Molecular Distance−Edge Vector (μ): An Extension from Alkanes to Alcohols

Shushen Liu,[†,‡] Hailing Liu,[‡] Zhining Xia,[§] ChenzhongCao,[||] and Zhiliang Li*[,§]

College of Bioengineering and College of Environmental and Chemical Engineering, Chongqing University, Chongqing, 400044, P R China, Department of Applied Chemistry, Guilin Institute of Technology, 541004, P R China, College of Environmental and Chemical Engineering, Chongqing University, Chongqing, 400044, P R China, and Department of Chemistry, Xiangtan Teacher's College, Xiangtan, 411100, P R China

Received February 14, 1999

The oxygen atom in the hydroxyl group OH of an alcohol molecule is considered a pseudo-carbon atom, like a carbon atom in alkanes, for our present case of extension from alkanes to alcohols. The ratio of electronegativity of oxygen (EN_O) to that of carbon (EN_C) is defined as the relative electronegativity, REN $= EN_O/EN_C$, of a oxygen atom under the condition of REN being 1 for a carbon atom; similarly, the value of bond length for the bond C−O (BL_O) over that of the bond C−C (BL_C) is regarded as the relative bond length, RBL $= BL_O/BL_C$, of the bond C−O with RBL being equal to 1 for C−C. The molecular distance−edge vector, i.e., vector $\boldsymbol{\mu}$, for alkanes can be extended to describe the molecular structure of alcohols instead of alkanes. A quantitative structure−property relationship (QSPR) equation can be modeled by using multiple linear regression (MLR). The first model, M1, between the modified $\boldsymbol{\mu}$ vector and boiling points of 106 alcohols is created with correlation coefficient being $R = 0.9951$ and root-mean-square error being rms $= 3.222$ °C between the values estimated by the above QSPR model and the experimental boiling points. With 30 compounds drawn from all 106 alcohols we construct an external prediction set, and the remaining 76 ones construct an internal calibration set, and the second model, M2, is developed with the statistics being $R = 0.9946$ and rms $= 3.480$ °C. The 30 boiling points in the external testing set can be predicted by model M2, and good prediction results are obtained with R being 0.9964 and rms being 2.586 °C.

INTRODUCTION

Estimation and prediction of the physicochemical properties such as boiling points of organic molecules based on their structures have become major interests of many chemists and scientists in recent years. It is necessary to describe the chemical structure of the examined compound as one or more parameters such as various topological indices in developing a quantitative structure−property relationship (QSPR) model. More than 100 topological indices such as the Wiener index, Hosoya index, and Randic indices were developed, and these structural descriptors have widely been applied in QSPR studies.[1−5] However, it seems to be unreasonable that a compound's structure is described by only using an index consisting of one number. The $\boldsymbol{\lambda}$ vector containing 10 elements being used to describe the structure of alkanes was proposed in our previous paper and this $\boldsymbol{\lambda}$ vector was related to physical properties such as boiling point[6] and thermodynamic functions of alkanes[7] with good results. In evaluating the values of elements in the $\boldsymbol{\lambda}$ vector, a geometric distance between the two types of atoms must be calculated and the $\boldsymbol{\lambda}$ element will be equal to zero if there exists a very large value in all distances. So, it is necessary to modify the $\boldsymbol{\lambda}$ vector, and a novel vector, called the $\boldsymbol{\mu}$ vector, was proposed and was related to eight physical properties and seven thermodynamic functions of alkanes in our other paper.[8]

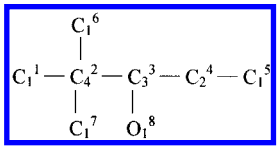

Figure 1. The molecular skeleton graph of 2,2-dimethyl-3-pentanol.

Some reports[9−14] have been contributed to structural description and QSPR models for a series of organic compounds, containing heteroatoms such as oxygen. In the present paper, the $\boldsymbol{\mu}$ vector will be extended to describe the structure of alcohols by introduction of relative electronegative (REN) and relative bond length (RBL) and will quantitatively be related to boiling points of 106 alcohols by using multiple linear regression (MLR). The results obtained show that the $\boldsymbol{\mu}$vector can be used to describe the structure of alcohols and to calculate, estimate, and/or predict the boiling points of these examined compounds.

THEORETICAL SECTION

Calculation of the Molecular Distance−Edge Vector ($\boldsymbol{\mu}$ Vector) for Alcohols. To calculate the molecular distance−edge vector, called the $\boldsymbol{\mu}$ vector, of alcohols, two simple concepts, relative electronegative (REN) and relative bond length (RBL), are introduced. The oxygen atom of a hydroxyl group in an alcohol molecule is considered a pseudo-carbon atom which has a different volume and electronegative (EN) from those of a real carbon atom. Thus, the REN value of an oxygen atom is defined as the ratio of Pauling's electronegativity of oxygen ($EN_O = 3.44$) to Pauling's electronegativity of carbon ($EN_C = 2.55$), and the RBL value of

[†] College of Bioengineering, Chongqing University.
[‡] Guilin Institute of Technology.
[§] College of Environmental and Chemical Engineering, Chongqing University.
[||] Xiangtan Teacher's College.

10.1021/ci990011f CCC: $18.00
Published on Web 10/01/1999

Table 1. Values of $d_{k,j}^{i,l}$, $q_{k,j}^{i,l}$, and μ_{kj} of 2,2-Dimethyl-3-pentanol

q	k	j	$d_{k,j}^{i,l}$ ($i \in k$, $l \in j$)	$q_{k,j}^{i,l}$	μ_{kj}
1	1	1	$2(C_1^1-C_1^6)$, $2(C_1^1-C_1^7)$, $2.9286(C_1^1-C_1^8)$, $4(C_1^1-C_1^5)$, $2.9286(C_1^6-C_1^8)$, $2(C_1^6-C_1^7)$, $4(C_1^6-C_1^5)$, $2.9286(C_1^7-C_1^8)$, $4(C_1^7-C_1^5)$, $2.9286(C_1^8-C_1^5)$	1, 1, 1.3490, 1, 1.3490, 1, 1, 1.3490, 1, 1.3490	1.5667
2	1	2	$3(C_1^1-C_2^4)$, $3(C_1^6-C_2^4)$, $3(C_1^7-C_2^4)$, $1.9286(C_1^8-C_2^4)$, $1(C_1^5-C_2^4)$	1, 1, 1, 1.3490, 1	1.6960
3	1	3	$2(C_1^1-C_3^3)$, $2(C_1^6-C_3^3)$, $2(C_1^7-C_3^3)$, $0.9286(C_1^8-C_3^3)$, $2(C_1^5-C_3^3)$	1, 1, 1, 1.3490, 1	2.5645
4	1	4	$1(C_1^1-C_4^2)$, $1(C_1^6-C_4^2)$, $1(C_1^7-C_4^2)$, $1.9286(C_1^8-C_4^2)$, $3(C_1^5-C_4^2)$	1, 1, 1, 1.3490, 1	3.4738
5	2	2	∞	0	0
6	2	3	$1(C_2^4-C_3^3)$	1	1.0000
7	2	4	$2(C_2^4-C_4^2)$	1	0.2500
8	3	3	∞	0	0
9	3	4	$1(C_3^3-C_4^2)$	1	1.0000
10	4	4	∞	0	0

a chemical bond linking oxygen and carbon (i.e. the bond C−O) is defined as the ratio of the bond length (BL) of C−O ($BL_O = 0.143$ nm) to that of C−C ($BL_C = 0.154$ nm).[9] So, the REN values of the elements O and C and the RBL values of bonds C−O and C−C are calculated as follows:

$$REN_O = \frac{EN_O}{EN_C} = \frac{3.44}{2.55} = 1.3490 \quad REN_C = \frac{EN_C}{EN_C} = 1$$

$$RBL_O = \frac{BL_O}{BL_C} = \frac{0.143\text{nm}}{0.154\text{nm}} = 0.9296 \quad RBL_C = \frac{BL_C}{BL_C} = 1$$

The $\boldsymbol{\mu}$ vector values of alcohols can be calculated after acquiring the RBL and REN values. Taking 2,2-dimethyl-3-pentanol as an example, the $\boldsymbol{\mu}$ vector values can be calculated by using the molecular graph (seeing Figure 1), which is illustrated as follows: Considering oxygen as a pseudo-carbon atom in alcohol, the carbon and/or oxygen atoms in the examined alcohol can be classified as four basic types, according to the connecting C−C or C−O bond number between it and the other atoms. Types 1, 2, 3 and 4 for primary (CH_3-), secondary ($CH_2<$), ternary ($-CH<$), and quaternary ($>C<$) carbons, being symbolized C_1, C_2, C_3, and C_4 in Figure 1, represent carbon atoms connecting one, two, three, and four bonds either C−C or C−O, respectively. Let $d_{k,j}^{i,l}$ expresses the sum of the RBL between C_k^i and C_j^l, which is actually the sum of the RBL of all chemical bonds, either C−C or C−O, between C_k^i and C_k^l, where the ith carbon atom or pseudo-carbon atom belongs to type C_k while the lth carbon atom belongs to the type C_j. Let $q_{k,j}^{i,l}$ refer to a kind of electrostatic action that equals 1.0000 when atom i and atom l are both carbon atoms and equals REN_O when atom i or atom l is an oxygen atom, i.e., a pseudo-carbon atom in our case. For 2,2-dimethyl-3-pentanol, the example compound, all values of $d_{k,j}^{i,l}$ and $q_{k,j}^{i,l}$ are listed in Table 1 where "∞" represents "no" distance or no linking between both atoms.

In general, every interaction energy of two arbitrary types of carbon or pseudo-carbon atoms is changing as a decrease of $(d^{i,l}{}_{k,j})^2$ and an increase of $q_{k,j}^{i,l}$. As described in our previous paper,[8] we can define a new function to approach this interaction energy and still call the function the $\boldsymbol{\mu}$ vector. A mathematical form of this "$\boldsymbol{\mu}$" vector is determined as

$$\mu_L = \mu_{kj} = \sum_{i \in k, l \in j}^{n_{kj}} \left(\frac{q_{k,j}^{i,l}{}_2}{d_{k,j}^{i,l}} \right) \qquad (k = 1, 2, 3, 4;\ j \geq k;\ L = (k-1)4 + j) \quad (1)$$

where n_{kj} refers to the item number of all terms $d_{k,j}^{i,l}$ or $q_{k,j}^{i,l}$ between atom i belonging to C_k and atom l to C_j. It is known from eq 1 that there are only 10 elements in the $\boldsymbol{\mu}$ vector:

$$\mu = (\mu_{11}, \mu_{12}, \mu_{13}, \mu_{14}, \mu_{22}, \mu_{23}, \mu_{24}, \mu_{33}, \mu_{34}, \mu_{44}) = (\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6, \mu_7, \mu_8, \mu_9, \mu_{10})$$

By applying the information given in Table 1 into eq 1, all 10 entries in the $\boldsymbol{\mu}$ vector for compound 2,2-dimethyl-3-pentanol can be obtained. For example, the $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are calculated as follows:

$$\mu_1 = \frac{1}{2^2} + \frac{1}{2^2} + \frac{1.3490}{2.9286^2} + \frac{1}{4^2} + \frac{1.3490}{2.9286^2} + \frac{1}{2^2} + \frac{1}{4^2} + \frac{1.3490}{2.9286^2} + \frac{1}{4^2} + \frac{1.3490}{2.9286^2} = 1.5624$$

$$\mu_2 = \frac{1}{3^2} + \frac{1}{3^2} + \frac{1}{3^2} + \frac{1.3490}{1.9286^2} + \frac{1}{1^2} = 1.6923$$

The other eight μ_L ($L = 3, 4, ..., 10$) values obtained in the same manner as above are also listed in Table 1.

Modeling Relationship between the $\boldsymbol{\mu}$ Vector and the Boiling Point of the Alcohol. A multiple linear regression (MLR) method was employed to develop a model for the quantitative relationship between a physicochemical property, such as boiling point (BP), of the interested alcohols and their 10 descriptors in the $\boldsymbol{\mu}$ vector. The model was given in the following form:

$$BP = b_0 = \sum_{L=1}^{10} b_L \mu_L \qquad (2)$$

where b_0 is the intercept term of the regression equation and b_L is the regression coefficient for the corresponding Lth descriptor μ_L, which is the contribution value of the Lth element μ_L to BP or the other physical properties of the compound.

RESULTS AND DISCUSSION

Data Set. The boiling points observed experimentally for all 106 alcohol compounds are taken from ref.[10] The range of the experimental BP values was from 64.60 to 245.00 °C, and the number of carbon atoms per alcohol spanned from 1 through 11. These 106 compounds construct a working data set. Of these 106 aliphatic alcohols, 30 compounds were randomly chosen as an external testing or prediction set, and the remaining 76 compounds comprised an internal training or calibration set. The external prediction set is never

Table 2. Values of μ Vector for 106 Alcohol Compounds

no.	alcohols	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_9	μ_{10}
1	methanol	1.5645	0	0	0	0	0	0	0	0	0
2	ethanol	0.3627	2.5645	0	0	0	0	0	0	0	0
3	propanol	0.1573	3.1772	0	0	1.0000	0	0	0	0	0
4	2-propanol	0.9754	0	3.5645	0	0	0	0	0	0	0
5	butanol	0.0874	3.4456	0	0	2.2500	0	0	0	0	0
6	2-methyl-1-propanol	0.5646	2.0645	2.3627	0	0	1.0000	0	0	0	0
7	2-butanol	0.6311	1.6127	2.8145	0	0	1.0000	0	0	0	0
8	2-methyl-2-propanol	1.8381	0	0	4.5645	0	0	0	0	0	0
9	pentanol	0.0555	3.5956	0	0	3.6111	0	0	0	0	0
10	3-methyl-1-butanol	0.4248	2.6495	2.1573	0	1.0000	1.2500	0	0	0	0
11	2-methyl-1-butanol	0.3558	3.3329	1.6127	0	0.2500	2.0000	0	0	0	0
12	2-pentanol	0.5126	2.1311	2.6757	0	1.0000	1.2500	0	0	0	0
13	3-pentanol	0.3771	2.9476	2.0645	0	0.2500	2.0000	0	0	0	0
14	3-methyl-2-butanol	1.1495	0	5.6772	0	0	0	0	1.0000	0	0
15	2-methyl-2-butanol	1.3549	1.8627	0	3.8145	0	0	1.0000	0	0	0
16	hexanol	0.0384	3.6911	0	0	5.0347	0	0	0	0	0
17	3-methyl-1-pentanol	0.2541	3.7994	1.4073	0	1.3611	2.2500	0	0	0	0
18	4-methyl-1-pentanol	0.3611	2.9318	2.0874	0	2.2500	1.3611	0	0	0	0
19	2-methyl-1-pentanol	0.2753	3.7329	1.4738	0	1.3611	2.2500	0	0	0	0
20	2-ethyl-1-butanol	0.2373	4.3236	0.8627	0	0.7500	3.0000	0	0	0	0
21	2,3-dimethyl-1-butanol	0.8043	2.0368	4.2700	0	0	1.2500	0	1.0000	0	0
22	2,3-dimethyl-1-butanol	1.0122	3.0106	0	3.1573	1.0000	0	1.2500	0	0	0
23	2-hexanol	0.4582	2.3921	2.6270	0	2.2500	1.3611	0	0	0	0
24	2,2-dimethyl-1-butanol	0.8742	3.8329	0	2.6127	0.2500	0	2.0000	0	0	0
25	3-hexanol	0.2847	3.3688	1.9257	0	1.3611	2.2500	0	0	0	0
26	3-methyl-2-pentanol	0.8921	1.5184	4.7884	0	0	1.2500	0	1.0000	0	0
27	4-methyl-2-pentanol	0.9125	1.1127	5.0552	0	0	2.0000	0	0.2500	0	0
28	2-methyl-3-pentanol	0.8469	1.5849	4.7884	0	0	1.2500	0	1.0000	0	0
29	3-methyl-3-pentanol	0.9620	3.4476	0	3.0645	0.2500	0	2.0000	0	0	0
30	2-methyl-2-pentanol	1.1878	2.4922	0	3.6757	1.0000	0	1.2500	0	0	0
31	3,3-dimethyl-2-butanol	1.9179	0	3.3145	3.6127	0	0	0	0	1.0000	0
32	2,3-dimethyl-2-butanol	1.9844	0	2.8627	4.0645	0	0	0	0	1.0000	0
33	heptanol	0.0281	3.7572	0	0	6.4983	0	0	0	0	0
34	4-methyl-1-hexanol	0.2050	4.0273	1.3374	0	2.6736	2.3611	0	0	0	0
35	5-methyl-1-hexanol	0.3268	3.0992	2.0555	0	3.6111	1.4236	0	0	0	0
36	3-methyl-1-hexanol	0.1883	4.1449	1.2684	0	2.5347	2.5000	0	0	0	0
37	2-methyl-1-hexanol	0.2357	3.9395	1.4252	0	2.6736	2.3611	0	0	0	0
38	2,4-dimethyl-1-pentanol	0.6434	2.8468	3.7834	0	0.2500	3.1111	0	0.2500	0	0
39	2-heptanol	0.4289	2.5502	2.6045	0	3.6111	1.4236	0	0	0	0
40	2-methyl-2-hexanol	1.1109	2.8157	0	3.6270	2.2500	0	1.3611	0	0	0
41	5-methyl-2-hexanol	0.8038	1.6033	4.8395	0	1.0000	2.5000	0	0.1111	0	0
42	2,3-dimethyl-2-pentanol	1.6784	1.6295	2.1127	3.9257	0	1.0000	0.2500	0	1.0000	0
43	2,4-dimethyl-2-pentanol	1.6502	1.3627	2.3795	3.7868	0	1.0000	1.0000	0	0.2500	0
44	2,3,3-trimethyl-2-butanol	2.8639	0	0	8.1005	0	0	0	0	0	1.0000
45	3-heptanol	0.2406	3.5848	1.8770	0	2.6736	2.3611	0	0	0	0
46	2-methyl-3-hexanol	0.7320	2.1172	4.6009	0	1.0000	1.6111	0	1.0000	0	0
47	3-methyl-3-hexanol	0.8210	3.9799	0	2.9257	1.3611	0	2.2500	0	0	0
48	5-methyl-3-hexanol	0.6621	2.4615	4.2566	0	0.2500	3.1111	0	0.2500	0	0
49	3-ethyl-3-pentanol	0.6594	4.7548	0	2.3145	0.7500	0	3.0000	0	0	0
50	2,2-dimethyl-3-pentanol	1.5667	1.6960	2.5645	3.4738	0	1.0000	0.2500	0	1.0000	0
51	2,3-dimethyl-3-pentanol	1.5429	1.8349	2.7238	3.3145	0	0.2500	1.0000	0	1.0000	0
52	2,4-dimethyl-3-pentanol	1.3792	0	7.7344	0	0	0	0	2.2500	0	0
53	4-heptanol	0.2026	3.7450	1.7868	0	2.5347	2.5000	0	0	0	0
54	octanol	0.0215	3.8058	0	0	7.9897	0	0	0	0	0
55	4-methyl-1-heptanol	0.1461	4.3435	1.1985	0	3.8872	2.6111	0	0	0	0
56	6-methyl-1-heptanol	0.3062	3.2103	2.0384	0	5.0347	1.4636	0	0	0	0
57	2,5-dimethyl-1-hexanol	0.5641	3.2226	3.6057	0	1.3611	3.5625	0	0.1111	0	0
58	2,4,4-trimethyl-1-pentanol	1.2614	3.1593	1.6960	3.1985	0.2500	2.0000	1.1111	0	0.2500	0
59	2-octanol	0.4112	2.6563	2.5923	0	5.0347	1.4636	0	0	0	0
60	2-methyl-2-heptanol	1.0693	3.0138	0	3.6045	3.6111	0	1.4236	0	0	0
61	3-methyl-2-heptanol	0.7373	2.2986	4.5297	0	2.2500	1.7847	0	1.0000	0	0
62	5-methyl-2-heptanol	0.6355	2.7614	4.0670	0	1.3611	3.5625	0	0.1111	0	0
63	2,3-dimethyl-2-hexanol	1.5529	2.2030	1.9738	3.8770	1.0000	1.2500	0.3611	0	1.0000	0
64	2,5-dimethyl-2-hexanol	1.4965	1.9644	2.2124	3.6895	1.0000	1.2500	1.2500	0	0.1111	0
65	3,4-dimethyl-2-hexanol	1.1613	1.5110	6.7304	0	0	1.3611	0	2.2500	0	0
66	2-methyl-3-ethyl-2-pentanol	1.4627	2.9813	1.3627	3.7868	0.2500	2.0000	0.5000	0	1.0000	0
67	2,3,4-trimethyl-2-pentanol	2.2520	0	4.9922	4.0368	0	0	0	1.0000	1.2500	0
68	3-octanol	0.2161	3.7184	1.8545	0	4.0747	2.4236	0	0	0	0
69	2-methyl-3-heptanol	0.6757	2.3957	4.5297	0	2.2500	1.7847	0	1.0000	0	0
70	3-methyl-3-heptanol	0.7544	4.2584	0	2.8770	2.6736	0	2.3611	0	0	0
71	4-methyl-3-heptanol	0.5229	3.4721	3.7120	0	1.1736	2.8611	0	1.0000	0	0
72	5-methyl-3-heptanol	0.4791	3.6775	3.4579	0	0.5625	4.2222	0	0.2500	0	0
73	6-methyl-3-heptanol	0.5739	2.8585	4.0670	0	1.3611	3.5625	0	0.1111	0	0
74	2-methyl-3-ethyl-3-pentanol	1.1917	3.3921	2.5849	2.5645	0.2500	0.5000	2.0000	0	1.0000	0
75	2,2,4-trimethyl-3-pentanol	2.1615	0	5.5106	3.5849	0	0	0	1.0000	1.2500	0

Table 2. (Continued)

no.	alcohols	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_9	μ_{10}
76	2,3,4-trimethyl-3-pentanol	2.1863	0	5.6698	3.5645	0	0	0	0.2500	2.0000	0
77	4-octanol	0.1634	3.9366	1.7382	0	3.8872	2.6111	0	0	0	0
78	2-methyl-4-heptanol	0.5678	2.9002	4.0952	0	1.3611	3.4236	0	0.2500	0	0
79	3-methyl-4-heptanol	0.5110	3.4934	3.7120	0	1.1736	2.8611	0	1.0000	0	0
80	4-methyl-4-heptanol	0.6903	4.4672	0	2.7868	2.5347	0	2.5000	0	0	0
81	nonanol	0.0169	3.8428	0	0	9.5015	0	0	0	0	0
82	2,2-diethyl-1-pentanol	0.3729	6.3530	0	0.9738	2.8333	0	4.2500	0	0	0
83	3,5,5-trimethyl-1-hexanol	1.1401	3.6833	1.4906	3.1666	1.3611	2.2500	1.1736	0	0.2500	0
84	2-nonanol	0.3998	2.7326	2.5850	0	6.4983	1.4914	0	0	0	0
85	2-methyl-2-octanol	1.0443	3.1477	0	3.5923	5.0347	0	1.4636	0	0	0
86	2,6-dimethyl-2-heptanol	1.4133	2.3018	2.1355	3.6445	2.2500	1.3611	1.3611	0	0.0625	0
87	3-nonanol	0.2010	3.8098	1.8423	0	5.5261	2.4636	0	0	0	0
88	2,6-dimethyl-3-heptanol	1.0368	1.5894	6.7997	0	1.0000	2.8611	0	1.1736	0	0
89	3,6-dimethyl-3-heptanol	1.1277	3.4696	2.1899	2.9395	1.3611	1.3125	2.2500	0	0.1111	0
90	2,2,3-trimethyl-3-hexanol	2.1878	2.6519	0	7.1009	1.0000	0	1.6111	0	0	1.0000
91	4-nonanol	0.1414	4.0554	1.7157	0	5.3161	2.6736	0	0	0	0
92	2,6-dimethyl-4-heptanol	0.9607	1.9754	6.4836	0	0.2500	4.2222	0	0.5625	0	0
93	3,5-dimethyl-4-heptanol	0.8819	3.0196	5.8594	0	0.0625	2.7222	0	2.2500	0	0
94	4-ethyl-4-heptanol	0.4399	5.5799	0	2.0368	3.2569	0	3.5000	0	0	0
95	5-nonanol	0.1267	4.1134	1.6895	0	5.2675	2.7222	0	0	0	0
96	7-methyl-1-octanol	0.2929	3.2895	2.0281	0	6.4983	1.4914	0	0	0	0
97	decanol	0.0137	3.8721	0	0	11.0289	0	0	0	0	0
98	3,7-dimethyl-1-octanol	0.4359	3.8369	3.3054	0	3.8872	4.0400	0	0.0625	0	0
99	2-decanol	0.3920	2.7901	2.5802	0	7.9897	1.5118	0	0	0	0
100	2,7-dimethyl-3-octanol	0.9730	1.8818	6.6739	0	2.2500	3.1458	0	1.1025	0	0
101	3,6-dimethyl-3-octanol	0.9521	4.6432	1.4399	2.9170	1.7622	2.3125	2.3125	0	0.1111	0
102	3-ethyl-3-octanol	0.4563	5.5725	0	2.1045	4.7883	0	3.4236	0	0	0
103	3-ethyl-2-methyl-3-heptanol	0.9857	4.2343	2.5138	2.3770	2.6736	0.6736	2.3611	0	1.0000	0
104	4-decanol	0.1279	4.1373	1.7034	0	6.7879	2.7136	0	0	0	0
105	2,6-dimethyl-4-octanol	0.7752	3.2070	5.6727	0	0.5625	5.3733	0	0.5625	0	0
106	undecanol	0.0113	3.8958	0	0	12.5687	0	0	0	0	0

used until after a model have been developed only by using the training set. The experimental error in determining BP, the root-mean-square (rms), falls in general within the approximate range from 5.4 to 11.4 K according to ref.[11]

Descriptors. The molecular distance−edge $\boldsymbol{\mu}$ vector with 10 descriptors has a good discriminating ability for the isomers of alcohols. For the examined 106 alcohols, according to eq 1, the values of elements in the $\boldsymbol{\mu}$ vector are different from each other (see Table 2). On the other hand, the MDE or $\boldsymbol{\mu}$ vector has a good correlation with physicochemical properties of the compounds as well. The results in our previous paper[8] showed that MDE vectors of 150 alkanes can be correlated satisfactorily with eight physical properties and seven thermodynamic functions of alkanes, such as boiling points, heat capacity, molar volume, and molar refraction.

Regression Analysis. MLR was used to develop linear models that link boiling point to the $\boldsymbol{\mu}$ vector of the examined compounds. This regression equation is given by eq 2 as mentioned above. By applying MLR, the best model containing 10 variables for 106 alcohols in the working data set is developed with root-mean-square (rms) error of 3.222 °C, correlation coefficient of $R = 0.9951$, and explained variance of EV = 99.46%, which is quite satisfactory and better than those reported in previous literature.[10] This model, noted model M1, can be represented as follows:

$$\mathrm{BP} = -130.7840 + 127.1989\mu_1 + 62.7453\mu_2 + 23.2333\mu_3 - 5.6334\mu_4 + 10.4526\mu_5 - 16.4463\mu_6 - 34.3960\mu_7 - 37.3037\mu_8 - 48.3073\mu_9 - 64.8995\mu_{10} \quad (3)$$

The boiling points observed, BP_{exp}, and the boiling points calculated, BP_{M1}, from eq 3, are listed in Table 3. In the same way, the BP_{M1} data are plotted against BP_{exp}, which is given in Figure 2. From Table 3 and Figure 2, it is shown that there is no observable pattern or deviation from the normal behavior. These results indicate that the model M1 has high internal stability.

To further validate the stability of the model, a new model obtained by using 76 compounds in the training set from all 106 compounds is developed to predict the boiling points of 30 compounds in the external prediction set. The new model, noted model M2, is given as follows:

$$\mathrm{BP} = -131.4423 + 127.9648\mu_1 + 63.0511\mu_2 + 23.2638\mu_3 - 5.7659\mu_4 + 10.3165\mu_5 - 16.6673\mu_6 - 34.5442\mu_7 - 36.9215\mu_8 - 48.9297\mu_9 - 65.3866\mu_{10} \quad (4)$$

$n = 76 \quad R = 0.9946 \quad \mathrm{EV} = 99.38\% \quad \mathrm{rms} = 3.480\ ^\circ\mathrm{C}$ (estimation)

$n = 30 \quad R = 0.9964 \quad \mathrm{rms} = 2.586\ ^\circ\mathrm{C}$ (prediction)

The boiling point estimated, BP_{M2}, for 76 compounds in the training set and the boiling points predicted by using eq 4, BP_{M2}*, for 30 compounds in the external prediction set are listed in Table 3 (see column M2). The 76 estimated BP_{M2} values are plotted against the responding 76 BP_{EXP} values, given in Figure 3, and the 30 predicted BP_{M2}* values are plotted against the responding 30 BP_{exp} values, given in Figure 4. These results indicate that the model M2 not only has high internal stability but also good predictive ability for the external data.

Moreover, a cross-validation (CV) procedure is employed to predict the BP of each compound in the data set of all 106 compounds. By leaving out each observation every time, a new model, M3, is developed by using the remaining 105 observations, and the corresponding observation is predicted by model M3. The average regression coefficient

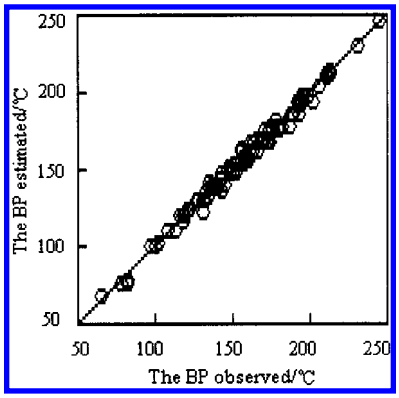

Figure 2. Plot of BP_{M1} estimated vs BP_{exp} for 106 alcohols.

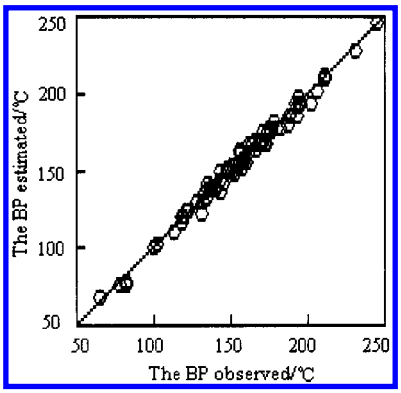

Figure 3. Plot of BP_{M2} estimated vs BP_{exp} for 76 alcohols.

of the 106 models obtained by cross-validation and their confident interval, at 95% level, are expressed in eq 5 as:

$$BP = (-130.8230 \pm 0.2985) + (127.3150 \pm 0.3076)\mu_1 + (62.7425 \pm 0.0885)\mu_3 + (23.2137 \pm 0.0546)\mu_3 + (-5.6703 \pm 0.0808)\mu_4 + (10.4560 \pm 0.0078)\mu_5 + (-16.4233 \pm 0.0652)\mu_6 + (-34.3762 \pm 0.076)\mu_7 + (-37.2954 \pm 0.0928)\mu_8 + (-48.2990 \pm 0.0986)\mu_9 + (-64.8866 \pm 0.2597)\mu_{10} \quad (5)$$

$n = 106$ $R = 0.9551 \pm 0.0003$ rms $= 3.237 \pm 0.041$
(average result of 106 estimations)

$n = 106$ $R = 0.9899$ rms $= 4.633$ °C
(prediction result)

The predicted BP_{M3} values of the 106 alcohols, each by model M3 with 106 estimates, are also listed in Table 3, and the plot of BP_{M3} by cross-validation vs BP_{exp} is given in Figure 5.

The CV results show that three boiling points (i.e. sample no. 1, no. 44, and no. 99) predicted using model M3 has observable difference from BP observed. This is because only methanol (No. 1) has C_1−OH while the other 105 alcohols all have C_2−OH or C_3−OH or C_4−OH and the μ_{10} values of 2,3,3-trimethyl-2-butanol (no. 44) and 2,2,3-trimethyl-3-pentanol (no. 90) are equal to 1.0000 while those of the other 104 alcohols all equal to zero among 106 alcohols. So, a set of models containing the first nine entries in the $\boldsymbol{\mu}$ vector can be developed by leaving out the three alcohols above, and regression equations as well as statistics are expressed as follows:

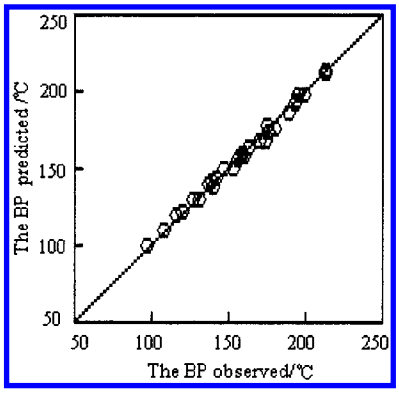

Figure 4. Plot of BP_{M2} predicted vs BP_{exp} for 30 alcohols.

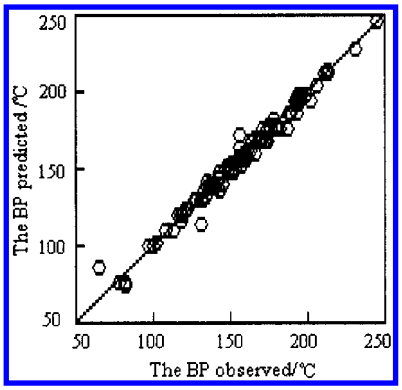

Figure 5. Plot of BP_{M3} predicted vs BP_{exp} for 106 alcohols.

$$BP = -133.8262 + 137.4218\mu_1 + 63.1098\boldsymbol{\mu}_2 + 21.7085\boldsymbol{\mu}_3 - 8.8503\boldsymbol{\mu}_4 + 10.6453\boldsymbol{\mu}_5 - 15.9456\boldsymbol{\mu}_6 - 33.5896\boldsymbol{\mu}_7 - 37.0171\boldsymbol{\mu}_8 - 47.9771\boldsymbol{\mu}_9 \quad (6)$$

$n = 103$, $m = 9$, $R = 0.9956$, EV $= 99.51\%$,
rms $= 2.990$ °C (estimation)

$$BP = -144.2952 + 141.7482\mu_1 + 66.0680\mu_2 + 22.7179\mu_3 - 8.3844\mu_4 + 10.5313\mu_5 - 17.2920\mu_6 - 36.0068\mu_7 - 38.2904\mu_8 - 51.0510\mu_9 \quad (7)$$

$n = 73$ $m = 9$ $R = 0.9960$ EV $= 99.54\%$
rms $= 2.906$ °C (estimation)

$n = 30$ $m = 9$ $R = 0.9936$ rms $= 3.420$ °C
(prediction)

$$BP = (-133.8780 \pm 0.3117) + (137.5240 \pm 0.3000)\mu_1 + (63.1112 \pm 0.0934)\mu_2 + (21.6945 \pm 0.0573)\mu_3 + (-8.8800 \pm 0.0813)\mu_4 + (10.6482 \pm 0.0093)\mu_5 + (-15.9260 \pm 0.0725)\mu_6 + (-33.5747 \pm 0.0828)\mu_7 + (-37.0137 \pm 0.0995)\mu_8 + (-47.9755 \pm 0.1041)\mu_9 \quad (8)$$

$n = 103$ $R = 0.9955 \pm 0.0004$ rms $= 3.010 \pm 0.0441$
(a average result of 106 estimations)

$n = 103$ $R = 0.9939$ rms $= 3.508$ °C
(prediction results)

The BP estimated by using eqs 6 and 7 and the BP predicted by using eqs 7 and 8 are all list in Table 3 (see columns E6, E7, and E8). It is obvious that the BP estimated and the BP predicted are similar to the BP estimated or predicted by using eq 3, 4, or 5.

Table 3. Experimental and BP estimated or predicted BP

no.	BP_{EXP}	M1	M2	M3*	E6	E7	E8*	no.	BP_{EXP}	M1	M2	M3*	E6	E7	E8*
1	64.60	68.22	68.76	86.65				54	195.30	194.26	193.69	194.16	194.36	194.34	194.28
2	78.00	76.26	76.67	76.00	77.86	76.55	77.83	55	188.00	185.87	185.58	185.77	186.13	186.39	186.04
3	97.10	99.03	99.33*	99.25	98.95	98.44	99.15	56	188.60	185.51	185.12*	185.37	185.36	185.23*	185.21
4	82.40	76.10	76.30	74.66	77.60	74.94*	75.98	57	179.50	178.44	178.38	178.36	178.92	178.97	178.87
5	117.60	120.05	120.20	120.28	119.59	119.43	119.80	58	168.50	168.71	168.81	168.73	168.86	170.23	168.91
6	108.10	109.02	109.27*	109.07	109.40	108.52	109.48	59	180.00	176.97	176.51*	176.85	176.85	176.09	176.72
7	99.50	99.62	99.81	99.63	99.83	98.36*	99.86	60	156.00	162.81	162.71	163.33	162.04	162.94*	162.56
8	82.40	77.31	77.45	75.34	78.37	77.98	75.65	61	166.10	169.33	169.76	169.45	169.37	169.53	169.50
9	138.00	139.63	139.62	139.76	139.16	139.16	139.27	62	170.00	169.30	169.17	169.26	169.64	169.10	169.62
10	131.00	129.51	129.64*	129.43	129.30	128.89	129.21	63	160.00	158.16	157.82*	157.98	157.75	158.57*	157.54
11	128.00	130.79	130.99	130.94	131.19	131.02*	131.37	64	154.50	154.98	154.98	155.02	154.57	155.18	154.58
12	119.30	120.19	120.25	120.24	119.91	118.87	119.94	65	165.50	161.79	163.25	160.90	162.24	163.36	161.43
13	116.20	119.82	119.94*	120.02	119.61	118.85	119.80	66	156.00	156.88	156.62*	157.02	157.40	158.21	157.65
14	112.90	110.03	110.80	109.63	110.37	109.33*	110.01	67	147.50	151.22	151.51	151.83	151.31	152.38*	151.94
15	102.30	102.55	102.84	102.58	102.57	102.84	102.61	68	171.00	175.83	175.44	176.02	175.53	175.14	175.71
16	157.60	158.33	158.14	158.38	157.99	158.03	158.02	69	167.50	167.59	168.00	167.59	167.03	167.21	167.01
17	153.00	149.85	149.91*	149.69	150.03	150.14*	149.88	70	163.00	162.90	163.02*	162.89	162.28	163.00*	162.22
18	151.90	148.73	148.71	148.59	148.38	148.17	148.20	71	170.00	167.74	168.24	167.54	167.59	168.14	167.38
19	149.00	149.92	149.98	149.97	150.19	150.26	150.26	72	172.00	168.36	168.38	168.08	168.57	168.48	168.31
20	147.00	149.23	149.34*	149.45	150.52	150.61	151.05	73	174.00	167.56	167.40*	167.22	167.30	166.78*	166.94
21	144.50	140.67	141.48	140.38	140.99	141.38*	140.72	74	160.00	156.54	156.50	156.03	156.96	156.88	156.50
22	143.00	136.54	136.84	135.90	135.99	137.14	135.25	75	150.50	154.30	154.60	154.93	154.12	155.12	154.72
23	140.00	139.76	139.66	139.75	139.38	138.53	139.36	76	156.50	153.02	152.59	151.58	152.95	152.85	151.46
24	136.50	140.01	140.52*	140.46	140.56	141.57*	141.10	77	176.30	175.08	174.69*	175.01	174.54	174.22*	174.44
25	135.00	138.77	138.74	138.94	138.32	137.81	138.49	78	164.00	167.15	167.10	167.29	166.78	166.39	166.91
26	134.30	131.35	132.09	131.21	131.59	131.35	131.46	79	162.00	167.56	168.06	168.04	167.30	167.86	167.77
27	131.60	130.33	130.52	130.22	130.39	129.25	130.28	80	161.00	162.12	162.27	162.20	161.30	162.00*	161.34
28	126.50	129.77	130.50*	129.95	129.58	129.34*	129.75	81	213.30	211.80	211.04*	211.61	212.16	212.05	212.01
29	122.40	124.46	124.86	124.59	124.31	124.77	124.44	82	192.00	193.22	193.64	193.77	197.14	196.94	202.78
30	121.10	123.43	123.63	123.62	122.81	123.43	122.96	83	193.00	186.92	186.87	186.19	186.83	188.38	186.09
31	120.40	121.52	121.33*	121.63	121.74	121.52*	121.88	84	193.50	194.98	194.33*	195.08	195.08	194.29*	195.18
32	118.40	116.94	116.72	116.78	117.07	116.90	116.92	85	178.00	181.60	181.33	181.94	180.97	181.90	181.27
33	176.40	176.46	176.09	176.47	176.33	176.35	176.32	86	173.00	173.79	173.66	173.88	173.29	174.03	173.32
34	173.00	168.17	168.06	167.96	168.35	168.55	168.14	87	195.00	193.88	193.30	193.82	193.77	193.35*	193.70
35	170.00	167.33	167.13*	167.22	167.04	166.90*	166.90	88	175.00	178.42	178.93*	178.70	178.15	178.27	178.41
36	169.00	168.09	167.98	168.04	168.29	168.52	168.25	89	173.00	174.56	174.63	174.68	174.29	174.84	174.39
37	164.00	168.61	168.49	168.81	168.94	169.10	169.16	90	156.00	164.03	164.06	172.95			
38	159.00	159.70	159.90*	159.75	160.18	160.20*	160.27	91	192.50	193.12	192.54*	193.16	192.74	192.41	192.76
39	159.00	158.63	158.35	158.62	158.34	157.57	158.31	92	174.50	178.19	178.32	178.58	178.13	177.77*	178.51
40	143.00	143.46	143.53	143.50	142.67	143.48	142.64	93	187.00	178.94	180.31	176.07	179.10	180.76	176.27
41	151.00	149.69	149.64	149.58	149.54	148.56*	149.42	94	182.00	177.47	177.62	176.64	177.85	177.91	177.07
42	139.70	138.57	138.36*	138.47	138.46	139.01	138.35	95	193.00	192.97	192.40*	192.97	192.52	192.21*	192.49
43	133.10	135.65	135.72	135.85	135.56	135.89	135.75	96	206.00	203.39	202.81	203.23	203.45	203.27	203.29
44	131.00	122.97	122.94	114.05				97	231.10	229.20	228.23	228.86	229.83	229.62	229.58
45	157.00	157.47	157.27*	157.49	157.03	156.62	157.03	98	212.50	214.06	213.61*	214.20	214.62	214.77*	214.83
46	143.00	148.72	149.30	148.97	148.20	148.25*	148.47	99	211.00	212.74	211.89	212.93	213.08	212.22	213.32
47	143.00	143.72	144.00	143.77	143.19	143.81	143.20	100	193.50	196.77	197.05	197.06	196.50	196.65	196.78
48	148.00	148.90	149.00	148.94	148.71	148.15	148.74	101	202.20	194.16	194.15	193.47	194.36	195.32	193.69
49	142.00	143.05	143.49*	143.16	143.60	143.79*	143.77	102	199.00	197.34	197.30*	197.06	197.91	198.06*	197.72
50	135.00	141.57	141.37	142.08	141.11	141.62	141.61	103	193.00	192.64	192.31	192.57	192.82	192.80	192.79
51	139.70	138.40	138.30	138.30	138.24	138.34	138.13	104	210.50	210.98	210.21	211.03	210.82	210.44	210.85
52	138.70	140.41	141.90*	140.90	140.32	140.76	140.78	105	195.00	197.37	197.41*	197.69	197.73	197.81*	198.12
53	156.00	156.86	156.66	156.90	156.27	155.90*	156.28	106	245.00	246.47	245.30	245.16	247.39	247.06	247.00

"*"refer to the BP predicted.

If a set of models is constructed by employing 106 alcohols, relative statistics are listed as follows:

For working set: $n = 106$ $m = 9$ $R = 0.9893$
EV = 98.83% rms = 4.77 ppm

For calibration set: $n = 76$ $m = 9$ $R = 0.9887$
EV = 98.71% rms = 5.025 ppm

For prediction set: $n = 30$ $R = 0.9902$ rms = 4.236

Cross-validation: $R = 0.9893 \pm 0.0003$
rms = 4.7799 ± 0.0335 (average result)

$n = 106$ $R = 0.9827$ rms = 6.059
(prediction result)

The results above show that various nine parameter models have lower estimation abilities and prediction capacities than model M1, M2, or M3.

So, the eq 6 and 7 may be employed to predict the BP of compounds not including μ_{10}, while the models M1, M2, and M3 having 10 parameters must be used for the compounds containing μ_{10}.

ACKNOWLEDGMENT

The project was partly supported by the State Educational Comission Foundation (SECF), Ministry of Mechanical Industrial Fund (MMIF), and an award by Fok Ying-Tung Educational Foundation (HYTF).

REFERENCES AND NOTES

(1) Wiener, H. Structure Determination of Paraffin Boiling Points. *J. Am. Chem. Soc*. **1947**, *69*, 17−20.

(2) Hosoya, H. Topological Index: A Newly Proposed Quantity Characteristizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons. *Bull. Chem. Soc. Jpn*. **1971**, *44*, 2332−2339.

(3) Randic, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc*. **1975**, *97*, 6609−6615.

(4) Balaban, A. T. Topological Indexes Based on Topological Distance in Molecular Graphs. *Pure Appl. Chem.* **1983**, *55*, 199−206.

(5) Randic, M, Generalized Molecular Descriptors. *J. Math. Chem*. **1991**, *7*, 155−168.

(6) Liu, S.; Cao, C.; Li, Z. Approach to Estimation and Prediction for Normal Boiling Point (NBP) of Alkanes Based on a Novel Molecular Distance−Edge (MDE) Vector. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 378−392.

(7) Liu, S.; Cao, C.; Li, Z. A Novel Molecular Distance−Edge (*MDE*) Vector () and Thermodynamical Properties of Alkanes. *Chem. Hong Kong*, **1998**, *2*, 113−123.

(8) Liu, S.; Xia, Z.; Liu, H.; Cao, C.; Li, Z. Estimation and Prediction of 15 Basic Physical Properties of Alkanes By Using A Novel Molecular Distance−Edge Vector Called μ Vector. Submitted to *Acta Chim. Sinica (in Chinese)*.

(9) Aylward, G. H.; Findlay, T. J. V. *SI Chemical Data*; 2nd ed.; John Wiley & Sons: New York, 1982.

(10) Li, Z.; Cao, C. Effective Length of Carbon Chain and Boiling Points of Fatty Alcohols. *Chin. J. Org. Chem*. (*in Chinese*), **1998**, *18* (4): 360−364.

(11) Wessel, M. D.; Jurs, P. C. Prediction of Normal Boiling Points for a Diverse Set of Industrially Important Organic Compounds from Molecular Structure. *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 841−850.

(12) Nikolic, S.; Trinajstic, N.; Mihalic, Z. Molecular Topological Index: An Extension to Heterosystems. *J. Math. Chem.* **1993**, **12** (1−4): 251−264.

(13) Lai, W. Y.; Chen, D. H.; Maddox, R. N. Application of Group-Contribution Model to the Prediction of Physical Constants. 1. Predicting Normal Boiling Points with Molecular Structure. *Ind. Eng. Chem. Res*. **1987**, *26* (6): 1072−1079.

(14) Anbrose D. Correlation of the Boiling Points of Alkanols. *J. Appl. Chem. Biotecnol*. **1976**, *26* (12): 712−714.

CI990011F