cross-validated statistical parameters (*r*<sub>cv</sub> and *s*<sub>cv</sub>) are also listed in Table 6. It is expected that the cross-validated *s*<sub>cv</sub> should be only slightly larger than *s* of the final models. One can see that for all three final models the *r*<sub>cv</sub> values are very close to the *r* values of the final models and the *s*<sub>cv</sub> values are only slightly larger than the *s* values. This cross-validation demonstrates that all three final models are statistically significant and reliable. The predicted *BP* and residuals from the cross-validation are also shown in Table 3 for an easy comparison. On the other hand, plots of calculated versus observed values and plots of residuals versus calculated values can also be used as evidence of the validity of a model. Plots of residuals versus calculated values show that the residuals are randomly distributed. Plots of calculated versus observed data show no obviously observable pattern (Figures 1–3). Therefore, the final models are statistically significant and successfully validated.

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

## 5. CONCLUSION

Multiple linear regression (MLR) using the modified Xu and atomic level AI indices can provide high quality QSPR models of the normal boiling points for two subsets of sulfides and thiols as well as a combined set of all compounds. A combined use of the modified Xu and atomic level AI indices significantly improves the quality of three final models compared with a single Xu index. The results indicate that the boiling points for compounds studied are dominated by the molecular size. On the other hand, atomic types in molecules are important due to interactions between atomic groups of the molecules. These high quality models not only offer a better understanding of the role of each structural feature in a molecule but also illustrate the high potential of these indices in QSPR/QSAR modeling. The cross-validation verifies the final models to be statistically significant and reliable.

## ACKNOWLEDGMENT

The author is thankful to Professor A. J. Hopfinger, associate editor of this journal, and the referees for their helpful and detailed comments, which have improved the quality of the paper.

## REFERENCES AND NOTES

- Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR. Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995.
- Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR. Hydrophobic, Electronic and Steric Constants*; American Chemical Society: Washington, DC, 1995.
- Consonni, V.; Todeschini, R.; Pavan, M. Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 1. Theory of the Novel 3D Molecular Descriptors. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 682–692.
- Consonni, V.; Todeschini, R.; Pavan, M.; Gramatica, P. Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 2. Application of the Novel 3D Molecular Descriptors to QSAR/QSPR Studies. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 693–705.
- Puri, S.; Chickos, J. S.; Welsh, W. J. Three-dimensional Quantitative Structure–Property Relationship (3D-QSPR) Models for Prediction of Thermodynamic Properties of Polychlorinated biphenyls (PCBs): Enthalpy of Sublimation. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 109–116.
- Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
- Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Studies*; Research Studies Press: Letchworth, 1986.
- Trinajstić, N. *Chemical Graph Theory*, 2nd ed.; CRC Press: Boca Raton, 1992.
- Balaban, A. T. Chemical Graphs: Looking Back and Glimpsing Ahead. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 339–350.
- Randic, M.; Basak, S. C. A New Descriptor for Structure–Property and Structure–Activity Correlations. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 650–656.
- Toropov, A.; Toropova, A.; Ismailov, T.; Bonchev, D. 3D Weighting of Molecular Descriptors for QSPR/QSAR by the Method of Ideal Symmetry (MIS): 1. Application to Boiling Points of Alkanes. *J. Mol. Struct. (THEOCHEM)*, **1998**, *424*, 237–247.
- Mihalic Z.; Nikolic, S.; Trinajstić, N. Comparative Study of Molecular Descriptors Derived from the Distance Matrix. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 28–37.
- Hosoya, H. Topological Index. A Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332–2339.
- Balaban, A. T. Highly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett.* **1982**, *89*, 399–404.
- Bonchev, D.; Trinajstić, N. Information Theory, Distance Matrix, and Molecular Branching. *J. Chem. Phys.* **1977**, *67*, 4517–4533.
- Schultz, H. P. Topological Organic Chemistry, 1. Graph Theory and Topological Indices. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 227–228.
- Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17–20.
- 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 University Technol. (in Chinese)* **1999**, *27*(5), 89–95.
- Estrada, E. Edge Adjacency Relationships in Molecular Graphs Containing Heteroatoms: A New Topological Index Related to Molar Volume. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 701–707.
- 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.
- Bogdanov, B.; Nikolic, S.; Trinajstić, N. On the Three-Dimensional Wiener Number. *J. Math. Chem.* **1989**, *3*, 299–309.
- Randic, M.; Jerman-Blazic, B.; Trinajstić, N. Development of 3-Dimensional Molecular Descriptors. *Comput. Chem.* **1990**, *14*, 237–246.
- Estrada, E. Three-dimensional Molecular Descriptors Based on Electron Charge Density Weighted Graphs. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 708–713.
- Hall, L. H.; Mohny, B.; Kier, L. B. The Electrotopylogical State: Structural Information at the Atomic Level for Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 76–82.
- 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.
- 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.
- Rose, K.; Hall, L. H.; Kier, L. B. Modeling Blood-Brain Barrier Partitioning Using the Electrotopylogical State. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 651–666.



- (31) Ren, B. Application of Novel Atom-Type AI Topological Indices to QSPR Studies of Alkanes. *Comput. Chem.* **2002**, 26, 357–369.
- (32) Ren, B. Novel Atom-Type AI Topological Indices for QSPR Studies of Alkanols. *Comput. Chem.* **2002**, 26, 223–235.
- (33) Ren, B. Novel Atomic-Level-Based AI Topological Descriptors: Application to QSPR/QSAR Modeling. *J. Chem. Inf. Comput. Sci.* **2002**, 42, 858–868.
- (34) Ren, B. Application of Novel Atom-Type AI Topological Indices in the Structure–Property Correlations. *J. Mol. Struct. (THEOCHEM)* **2002**, 586, 137–148.
- (35) Lucic, 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.
- (36) Xu, L. *Chemometrical Method (in Chinese)*; Scientific Press of China: Beijing, 1996.
- (37) 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.
- (38) 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.
- (39) Randić, M.; Basak, S. Construction of High-Quality Structure–Property-Activity Regressions: The Boiling Points of Sulfides. *J. Chem. Inf. Comput. Sci.* **2000**, 40, 899–905.
- (40) Zefirov, N. S.; Palyulin, V. A. QSAR for Boiling Points of “Small” Sulfides. Are the “High-Quality Structure-Property-Activity Regressions” the Real High Quality QSAR Models? *J. Chem. Inf. Comput. Sci.* **2001**, 41, 1022–1027.
- (41) 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.
- (42) Weast, R. *CRC Handbook of Chemistry and Physics*, 70th ed.; CRC Press: Boca Raton, FL, 1989–1990.
- (43) Lide, D. R.; Milne, G. W. A. *Handbook of Data on Common Organic Compounds*; CRC Press: Boca Raton, FL, 1992.
- (44) Dean, J. A. *Lange’s Handbook of Chemistry*, 15th ed.; McGraw-Hill: Beijing, 1999.
- (45) Yaws, C. L. *Chemical Properties Handbook*; McGraw-Hill: Beijing, 1999.
- (46) Mihalic, Z.; Trinajstić, N. A Graph-Theoretical Approach to Structure–Property Relationships. *J. Chem. Educ.* **1992**, 69, 701–712.
- (47) 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.
- (48) 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.
- (49) Pitzer, K. S.; Lippmann, D. Z.; Curl, R. F.; Huggins, C. M.; Petersen, D. E. The Volumetric and Thermodynamic Properties of Fluids. II. Compressibility Factor, Vapor Pressure and Entropy of Vaporization. *J. Am. Chem. Soc.* **1955**, 77, 3433–3400.

CI020382N