

Application of Molecular Fragments Variable Connectivity Index to Predicting Boiling Points of Alcohols

Q. Chen^a, Y.-W. Kou^b, F. Pan^a and J.-Y. Yuan^{a,*}

^aCollege of Atmospheric Sciences, Lanzhou University, P. R. China

^bCollege of Earth and Environmental Science, Lanzhou University, P. R. China

(Received 28 March 2009, Accepted 8 April 2010)

Molecular fragments variable connectivity index (mfVCI) is proposed as a variable molecular descriptor. Having in mind that the molecular structure unit exerts strong effect on the boiling point, molecular fragments could be defined as the atoms or functional groups having different characteristics due to different chemical bonding. Each molecular fragment is regarded as a vertex of topological diagram endowed with variable weights to substitute for topological matrix diagonal. The quantitative structure-property relationship (QSPR) model, obtained by mfVCI, shows its desirable robustness and predictivity.

Keywords: QSPR, Molecular fragments variable connectivity index (mfVCI), Boiling point, Prediction

INTRODUCTION

Being firstly and easily measured as a property of a compound, the normal boiling point (NBP) is used to characterize and identify a new compound, and to estimate other physical properties, such as critical temperatures, vapor pressures, enthalpies of vaporization, *etc.* [1-3]. It is well-known that NBP is determined by the intermolecular interactions in the liquid and by the difference in the molecular internal partition function between gas and liquid phases at the boiling point. Consequently, the NBP depends indirectly on the chemical structure of the molecules. It is not surprising that numerous methods have been developed to estimate the NBP of a compound from its structure [1]. The quantitative structure-property relationship (QSPR) model has been quite extensively reported in the literature to predict the boiling point [4-7].

A key point in the QSPR studies is the accurate and simple

characterization of the structural features that are related to the observed property, known as the molecular descriptors [7,8]. Most topological descriptors are characterized by fixed numerical values, which are independent of the property considered [9]. Based on the idea that a variable parameter undergoes change during the regression analysis, Randić proposed the variable connectivity index [10,11], which has improved the regression results. In these descriptors, the optimal weights which represent relative contributions to individual atoms are optimized during the modeling procedure to best suit the modeled property. In the past research, the influence of the functional groups and their characteristics on chemical bonding was ignored, as the weights were calculated. In fact, they play an important role in the property.

In this work, we have tried to remedy this obvious shortcoming. The proposed molecular fragments variable connectivity index, which considers the atoms or functional groups to have different characteristics due to different chemical bonding as a chemical unit, is applied to study the boiling points of alcohols.

*Corresponding author. E-mail: chenqhq@lzu.edu.cn

THEORY

Atoms and functional groups which have different characteristics due to their different chemical bonding, exert significant effect on the boiling point of compounds. In order to explore factors which affect the boiling point of alcohols, molecular fragments were defined as the atoms or functional groups of different characteristics of chemical bonding. In this paper, they are considered as the unit of molecular structure. The different types of the molecular fragments are expressed in boldface and in italics in Table 1 for the case of 6-methyl-3-heptanol structure see (Fig. 1). To a certain property, the different fragments are endowed with different molecular fragment weights (w_i).

With the molecular fragments being regarded as the vertex of topological diagram, the augmented adjacency matrix is illustrated in Table 2. The zero diagonal elements in the augmented adjacency matrix are replaced by the variable weight (w_i) of molecular fragments. The variable weight stands for their contribution to the boiling point. The sum of the row in the augmented adjacency matrix (δ_i^{mf}) is the augmented valence of the molecular fragment.

Based on the augmented adjacency matrix, the molecular fragments variable connectivity index (mfVCI) was developed from the connectivity index [12]. The expression of the first order mfVCI (${}^1\chi^{mf}$) is as follows:

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

Table 1. Molecular Fragments for 6-Methyl-3-heptanol

Name	Diagrams and signs						
Molecular Fragments							
Expressions	CH₃(CH)	2(CH₃)CH(CH₂)	(CH)CH₂(CH₂)	2(CH₂)CH(OH)	CH₃CH₂(CH)	CH₃(CH₂)	(CH)OH
Weights	w_1	w_2	w_3	w_4	w_5	w_6	w_7

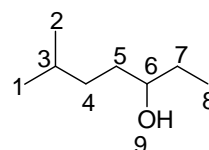


Fig. 1. Atom numbering scheme for 6-methyl-3-heptanol.

The summation is made over all edges m (the paths of length one); therefore, ${}^1\chi_j^{mf}$ represents the contribution of the adjacent vertices by one edge j to the connectivity index, and δ_i^{mf} is the sum of the row in the augmented adjacency matrix, which is called the molecular fragment valence. It can be calculated as the Randić valence of the molecular fragment (δ_i) and the corresponding diagonal weight (w_i). The contributions of molecular fragments to the mfVCI are also defined for real numbers. Therefore, the weights (w_i) have to be varied in such a way that the sum of the row in the augmented adjacency matrix remains positive ($\delta_i^{mf} = \delta_i + w_i > 0$). In this way, the weight of an individual molecular fragment varies from negative Randić delta to infinity.

The mfVCI of order one (${}^1\chi^{mf}$) was optimized by using *Solver* in Microsoft Office Excel. The target function is the summation of the root-mean-square error (RMSE) for training set. The aim of this optimal procedure was to optimize the weights of the molecular fragments that would minimize the standard error of a linear regression equation (Property = $a \times {}^1\chi^{mf} + b$).

The optimization was carried out by running the *Solver* program of *Quadratic in Estimates* to solve the nonlinear

Table 2. Augmented Adjacency Matrix of 6-Methyl-3-heptanol

	1	2	3	4	5	6	7	8	9	Row Sum
1	w_1	0	1	0	0	0	0	0	0	$1+w_1$
2	0	w_1	1	0	0	0	0	0	0	$1+w_1$
3	1	1	w_2	1	0	0	0	0	0	$3+w_2$
4	0	0	1	w_3	1	0	0	0	0	$2+w_3$
5	0	0	0	1	w_3	1	0	0	0	$2+w_3$
6	0	0	0	0	1	w_4	1	0	1	$3+w_4$
7	0	0	0	0	0	1	w_5	1	0	$2+w_5$
8	0	0	0	0	0	0	1	w_6	0	$1+w_6$
9	0	0	0	0	0	1	0	0	w_7	$1+w_7$

problem. The initial weights were randomly assigned to the diagonal elements of the augmented adjacency matrix. The first stage of the optimizing variable connectivity was to use the program of Forward in *Derivatives and Newton in Search* to find the probable values of the weights quickly. Then, the program of *Central in Derivatives and Newton in Search* was employed to calculate the weights. Finally, the program of *Central and Conjugate* was used to obtain the optimal weights. The chances of selecting a local minimum were reduced by ending with different initial weights at the beginning of the optimization procedure for several times [13].

RESULTS AND DISCUSSION

The boiling points of 100 alcohols were taken from the literature [14] (Table 3) and then divided into two subsets - the training set and the test set. The training set (70 alcohols), which was selected randomly, was used to create the model by means of mfVCI. The test set (30 alcohols) was employed to evaluate the predictive capability of the created model.

Assuming that there was a linear relationship between BP and ${}^1\chi^{mf}$, the *solver* in Excel was used to establish the model equation by means of the training set. Both the precision and the convergence were ended with 10^{-8} . In the process of calculation, the endpoint of the optimization program was judged by the lowest RMS error. The model equation is as follows:

$$\text{BP} = 42.94 \times {}^1\chi^{mf} - 70.49 \quad (2)$$

$$n = 70 \quad R^2 = 0.9979 \quad \text{RMS} = 1.382 \quad F = 33228$$

The calculated boiling points of 100 alcohols are listed in Table 3.

The molecular fragments, variables and optimum weights for all 100 alcohols are shown in Table 4.

The leave one-out (LOO) and the leave five-out (L5O) cross-validation method [15] was applied to validate the robust power of the model by MS Modeling 4.0 (purchased from Accelrys Inc.). The groups of five compounds were randomly selected from the training set. Each group was left out and predicted by the model developed from the remaining observations.

It is obvious that the model is quite robust to the inclusion-exclusion of the compound as indicated by the L_{00} and L_{50} correlation coefficients, which are presented below: $Q^2_{LOO} = 0.9976$ and $Q^2_{L5O} = 0.9977$.

According to Golbraikh and Tropsha [16], the predictive power of a QSAR model can be conveniently estimated by an external R^2_{Cvext} . Furthermore, Golbraikh and Tropsha [16] proposed the predictive power of the QSAR model, the following conditions being satisfied:

$$R^2_{Cvext} > 0.5; \quad R^2 > 0.6; \quad (R^2 > \max\{R_0^2, R_0'^2\}) \quad \text{and}$$

$$\frac{(R^2 - R_0^2)}{R^2} < 0.1 \quad \text{or} \quad \frac{(R^2 - R_0'^2)}{R^2} < 0.1$$

$$0.85 \leq k \leq 1.15 \quad \text{or} \quad 0.85 \leq k' \leq 1.15$$

Table 3. Listed of the Experimental/Calculated Normal Boiling Points and the First order of Molecular Fragments Variable Connectivity Index

Name	${}^1\chi^{mf}$	BP (°C)		Name	${}^1\chi^{mf}$	BP (°C)	
		Exp.	Pred.			Exp.	Pred.
Ethanol	3.48	78.0	78.9	2,3,3-Trimethyl-2-butanol	4.69	131.0	131.0
Propanol	3.85	97.1	94.8	Octanol	6.18	195.1	195.0
2-Propanol	3.56	82.4	82.4	6-Methyl-1-heptanol ^b	6.01	188.6	187.6
Butanol	4.42	117.6	119.1	4-Methyl-1-heptanol ^b	5.98	188.0	186.3
2-Methyl-1-propanol	4.14	108.1	107.1	2-Octanol	5.83	180.0	179.9
2-Butanol	4.01	99.5	101.7	2,5-Dimethyl-1-hexanol	5.79	179.5	178.1
2-Methyl-2-propanol	3.56	82.4	82.4	4-Octanol ^b	5.70	176.3	174.1
Pentanol	4.85	138.0	137.8	6-Methyl-3-heptanol ^b	5.53	174.0	166.9
3-Methyl-1-butanol ^b	4.71	131.0	132.0	5-Methyl-3-heptanol	5.55	172.0	167.8
2-Methyl-1-butanol ^b	4.73	128.0	132.8	3-Octanol ^b	5.70	171.0	174.3
2-Pentanol	4.43	119.3	119.8	5-Methyl-2-heptanol ^b	5.61	170.0	170.5
3-Pentanol	4.40	116.2	118.4	4-Methyl-3-heptanol	5.58	170.0	169.2
3-Methyl-2-butanol	4.30	112.9	114.2	2,4,4-Trimethyl-1-pentanol	5.57	168.5	168.5
2-Methyl-2-butanol	4.02	102.3	102.4	2-Methyl-3-heptanol	5.47	167.5	164.3
Hexanol	5.31	157.6	157.7	3-Methyl-2-heptanol	5.52	166.1	166.5
3-Methyl-1-pentanol ^b	5.20	153.0	152.6	3,4-Dimethyl-2-hexanol	5.50	165.5	165.5
4-Methyl-1-pentanol ^b	5.14	151.9	150.3	2-Methyl-4-heptanol	5.50	164.0	165.7
2-Methyl-1-pentanol	5.11	149.0	148.9	3-Methyl-3-heptanol	5.44	163.0	163.1
2-Ethyl-1-butanol	5.06	147.0	147.0	3-Methyl-4-heptanol	5.46	162.0	164.1
2,3-Dimethyl-1-butanol ^b	4.90	144.5	140.2	4-Methyl-4-heptanol ^b	5.44	161.0	162.9
3,3-Dimethyl-1-butanol	4.97	143.0	143.0	2-Methyl-3-ethyl-3-pentanol ^b	4.65	160.0	159.9
2-Hexanol	4.87	140.0	138.6	2,3-Dimethyl-2-hexanol	5.37	160.0	160.0
2,2-Dimethyl-1-butanol	4.82	136.5	136.5	2,3,4-Trimethyl-3-pentanol	5.29	156.5	156.5
3-Hexanol	4.83	135.0	136.9	2-Methyl-3-ethyl-2-pentanol	5.27	156.0	156.0
3-Methyl-2-pentanol	4.73	134.3	132.6	2-Methyl-2-heptanol	5.37	156.0	160.1
4-Methyl-2-pentanol ^b	4.48	131.6	122.0	2,5-Dimethyl-2-hexanol	5.20	154.5	152.8
2-Methyl-3-pentanol	4.60	126.5	127.1	2,2,4-Trimethyl-3-pentanol	5.15	150.5	150.5
3-Methyl-3-pentanol ^b	4.50	122.4	122.8	2,4,4-Trimethyl-2-pentanol	5.08	147.5	147.5
2-Methyl-2-pentanol	4.46	121.1	121.2	Nonanol ^b	6.62	213.3	213.6
3,3-Dimethyl-2-butanol	4.45	120.4	120.4	7-Methyloctanol	6.44	206.0	206.2
2,3-Dimethyl-2-butanol	4.40	118.4	118.4	3-Nonanol	6.13	195.0	192.9
Heptanol	5.75	176.4	176.3	2-Nonanol	6.15	193.5	193.5
4-Methyl-1-hexanol	5.62	173.0	170.9	5-Nonanol ^b	6.13	193.0	192.9
5-Methyl-1-hexanol	5.58	170.0	168.9	4-Nonanol	6.11	192.5	192.1
3-Methyl-1-hexanol	5.55	169.0	167.9	4-Ethyl-4-heptanol	5.88	182.0	182.2

Table 3. Continued

2-Methyl-1-hexanol	5.53	164.0	166.9	2-Methyl-2-octanol	5.80	178.0	178.7
2-Heptanol ^b	5.30	159.0	157.3	2,6-Dimethyl-3-heptanol	5.73	175.0	175.6
2,4-Dimethyl-1-pentanol	5.33	159.0	158.5	2,6-Dimethyl-4-heptanol ^b	5.74	174.5	176.0
3-Heptanol ^b	5.27	157.0	155.7	2,6-Dimethyl-2-heptanol	5.63	173.0	171.3
4-Heptanol	5.26	156.0	155.4	3,6-Dimethyl-3-heptanol ^b	5.70	173.0	174.4
5-Methyl-2-hexanol	5.13	151.0	149.9	2,2,3-Trimethyl-3-hexanol	5.27	156.0	156.0
5-Methyl-3-hexanol ^b	5.07	148.0	147.2	Decanol ^b	7.05	231.1	232.2
2-Methyl-2-hexanol	4.94	143.0	141.5	3,7-Dimethyl-1-octanol	6.68	212.5	216.6
2-Methyl-3-hexanol ^b	5.03	143.0	145.6	2-Decanol ^b	6.60	211.0	213.2
3-Methyl-3-hexanol	4.97	143.0	142.9	4-Decanol ^b	6.56	210.5	211.4
3-Ethyl-3-pentanol ^b	4.89	142.0	139.6	3,6-Dimethyl-3-octanol ^b	6.18	202.2	195.1
2,3-Dimethyl-3-pentanol	4.89	139.7	139.7	3-Ethyl-3-octanol ^b	5.42	199.0	198.9
2,4-Dimethyl-3-pentanol	4.87	138.7	138.7	2,6-Dimethyl-4-octanol	6.22	195.0	196.6
2,2-Dimethyl-3-pentanol	4.78	135.0	135.0	2,7-Dimethyl-3-octanol	6.16	193.5	194.2
2,4-Dimethyl-2-pentanol ^b	4.77	133.1	134.3	3-Ethyl-2-methyl-3-heptanol	6.14	193.0	193.0

Note: The compounds representing the test set are marked with the italic letter b.

Table 4. Optimum weights for atom in different fragments

No.	Fragments	Variables	Weights	No.	Fragments	Variables	Weights
1	$\text{CH}_3(\text{CH}_2)$	w_1	65.352	27	$2(\text{CH})\text{CH}(\text{OH})$	w_{27}	-2.193
2	$\text{CH}_3(\text{CH})$	w_2	6.793	28	$2(\text{CH}_2)\text{CH}(\text{CH}_3)$	w_{28}	-0.538
3	$\text{CH}_3(\text{C})$	w_3	0.087	29	$2(\text{CH}_2)\text{CH}(\text{OH})$	w_{29}	-2.166
4	$(\text{CH}_3)\text{CH}_2(\text{CH}_2)$	w_4	4.293	30	$2(\text{CH}_2)\text{CH}(\text{C})$	w_{30}	-2.718
5	$(\text{CH}_3)\text{CH}_2(\text{CH})$	w_5	1.829	31	$(\text{CH}_2)(\text{CH})\text{CH}(\text{OH})$	w_{31}	-2.166
6	$(\text{CH}_3)\text{CH}_2(\text{C})$	w_6	0.742	32	$(\text{CH}_2)(\text{C})\text{CH}(\text{OH})$	w_{32}	1.670
7	$(\text{CH}_3)\text{CH}_2(\text{OH})$	w_7	-1.263	33	$2(\text{CH})\text{CH}(\text{CH}_3)$	w_{33}	-0.930
8	$2(\text{CH}_2)\text{CH}_2$	w_8	0.305	34	$(\text{CH})(\text{C})\text{CH}(\text{OH})$	w_{34}	1.138
9	$(\text{CH}_2)\text{CH}_2(\text{CH})$	w_9	0.214	35	$(\text{CH})\text{C}(\text{C})(\text{CH}_3)$	w_{35}	-2.506
10	$(\text{CH}_2)\text{CH}_2(\text{C})$	w_{10}	-0.449	36	$3(\text{CH}_3)\text{C}(\text{CH}_2)$	w_{36}	26.404
11	$(\text{CH}_2)\text{CH}_2(\text{OH})$	w_{11}	-1.274	37	$3(\text{CH}_3)\text{C}(\text{CH})$	w_{37}	-2.878
12	$2(\text{CH})\text{CH}_2$	w_{12}	0.244	38	$3(\text{CH}_3)\text{C}(\text{C})$	w_{38}	-2.684
13	$(\text{CH})\text{CH}_2(\text{C})$	w_{13}	-0.626	39	$3(\text{CH}_3)\text{C}(\text{OH})$	w_{39}	-1.406
14	$(\text{CH})\text{CH}_2(\text{OH})$	w_{14}	-1.277	40	$2(\text{CH}_3)2(\text{CH}_2)\text{C}$	w_{40}	-1.107
15	$2(\text{C})\text{CH}_2$	w_{15}	-1.238	41	$2(\text{CH}_3)(\text{CH}_2)\text{C}(\text{OH})$	w_{41}	-2.145
16	$(\text{C})\text{CH}_2(\text{OH})$	w_{16}	-0.881	42	$2(\text{CH}_3)(\text{CH})\text{C}(\text{OH})$	w_{42}	0.269
17	$3(\text{CH}_2)\text{CH}$	w_{17}	-1.044	43	$2(\text{CH}_3)(\text{C})\text{C}(\text{OH})$	w_{43}	1.339

Table 4. Continued

18	2(CH ₃)CH(CH ₂)	w ₁₈	3.060	44	(CH ₃)(CH ₂)(CH)C(OH)	w ₄₄	-1.032
19	2(CH ₃)CH(CH)	w ₁₉	1.825	45	(CH ₃)(CH ₂)(C)C(OH)	w ₄₅	1.247
20	2(CH ₃)CH(C)	w ₂₀	-2.669	46	3(CH ₂)C(OH)	w ₄₆	-3.000
21	2(CH ₃)CH(OH)	w ₂₁	-1.980	47	2(CH ₂)(CH ₃)C(OH)	w ₄₇	-2.668
22	(CH ₃)(CH ₂)CH(CH)	w ₂₂	-0.292	48	2(CH ₂)(CH)C(OH)	w ₄₈	-1.497
23	(CH ₃)(CH ₂)CH(C)	w ₂₃	-2.694	49	2(CH)(CH ₃)C(OH)	w ₄₉	2.813
24	(CH ₃)(CH ₂)CH(OH)	w ₂₄	-2.097	50	(CH ₂)OH	w ₅₀	-0.878
25	(CH ₃)(CH)CH(OH)	w ₂₅	-2.137	51	(CH)OH	w ₅₁	-0.879
26	(CH ₃)(C)CH(OH)	w ₂₆	2.852	52	(C)OH	w ₅₂	-0.877

Mathematical definitions of R^2 , R^2_{ext} , k and k' are based on regression of the observed activities against the predicted activities and vice versa (regression of the predicted activities against observed activities). The definitions are clearly presented [16] and are not repeated here for brevity. More specifically, the proposed model passed all the tests for the predictive ability as:

$$R^2_{C_{ext}} = 0.95 > 0.5; R^2 = 0.997 > 0.6; R^2_0 = 0.899 \text{ and}$$

$$R^2_0 = 0.856;$$

$$\frac{(R^2 - R^2_0)}{R^2} = 0.099 < 0.1 \text{ or } \frac{(R^2 - R^2_0)}{R^2} = 0.041 < 0.1$$

$$k = 1 \text{ and } k' = 0.9999$$

The results of the validation show that the proposed model is reliable and applicable for both external prediction and regulatory purposes.

According to the definition of mfVCI, the weights of the molecular fragments, which were assigned to diagonal elements of augmented matrix, represent the contributions of the molecular fragments to the overall boiling point of the alcohols. The results presented in Table 4 show that the weights are different for the same atom in different molecular fragments. The molecular fragments play a dominant role in the boiling points of alcohols. It is clear that the negative values of the weights which increase the boiling points derive from all molecular fragments containing hydroxyl groups and

almost all carbon atoms bonding with hydroxyl groups. So, the weights of the molecular fragments provide quite adequate information for the user who can gain direct knowledge about which molecular fragment is an enhancer and which is a suppressor to the boiling point. The first order of the mfVCI is a summation of all edges in Eq. (1). The contribution of the adjacent vertices by one edge to the index is expressed as:

$${}^1\chi_p^{mf} = \frac{1}{\sqrt{\delta_i^{mf} \delta_j^{mf}}} \quad (3)$$

Here, i and j are the two conjoint fragments by edge p . Therefore, the ${}^1\chi_p^{mf}$ should be used to interpret the influence of molecular fragments on the property of the compounds. As an example, the numerical results in the case of 6-methyl-3-heptanol are illustrated in Table 5.

The molecular fragments can be differentiated by their characteristics which are determined by chemical bonding. Based on the different characteristics, the mfVCI, which offers quite comprehensive and exact molecular structural information, is proposed in this paper. The negative values show that the contribution of the CH_x-OH bond is much greater to the boiling point than others. In Table 6, the contribution of every bond between two molecular fragments on boiling points indicates the following rank order: 6-9 > 5-6 > 6-7 > 4-5 > 3-4 > 1-3 > 2-3 > 7-8. The single C-O (6-9) bond makes the biggest contribution. The C-C bond (5-6 and 6-7), which contains a connected oxygen carbon, is a secondary contribution to the molecular. This rank order can

Table 5. All Bonds' Numerical Results of 6-Methyl-3-heptanol

No.	Bond parts	Formula	Contribution
1	1-3	$\frac{1}{\sqrt{(1+w_1)(3+w_2)}}$	0.146
2	2-3	$\frac{1}{\sqrt{(1+w_1)(3+w_2)}}$	0.146
3	3-4	$\frac{1}{\sqrt{(3+w_2)(2+w_3)}}$	0.273
4	4-5	$\frac{1}{2+w_3}$	0.452
5	5-6	$\frac{1}{\sqrt{(2+w_3)(3+w_4)}}$	0.736
6	6-7	$\frac{1}{\sqrt{(3+w_4)(2+w_5)}}$	0.560
7	7-8	$\frac{1}{\sqrt{(2+w_5)(1+w_6)}}$	0.063
8	6-9	$\frac{1}{\sqrt{(3+w_4)(1+w_7)}}$	3.153

Note: The number bond is the atoms numbering in Fig. 1.

be applied to any of the compounds in this study.

Moreover, the mfVCI can better differentiate between position isomers. For example, the boiling points for 3-heptanol and 4-heptanol were reported as 157 °C and 156 °C, respectively; however, they were predicted both as 155.3 °C and in ref. [14]. Using the mfVCI, the predicted values for the boiling points of the two alcohols were 155.7 and 155.4, respectively, which are closer to the experimental values. This shows that the relative contribution to the variable connectivity index of the atom in different fragments is different.

All the contributions of CH_x-OH bonds to the mfVCI were calculated and are listed in Table 6.

Herein, we present the relative magnitudes of BP: the primary alcohols will have the highest boiling point, the secondary alcohols will be in the middle, and the tertiary alcohols will have the lowest boiling point. Nevertheless, this regularity is violated in few cases. For example, the boiling points for 2-hexanol and 2,2-dimethyl-1-butanol are reported as 140 °C and 136.5 °C, despite the fact that the former is a

secondary alcohol and the latter a primary alcohol. Analyzing the structure of the two alcohols, we find that the former CH_x-OH bond contains "(CH₃)(CH₂)CH(OH)" fragments, and the latter CH_x-OH bond contains "(C)CH₂(OH)" fragments. It is clear that the contributions of the two bonds are 9.152 and 7.325, respectively. This violation may be related to the complexity of molecular structures. The simpler the alcohol isomer is, the more potent the boiling point.

CONCLUSIONS

The Molecular Fragments Variable Connectivity Index which contains more comprehensive and exact molecular structure information has been proposed in this paper, with molecular fragments being defined as the atoms or functional groups in different characteristics of chemical bonding. The results of prediction about the boiling points of alcohols are very encouraging and powerful. With the help of fewer descriptors in simple linear regression, the mfVCI is capable of accounting for the enhancing or suppressing contributions

Table 6. All the Contribution of CH_x-OH Bonds in Alcohol Boiling Points

Sorts of alcohols	Conjunction fragments	Contribution
Primary alcohols	(C)CH ₂ (OH)	7.325
	(CH ₂)CH ₂ (OH)	11.290
	(CH)CH ₂ (OH)	11.337
Secondary alcohols	(CH ₃)(C)CH(OH)	1.412
	(CH ₂)(C)CH(OH)	1.770
	(CH)(C)CH(OH)	1.997
	2(CH ₃)CH(OH)	8.102
	(CH ₃)(CH ₂)CH(OH)	9.152
	(CH ₃)(CH)CH(OH)	9.576
	2(CH ₂)CH(OH)	9.909
	(CH ₂)(CH)CH(OH)	9.909
	2(CH)CH(OH)	10.241
Tertiary alcohols	2(CH)(CH ₃)C(OH)	1.193
	2(CH ₃)(C)C(OH)	1.523
	(CH ₃)(CH ₂)(C)C(OH)	1.549
	2(CH ₃)(CH)C(OH)	1.904
	(CH ₃)(CH ₂)(CH)C(OH)	2.739
	2(CH ₂)(CH)C(OH)	3.248
	2(CH ₃)(CH ₂)C(OH)	4.383
	2(CH ₂)(CH ₃)C(OH)	6.104
3(CH ₂)C(OH)	8.130	

of molecular fragments in alcohols. Currently, its applications to other physical properties and to more complex compounds are under investigation, the results of which will be reported in our future papers.

ACKNOWLEDGMENTS

This work was supported by the Technological Project of Gansu (No. 0804GKCA029), State Key Laboratory of Environmental Chemistry and Ecotoxicology, Research Center for EES of CAS (KF2008-03), and the Fundamental Research Funds for the Central Universities (Izujbky-2009-7).

REFERENCES

- [1] C.E. Rechsteiner, in: W.J. Lyman, W.F. Reehl, D.H. Rosenblatt (Eds.), *Handbook of Chemical Property Estimation Methods*, McGraw-Hill: New York, 1982, Chap. 12.
- [2] E. Voutsas, M. Lampadariou, K. Magoulas, D. Tassios, *Fluid Phase Equilib.* 198 (2002) 81.
- [3] E. Panteli, E. Voutsas, K. Magoulas, D. Tassios, *Fluid Phase Equilib.* 248 (2006) 70.
- [4] A. Komasa, *Polish J. Chem.* 77 (2003) 1491.
- [5] M. Kompany-Zareh, *Acta Chim. Slov.* 50 (2003) 259.

Application of Molecular Fragments Variable Connectivity Index

- [6] T. Öberg, *J. Chem. Inf. Comput. Sci.* 44 (2004) 187.
- [7] M. Pérez González, A.A. Toropov, P.R. Duchowicz, D.A. Castro, *Molecules* 9 (2004) 1019.
- [8] Q. Li, X. Chen, Z. Hu, *Chemom. Intel. Labl. Syst.* 72 (2004) 93.
- [9] M. Randic, *New. J. Chem.* 24 (2000) 165.
- [10] M. Randic, *Chemom. Intel. Labl. Syst.* 10 (1991) 213.
- [11] M. Randic, *J. Chem. Inf. Comput. Sci.* 31 (1991) 970.
- [12] M. Randic, *J. Am. Chem. Soc.* 97 (1975) 6609.
- [13] M. Pompe, M. Randić, *J. Chem. Inf. Model.* 46 (2006) 2.
- [14] M. Randic, M. Pompe, *J. Chem. Inf. Comput. Sci.* 41 (2001) 631.
- [15] P. Gramatica, P. Pilutti, E. Papa, *J. Chem. Inf. Comput. Sci.* 44 (2004) 1794.
- [16] A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model.* 20 (2002) 269.

Archive of SID