1504 Bull. Chem. Soc. Jpn. Vol. 80, No. 8, 1504–1510 (2007)

Molecular Descriptors of Topology and a Study on Quantitative Structure and Property Relationships

Congyi Zhou*[1,2] **and Changming Nie**[1,3]

[1]School of Chemistry and Chemical Engineering, Nanhua University, Hengyang, 421001, P. R. China

[2]Department of Chemistry, University of Montana, Missoula, 59812, U.S.A.

[3]Department of Applied Chemical, School of Science and Engineering, Waseda University,
3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555

Received October 31, 2006; E-mail: kentchowzcy@163.com

The novel topological descriptor *PE* was proposed, using a revised distance matrix involving relative bond length as well as equilibrium electronegativity. Along with *PE*, the quantitative structure–property relationships were included with path number P_3. The analyzed models could be applicable to the condensed-type properties of hydrocarbons. Moreover, the models also were applicable to the Kováts retention indexes under different conditions.

Quantitative structure–property/retention relationships (QSPR/QSRR) have gained lots of attention,[1–17] because of great convenience and importance for practical use and molecular design science the physicochemical properties and biological activities can be predicted from the relationship between the properties and the structures of the compounds. Among them, developing topological indexes, which refers to numerical descriptors that are derived from molecular graphs of chemical compounds, is a promising approach. In 1947, the first topological index *W* was proposed by Wiener,[18] which can be used to characterize molecular structures, and since then, many topological indexes have been suggested, such as Kier–Hall's molecular connectivity index,[19,20] Hosoya index,[21] Balaban index,[22] Bonchev's index,[23] Estrada index,[24] Randic index,[25] and Xu index.[26] To date, the third generation of topological indexes has been regarded as vertex invariants expressed by real number.[19,27]

Based on our previous work,[11–16] which includes new topological descriptors, such as the *OET* index,[11] the *AEI* index,[14] and the *QTI* index,[15] in the present study, an equilibrium-electronegative index *PE* is proposed on the basis of the distance matrix with entries employing relative bond lengths and equilibrium electronegativities. Along with the path number P_3 and $(PE)^{1/2}$, the study was extended the condensed-type properties and Kováts retention indexes (I_R) of hydrocarbons.

Principle and Method

Group Electronegativities. As one of the main properties of atoms, electronegativities represent the ability of atoms to gain or lose electrons. The larger the electronegativity of an atom, the stronger the ability of the atom to attract electrons. Based on Pauling electronegativities, group electronegativities can be calculated by step-wise addition.

$$\chi_0 = \frac{1}{n_{1l}}\sum_{l=1}^{n_{1l}}\chi_{1l} \quad \text{the equilibrium of the first level (1 in Fig. 1)}$$

$$\chi_{1l} = \frac{1}{n_{2l}}\sum_{l=1}^{n_{2l}}\chi_{2l} \quad \text{the equilibrium of the second level (2 in Fig. 1)}$$

$$\vdots \qquad \vdots$$

$$\chi_{(k-1)l} = \frac{1}{n_{kl}}\sum_{l=1}^{n_{kl}}\chi_{kl} \quad \text{the equilibrium of the } k\text{th level (} k \text{ in Fig. 1)}$$

$$\vdots \qquad \vdots$$

$$\chi_G = \left\{\frac{1}{n_{1l}}\sum_{l=1}^{n_{1l}}\left[\frac{1}{n_{2l}}\sum_{l=1}^{n_{2l}}\cdots\left(\frac{1}{n_{kl}}\sum_{l=1}^{n_{kl}}\chi_{kl}\right)\cdots\right]\right\}, \tag{1}$$

where $n_{1l}, n_{2l}, \ldots, n_{kl}$ are the sum of atom or group directly attached to the ground atom, which is the left atom next to the dotted line labeled $1, 2, 3, \ldots, k, \ldots$ of each level in Fig. 1, and $\sum_{l=1}^{n_{1l}}\chi_{1l}, \sum_{l=1}^{n_{2l}}\chi_{2l}, \ldots, \sum_{l=1}^{n_{kl}}\chi_{kl}$ are the sum of electronegativity of atom or group directly attached to the ground atom. Some groups and their electronegativities are listed in Table 1.

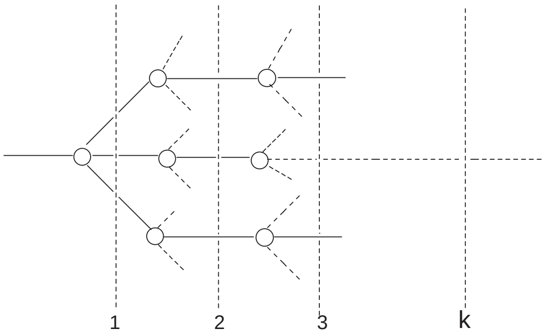

Fig. 1. Plot of group structure.

Published on the web August 10, 2007; doi:10.1246/bcsj.80.1504

Table 1. Some Groups and Their Electronegativities

Groups	χ_G	Groups	χ_G	Groups	χ_G	Groups	χ_G
$-CH_3$	2.2875	$-CH_2CH_3$	2.3094	$-CH{=}CH_2$	2.3556	$-C{\equiv}CH$	2.4625
–CN	2.7950	–NCO	3.0100	–OCN	3.1175	$-C_6H_5$	2.4333
–CHO	2.7300	–COOH	2.9467	$-SiH_3$	2.1250	$-NH_2$	2.4800
–NO	3.2400	$-NO_2$	3.3067	–OPh	2.9367	$-OCH_3$	2.8638
–OH	2.8200	$-CCl_3$	3.0075	–COMe	2.7592	–COPh	2.8078

We presumed that the electronegativities change in the formation of a molecule. As long as a molecule is formed, the electronegativities of atoms in the molecule are fixed, that is, the electronegativities are in the state of equilibrium, which is called the equilibrium electronegativities. The definition of the equilibrium electronegativities for atom i is the following:

$$\chi_i = \left(\chi_{iA} + \sum \chi_G\right) \Big/ \left(1 + \sum l\right), \tag{2}$$

where χ_{iA} is the Pauling electronegativity of atom i, χ_G is the electronegativities of group(s) directly connected to atom i calculated by Eq. 1, and l is the sum of group(s) directly connected to atom i. Taking into account the equilibrium electronegativities, we proposed the index PE.

Calculations of Topological Index *PE*. Assume that a graph $G = \{V, E\}$ is a hydrogen-suppressed graph with n vertices of atoms. For the two sets V and E, the former represents the atoms of a molecule and the latter symbolizes the covalent bonds between pairs of atoms. Theoretically, such a graph depicts the topology of chemical species. Invariants derived from the graphs can be used to characterize chemical structure.

The distance matrix D of n atoms in a molecule, which is a square symmetric matrix, can be given as $D = [d_{ij}]_{n \times n}$, where d_{ij} are the length of the shortest path between the vertices i and j in a G. Instead, in this paper, d_{ij} was revised by using the relative bond length, which can be expressed as:

Fig. 2. Description of the hydrogen-suppressed graph of 1,2-butadiene.

$$d_{ij} = \sum L / L_{C\text{–}C}, \tag{3}$$

where $\sum L$ is the sum of bond length between the vertices i and j, and $L_{C\text{–}C}$ is the bond length of carbon–carbon single bond. For carbon–carbon double bond and carbon–carbon triple bond, the revised d_{ij} is 0.134/0.154 and 0.120/0.154, respectively. Clearly, the relative bond length can distinguish saturated and unsaturated bonds. Then, the revised distance matrix D_1 is obtained by adding equilibrium electronegativities into the diagonal of D. Based on D_1, the new topological index PE is defined as:

$$PE = \log \lambda_{\max}, \tag{4}$$

where $\lambda_{\max}$ is the maximum eigenvalue of matrix D_1.

For instance, the hydrogen-suppressed graph and D_1 of 1,2-butadiene are given in Fig. 2 and Eq. 5.

$$D_1 = \begin{bmatrix} \chi_1 & L_{12}/L_{C\text{–}C} & (L_{12}+L_{23})/L_{C\text{–}C} & (L_{12}+L_{23}+L_{34})/L_{C\text{–}C} \\ L_{12}/L_{C\text{–}C} & \chi_2 & L_{23}/L_{C\text{–}C} & (L_{23}+L_{34})/L_{C\text{–}C} \\ (L_{12}+L_{23})/L_{C\text{–}C} & L_{23}/L_{C\text{–}C} & \chi_3 & L_{34}/L_{C\text{–}C} \\ (L_{12}+L_{23}+L_{34})/L_{C\text{–}C} & (L_{23}+L_{34})/L_{C\text{–}C} & L_{34}/L_{C\text{–}C} & \chi_4 \end{bmatrix}$$
$$= \begin{bmatrix} 2.3495 & 0.8701 & 1.7402 & 2.7402 \\ 0.8701 & 2.4042 & 0.8701 & 1.8701 \\ 1.7402 & 0.8701 & 2.3677 & 1.0000 \\ 2.7402 & 1.8701 & 1.0000 & 2.3089 \end{bmatrix}. \tag{5}$$

According to the method introduced, PE for 1,2-butadiene was 0.8482. PE for some alkanes are presented in Table 2.

Path Number P_3. The path number P_3 was defined as the sum of 3-bond paths between all pairs of non-hydrogen atoms in a molecular introduced by Wiener, which characterizes the shape of molecules.[18] For example, P_3 for 1,2-butadiene is 1.

Results

Model Development. The normal boiling point (T_B) is a physical property universally and precisely measured for low molecular weight compounds. Consequently, T_B is used to test the performances of the topological indices as a starting point. Interestingly, both Figure 3 and Figure 4, which are plots of the T_B and PE versus the number of carbon atoms of alkanes, respectively, show the same regularity. Therefore, the functional relationship between T_B and PE may be expressed as the following:

$$T_B = a_1 PE + b_1, \tag{6}$$

where a_1 is the contribution coefficient of PE and b_1 is a constant. The regression result proves the rationality of our assumption (see Table 3). In that properties of alkanes are closely connected with the shape of molecules, P_3 takes part in the modeling as well. If the square root of PE is included, the quality is further improved. In determining the quality of these models, the correlation coefficient (R), standard error (S) and F

Table 2. Topological Indexes *PE* and Path Number P_3 for Alkanes

No.	Compound	*ET*	P_3	No.	Compound	*ET*	P_3
1	Ethane	0.5169	0	38	2,3,4-Trimethylpentane	1.2814	8
2	Propane	0.7013	0	39	2,2,3,3-Tetramethylbutane	1.2367	9
3	Butane	0.8728	1	40	Nonane	1.4778	6
4	2-Methylpropane	0.8417	0	41	2,2-Dimethylheptane	1.4195	6
5	Pentane	1.0249	2	42	2,3-Dimethylheptane	1.4127	8
6	2-Methylbutane	0.9895	2	43	2,4-Dimethylheptane	1.4125	7
7	2,2-Dimethylpropane	0.9497	0	44	2,5-Dimethylheptane	1.4189	7
8	Hexane	1.1587	3	45	2,6-Dimethylheptane	1.4329	6
9	2-Methylpentane	1.1260	3	46	3,3-Dimethylheptane	1.3976	8
10	3-Methylpentane	1.1155	4	47	3,4-Dimethylheptane	1.3977	9
11	2,2-Dimethylbutane	1.0783	3	48	3,5-Dimethylheptane	1.4044	8
12	2,3-Dimethylbutane	1.0900	4	49	4,4-Dimethylheptane	1.3905	8
13	Heptane	1.2771	4	50	3,3-Diethylpentane	1.3560	12
14	2-Methylhexane	1.2482	4	51	3-Ethylheptane	1.4217	8
15	3-Methylhexane	1.2347	5	52	4-Ethylheptane	1.4147	8
16	3-Ethylpentane	1.2202	6	53	3-Ethyl-2,2-dimethylpentane	1.3542	10
17	2,2,3-Trimethylbutane	1.1674	6	54	3-Ethyl-2,3-dimethylpentane	1.3457	12
18	2,2-Dimethylpentane	1.2025	4	55	3-Ethyl-2,4-dimethylpentane	1.3628	10
19	2,3-Dimethylpentane	1.2025	6	56	3-Ethyl-2-methylhexane	1.3896	9
20	2,4-Dimethylpentane	1.2121	4	57	3-Ethyl-3-methylhexane	1.3737	10
21	3,3-Dimethylpentane	1.1869	6	58	4-Ethyl-3-methylhexane	1.3702	10
22	Octane	1.3828	5	59	4-Ethyl-2-methylhexane	1.3967	8
23	2,2-Dimethylhexane	1.3164	5	60	2-Methyloctane	1.4561	6
24	2,3-Dimethylhexane	1.3115	7	61	3-Methyloctane	1.4428	7
25	2,4-Dimethylhexane	1.3160	6	62	4-Methyloctane	1.4366	7
26	2,5-Dimethylhexane	1.3308	5	63	2,2,3-Trimethylhexane	1.3717	9
27	3,3-Dimethylhexane	1.2956	7	64	2,2,4-Trimethylhexane	1.3787	7
28	3,4-Dimethylhexane	1.3006	8	65	2,2,5-Trimethylhexane	1.3940	6
29	3-Ethylhexane	1.3240	7	66	2,3,3-Trimethylhexane	1.3637	10
30	3-Ethyl-2-methylpentane	1.2950	8	67	2,3,4-Trimethylhexane	1.3712	10
31	3-Ethyl-3-methylpentane	1.2782	9	68	2,3,5-Trimethylhexane	1.3869	8
32	2-Methylheptane	1.3577	5	69	2,4,4-Trimethylhexane	1.3709	8
33	3-Methylheptane	1.3438	6	70	3,3,4-Trimethylhexane	1.3546	11
34	4-Methylheptane	1.3394	6	71	2,2,3,3-Tetramethylpentane	1.3254	12
35	2,2,3-Trimethylpentane	1.2701	8	72	2,2,3,4-Tetramethylpentane	1.3431	10
36	2,2,4-Trimethylpentane	1.2865	5	73	2,2,4,4-Tetramethylpentane	1.3510	6
37	2,3,3-Trimethylpentane	1.2642	9	74	2,3,3,4-Tetramethylpentane	1.3347	12

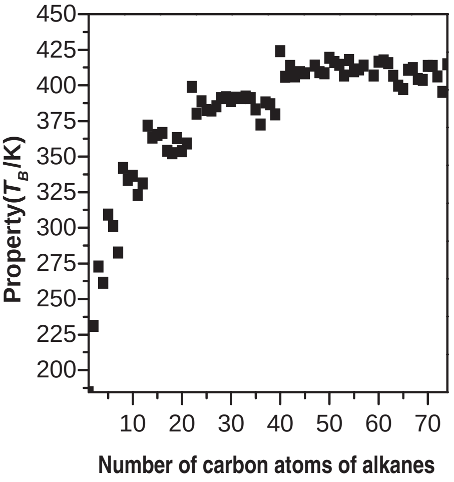

Fig. 3. Scatter plot of number of carbon atoms of alkanes versus property T_B.

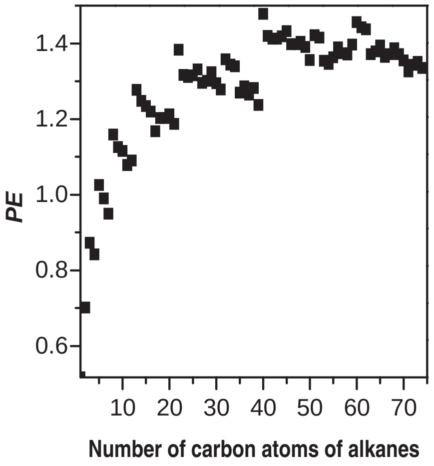

Fig. 4. Scatter plot of number of carbon atoms of alkanes versus *PE*.

Table 3. Statistical Results of the Models and the Quality of Their Correlations with T_B

Model	Parameter	*R*	*S*	*Q*	*N*	*F*
1	*PE*	0.9840	8.3012	0.1185	74	2202.31
2	*PE*, P_3	0.9991	1.9975	0.5002	74	19603.97
3	*PE*, P_3, $(PE)^{1/2}$	0.9993	1.7955	0.5566	74	16182.03

Table 4. Statistical Results of Regression for the Condensed-Type Properties of Alkanes

No.	*P*	$a(t_1)$[a)]	$b(t_2)$[a)]	$c(t_3)$[a)]	*d*	*R*	*S*	*N*	*F*	*MRE* /%
1	T_B	142.9899 (8.6219)	142.0095 (4.2282)	4.03754 (38.3279)	10.0272	0.9993	1.7955	74	16182.03	0.37
2	T_c	−168.1930 (−5.7080)	801.0220 (13.4274)	7.3188 (38.8633)	−183.7580	0.9985	3.1820	72	7521.81	0.40
3	$\Delta_{vap}H_m$	35.2998 (8.8302)	−27.8502 (−3.4411)	0.2515 (9.8424)	16.9197	0.9954	0.4317	72	2450.25	1.10
4	P_c	−2.1250 (−0.3946)	−53.2900 (−4.8878)	0.2588 (7.5207)	88.2124	0.9919	0.5815	72	1376.49	1.53
5	n_D	−1.8699 (−13.6592)	4.2529 (15.3892)	0.0025 (2.7021)	−1.0325	0.9608	0.0146	70	264.06	0.62
6	γ	−47.9578 (−9.8449)	129.4194 (13.1402)	0.4235 (13.1526)	−67.7585	0.9905	0.5212	71	1153.00	2.26
7	*D*	−1.3876 (−25.656)	3.3596 (30.6851)	0.0071 (20.7667)	−1.3688	0.9960	0.0058	73	2853.15	0.58
8	η	15.1740 (0.8410)	−112.8470 (−3.1239)	0.1519 (1.1394)	166.8084	0.9902	1.3473	32	470.22	1.52
9	*K*	−0.0171 (−1.9443)	0.0131 (0.7188)	−0.0002 (−4.7110)	0.0190	0.9919	0.0004	30	529.19	2.46

a) t_1, t_2, and t_3 mean the *t*-values of *PE*, its squared term and P_3, respectively.

Table 5. The Results of Leave-One-Out Cross Validation and Prediction

	Cross validation of the training set				Predication of the test set			
Property	*R*	*N*	*RMS*	*MRE*/%	*R*	*N*	*RMS*	*MRE*/%
T_B	0.9991	48	2.6208 K	0.53	0.9996	26	1.6463 K	0.38
T_c	0.9989	48	2.9692 K	0.41	0.9983	24	4.1849 K	0.52
$\Delta_{vap}H_m$	0.9955	48	0.5107 kJ mol^{-1}	1.36	0.9959	24	0.4244 kJ mol^{-1}	1.07
P_c	0.9963	48	0.4766 bar	1.34	0.9798	24	0.8181 bar	2.23

value (*F*) can play important roles. Mihalic and Trinajstic have suggested that a good QSPR model for T_B must have $R > 0.99$ and $S < 5.0\,^{\circ}\mathrm{C}$.[28] Moreover, the quality factor (*Q*) is used to measure the predictive power of the model.[29] *Q* is defined as the ratio of *R* to *S*, that is, $Q = R/S$. Hence, the larger the value of *R*, the smaller the value of *S*, the higher *Q* will be, and the better will be the predictive power of the model. Clearly, the comparison is in favor of model 3, which can be expressed as:

$$Property = aPE + b(PE)^{1/2} + cP_3 + d, \tag{7}$$

where *a*, *b*, and *c* are the contribution coefficients of *PE*, $(PE)^{1/2}$, and P_3, respectively, and *d* is a constant.

Consequently, the tri-parameter regression in Eq. 7 is used to study the other condensed-typed properties of alkanes,[30] which are critical temperature (T_c), enthalpy of vaporization (ΔH_{vap}), critical pressure (P_c), refractive index (n_D), surface tension (γ), density at 25 °C (*D*), viscosity of a gas at 25 °C (η), and thermal conductivity of a gas (*K*). The statistical results in Table 4 show that Eq. 7 is a good model for the condensed-type properties of alkanes.

Model Validation. To validate the quality of the models in Table 4, the sample was divided into training set and test set by random sampling. For the training set, the leave-one-out method was used to perform the cross validation. Each time one compound is left out from the training set, and then the model based on the others is used to predict the compound extracted, that is, a model is built with $N - 1$ compounds and the *N*th compound is predicted. For the test set, the predicted values are obtained from the model using the whole training set. The quality of the model can be estimated by the correlation coefficient (*R*), the root-mean-square error (*RMS*), and the mean relative error (*MRE*).

$$RMS = \sqrt{\frac{\sum (y_{pre} - y_{exp})^2}{n}}, \tag{8}$$

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_{pre} - y_{exp}|}{y_{exp}} \times 100\%, \tag{9}$$

where y_{pre}, y_{exp}, and *n* are the predicted values, experimental values, and the number of the sample, respectively. Because the sample number in this paper is large, just a few of the validation results are listed. From Table 5, one can see that the quality of the models for the four properties is satisfactory. For T_B, both the results of validation and prediction meet the requirements stated by Mihalic and Trinajstic. Next, using the leave-one-out method employed, the frequency of *R* for the whole sample was investigated. For example, Figure 5 shows that *R* of *D* is distributed normally, which shows the stability of the model. For the properties not listed, all the correlation coefficients of cross validation were close to the corresponding ones in Table 4.

Application to Alkenes and Alkynes. To validate the application to alkenes and alkynes, their indexes were also computed, and some of them are listed in Table 6. The relationship

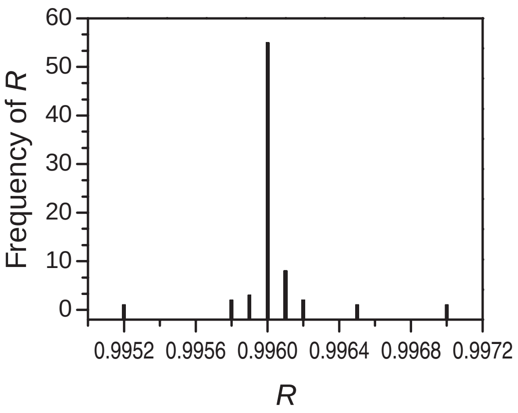

Fig. 5. The distribution plot of correlation coefficient (R) versus frequency of R for density of alkanes.

between the properties and the indexes were investigated using Eq. 7. One can observe that the method also has good applicability to alkenes and alkynes. It must be mentioned that R for the T_B of alkenes was 0.9996, which is close to 1, and R^2 was 0.9991, which indicates that the model explains more than 99.9% of the variance in the experimental values of T_B for these alkenes. The other results validate wide application of this method as well (Tables 7 and 8).

Correlation with Kováts Retention Indexes. In addition, Eq. 7 can also be used to correlate the Kováts retention indexes (I_R) of alkanes on different stationary phases of squalane at 30, 50, and 70 °C, H-P PONA at 60 °C and J&W DB-5 at 60 °C.[31] The statistical results in Table 9 show that the modeling was successful, and the R values are greater than 0.994. Our study shows that the method has wide application to I_R of other various stationary phases at different temperatures. Due to the limitation of the length of the paper, only a few are listed.

Table 6. PE and P_3 for Alkenes and Alkynes (Including Dienes)

No.	Compound	PE	P_3	No.	Compound	PE	P_3
1	Ethene	0.5034	0	27	1-Tetradecene	1.8433	11
2	Propene	0.6878	0	28	1-Pentadecene	1.9019	12
3	1-Butene	0.8616	1	29	1-Hexadecene	1.9569	13
4	1-Pentene	1.0160	2	30	1-Heptadecene	2.0086	14
5	2-Methyl-1-butene	0.9806	2	31	1-Octadecene	2.0575	15
6	3-Methyl-1-butene	0.9800	2	32	1-Nonadecene	2.1039	16
7	1-Hexene	1.1517	3	33	1-Icosene	2.1479	17
8	2-Methyl-1-pentene	1.1189	3	34	Acetylene	0.4989	0
9	3-Methyl-1-pentene	1.1078	3	35	Propyne	0.6810	0
10	4-Methyl-1-pentene	1.1182	3	36	Butyne	0.8553	1
11	2,3-Dimethyl-1-butene	1.0825	4	37	Propadiene	0.6752	0
12	3,3-Dimethyl-1-butene	1.0699	3	38	3-Methyl-1-butyne	0.9742	2
13	2-Ethyl-1-butene	1.1086	4	39	1,2-Butadiene	0.8482	1
14	1-Heptene	1.2715	4	40	1,3-Butadiene	0.8503	1
15	2,3,3-Trimethyl-1-butene	1.1607	6	41	1-Pentyne	1.0107	2
16	2-Ethyl-1-pentene	1.2293	5	42	2-Pentyne	1.0050	2
17	3-Ethyl-2-pentene	1.2108	6	43	1-Hexyne	1.1474	3
18	1-Octene	1.3781	5	44	2-Hexyne	1.1435	3
19	2,4,4-Trimethyl-1-pentene	1.2811	5	45	3-Hexyne	1.1402	3
20	2,4,4-Trimethyl-2-pentene	1.2758	5	46	1-Heptyne	1.2679	4
21	2-Ethyl-1-hexene	1.3394	6	47	1-Nonyne	1.4715	6
22	1-Nonene	1.4740	6	48	1-Decyne	1.5587	7
23	1-Decene	1.5608	7	49	1-Undecyne	1.6383	8
24	1-Undecene	1.6401	8	50	1-Dodecyne	1.7115	9
25	1-Dodecene	1.7131	9	51	1-Tridecyne	1.7792	10
26	1-Tridecene	1.7805	10	52	1-Pentadecyne	1.9008	12

Table 7. Statistical Results of Regression for the Condensed-Type Properties of Alkenes

No.	P	$a(t_1)$[a)]	$b(t_2)$[a)]	$c(t_3)$[a)]	d	R	S	N	F	MRE /%
1	T_B	187.2972 (6.9105)	78.42632 (1.6449)	4.2518 (5.4668)	25.5550	0.9996	3.5357	33	10990.07	0.80
2	T_c	−155.7670 (−4.0830)	834.0023 (12.4273)	6.6660 (6.0892)	−227.6790	0.9992	4.9768	33	5952.81	0.72
3	P_c	2.6446 (0.3218)	−62.7297 (−4.3390)	0.2175 (0.9224)	94.6060	0.9942	1.0721	33	832.45	1.99
4	$\Delta_{vap}H_m$	36.5022 (8.1487)	−29.2505 (−3.7120)	0.3294 (2.5624)	16.9223	0.9989	0.5844	33	4272.86	1.51
5	D	−0.9941 (−17.3173)	2.4276 (21.9228)	0.0115 (8.0317)	−0.8232	0.9967	0.0056	30	1290.08	0.55

a) t_1, t_2, and t_3 mean the t-values of PE, its squared term and P_3, respectively.

Table 8. Statistical Results of Regression for the Condensed-Type Properties of Alkynes

No.	P	$a(t_1)$[a)]	$b(t_2)$[a)]	$c(t_3)$[a)]	d	R	S	N	F	MRE /%
1	T_B	87.4729 (0.3347)	216.9084 (0.5455)	6.1289 (0.6922)	−0.7638	0.9978	7.2581	19	1116.35	1.74
2	T_c	249.3782 (0.4831)	226.2003 (0.2880)	−9.3584 (−0.5351)	31.2857	0.9932	14.3345	19	364.17	2.12
3	P_c	42.1089 (0.7516)	−137.5980 (−1.6143)	−1.1675 (−0.6151)	138.8593	0.9943	1.5559	19	434.48	2.44

a) t_1, t_2, and t_3 mean the t-values of PE, its squared term and P_3, respectively.

Table 9. Statistical Results of Regression for Kováts Retention Indexes of Alkanes on Different Stationary Phases at Different Temperatures

No.	P	$a(t_1)$[a)]	$b(t_2)$[a)]	$c(t_3)$[a)]	d	R	S	N	F	MRE/%
1	Squalane	1588.0620	−2037.6800	15.8153	895.3336	0.9945	11.4795	67	1888.44	1.09
	(30 °C)	(3.6814)	(−2.1131)	(21.4186)						
2	Squalane	1560.0920	−1993.5400	16.5004	878.7707	0.9948	11.1811	67	2012.35	1.06
	(50 °C)	(3.7131)	(−2.1225)	(22.9429)						
3	Squalane	1544.2840	−1976.3300	17.1938	877.0268	0.9951	10.9037	67	2140.32	1.02
	(70 °C)	(3.7690)	(−2.1577)	(24.5150)						
4	H-P PONA	3244.3200	−5717.7900	16.1756	2934.5990	0.9953	8.4837	52	1676.51	0.92
		(4.7698)	(−3.6974)	(24.7281)						
5	J&W DB-5	3266.1770	−5782.0800	16.6826	2975.9440	0.9952	8.6071	52	1638.85	0.94
		(4.7331)	(−3.6854)	(25.1376)						

a) t_1, t_2, and t_3 mean the t-values of PE, its squared term and P_3, respectively.

Discussion

An important attribute for measuring the quality of a topological index is the ability to distinguish the structures of compounds. In order to achieve that, Razinger et al. quantitatively described the degree of degeneracy by $C(s)$, which is the selectivity coefficient of topological indexes for the structures of compounds.[32] The definition of $C(s)$ is given as:

$$C(s) = N_{\mathrm{val}}/N_{\mathrm{str}}, \quad (10)$$

where N_{val} is the number of isomers that can be structurally distinguished by indexes, and N_{str} is the number of given isomers. Figure 6 illustrates that the structural selectivity of PE is better than the other important indexes. In the case of the same N_{str}, the larger $C(s)$ is, the better the index. For W index, when N_{str} is 9, its $C(s)$ is 7/9, which means that the index can only distinguish structures of seven isomers. For Randic index, when N_{str} is 75, its $C(s)$ is 49/75, which means that the index can only distinguish the structures of 49 isomers. However, the $C(s)$ of PE was 1 when N_{str} was 355. The $C(s)$ for the other five indexes are less than 1 in the range of N_{str} ranging from 1 to 355, which indicates the superiority of PE.[33]

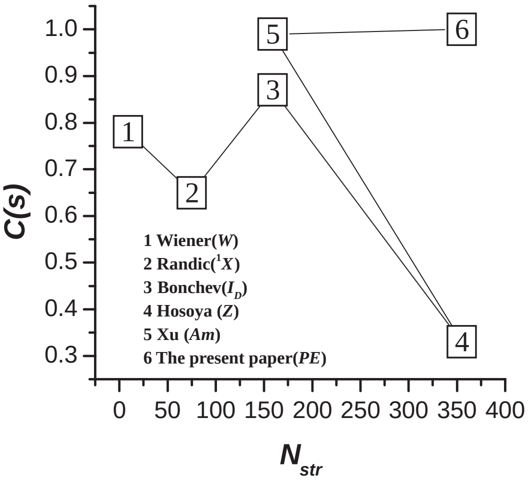

Fig. 6. The plot of N_{str} versus $C(s)$ of alkanes for six important topological indexes.

Through computation, we found that PE could not only differentiate hydrocarbons, not including cis–trans-isomers, but also organic compounds containing hetero atoms or/and multiple bonds. The good structural selectivity of PE comes from the fact that equilibrium electronegativities characterize the electronegativities of each atom, which are closely connected to their chemical environment. Therefore, equilibrium electronegativities reflect the chemical information at the atomic and group levels. For example, the equilibrium electronegativities of each vertex atom for o-bromochlorobenzene was 2.67830, 2.64500, 2.46500, 2.42275, 2.42278, 2.47610, 2.83890, and 2.75000, which is in the accordance with the fact that chemical environment of each vertex atom is different. In addition, proposed on the revised distance matrix with relative bond length, PE shows the interaction between bonding atoms and structural information of molecules. Consequently, PE has better structural selectivity than other topological indexes.

Alkanes are non-polar compounds having only dispersion force in molecules. The dispersion force changes positively with an increase in the molecular weight or atom number (N). However, the rate of increasing decreases gradually. Similarly, PE increases with N but the changing rate gets lower. For instance, the value of PE for ethane, propane, butane, and pentane was 0.5169, 0.7013, 0.8728, and 1.0249, respectively, and the increment between them is 0.1844, 0.1715, and 0.1521, respectively, which shows the same regularity as that of the dispersion force. Therefore, index *PE* conveys the inherent chemical information of compounds which leads to successful modeling.

Conclusion

To summarize, the new equilibrium-electronegative index *PE* described the relationship between structure–property/retention and the properties of organic compounds with a wide range of data subsets. The good structural selectivity of *PE* was attributed to the fact that the index uses equilibrium electronegativities of each atom in a molecule, which is calculated on the basis of step-wise addition as well as replacing the topological distance with relative bond length. Using P_3 and $(PE)^{1/2}$, the quality of the modeling was shown. Moreover, this method once again demonstrates the wider application of the traditional distance matrix to the organic compounds including heteroatoms and/or multiple bonds as well as inorganic compounds. Studies involving other compounds will be presented in a later paper.

Supporting Information

Tables S1–S10 are the corresponding data. This materials is available free of charge on the web at http://www.csj.jp/journals/bcsj/.

References

1 L. H. Hall, B. Mohney, L. B. Kier, *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 76.

2 F. A. L. Ribeiro, M. M. C. Ferreira, *THEOCHEM* **2003**, *663*, 109.

3 S. Yang, W. Lu, N. Chen, Q. Hu, *THEOCHEM* **2005**, *719*, 119.

4 H.-S. Kim, K.-W. Chi, *THEOCHEM* **2005**, *722*, 1.

5 K. Rose, L. H. Hall, L. B. Kier, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 651.

6 X. Q. Kong, D. Shea, W. A. Gebreyes, X.-R. Xia, *Anal. Chem.* **2005**, *77*, 1275.

7 M. Clark, *J. Chem. Inf. Model.* **2005**, *45*, 30.

8 Y. H. Zhao, M. H. Abraham, *J. Org. Chem.* **2005**, *70*, 2633.

9 M. Randić, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 373.

10 A. G. Fragkaki, M. A. Koupparis, C. G. Georgakopoulos, *Anal. Chim. Acta* **2004**, *512*, 165.

11 C. Y. Zhou, C. M. Nie, S. Li, S. N. Wen, G. W. Peng, Z. H. Li, *Chin. J. Inorg. Chem.* **2007**, *23*, 25.

12 C. M. Nie, S. N. Wen, Z. H. Li, S. B. Xie, H. Q. Wang, *Acta Chim. Sin.* **2002**, *60*, 207.

13 C.-M. Nie, Y.-M. Dai, S.-N. Wen, Z.-H. Li, C.-Y. Zhou, G.-W. Peng, *Acta Chim. Sin.* **2005**, *63*, 1449.

14 Z.-H. Li, Y.-M. Dai, S.-N. Wen, C.-M. Nie, C.-Y. Zhou, *Acta Chim. Sin.* **2005**, *63*, 1348.

15 Y. M. Dai, S. N. Wen, C. M. Nie, Z. H. Li, *Chin. J. Inorg. Chem.* **2005**, *7*, 1015.

16 C. M. Nie, *J. Wuhan Univ., Nat. Sci. Ed.*, **2000**, *2*, 176.

17 D. J. Livingstone, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 195.

18 H. Wiener, *J. Am. Chem. Soc.* **1947**, *69*, 17.

19 L. B. Kier, L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, **1976**.

20 L. B. Kier, L. H. Hall, *Molecular Connectivity in Structure-Activity Studies*, Research Studies Press, Letchworth, U.K., **1986**.

21 H. Hosoya, *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332.

22 A. T. Balaban, *Chem. Phys. Lett.* **1982**, *89*, 399.

23 D. Bonchev, N. Trinajstić, *J. Chem. Phys.* **1977**, *67*, 4517.

24 E. Estrada, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 31.

25 M. Randic, *J. Am. Chem. Soc.* **1975**, *97*, 6609.

26 Y. Yao, L. Xu, X. Yuan, *Acta. Chim. Sin.* **1993**, *51*, 1041.

27 Y. Z. Liang, R. Q. Yu, *Chemometrics*, High Education Press, Beijing, P. R. China, **2003**.

28 M. Mihalic, N. Trinajstic, *J. Chem. Educ.* **1992**, *69*, 701.

29 M. Jaiswal, P. Khadikar, *Bioorg. Med. Chem.* **2004**, *12*, 1793.

30 C. L. Yaws, *Chemical Properties Handbook*, McGraw-Hill, Beijing, P. R. China, **1999**.

31 H. C. Li, *Handbook for Analytical Chemistry: Gas Chromatography Analysis*, Chemistry Industry Press of China, Beijing, P. R. China, **1999**.

32 M. Razinger, J. R. Chretien, J. E. Dubois, *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 23.

33 L. Xu, *Chemometrical Method*, Scientific Press of China, Beijing, P. R. China, **1999**.