Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Molecular Graphics and Modelling 24 (2006) 219–226
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Quantitative study of electrostatic and steric effects on physicochemical property and biological activity

Yi-Yu Cheng *, Hua Yuan

Pharmaceutical Informatics Institute, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310027, China

Received 7 July 2005; received in revised form 18 August 2005; accepted 18 August 2005
Available online 22 September 2005

Abstract

The physicochemical properties and biological activities are highly dependent on the electrostatic and steric effects of compounds. The characterization of these structural features will help to elucidate the quantitative structure–property/activity relationship (QSPR/QSAR). In this paper, two novel structural descriptors, lone-pair electrons index (LEI) and molecular volume index (MVI) were developed to quantify the molecular electrostatic and steric effects, respectively. One of the most attractive traits of these two descriptors is that the calculation is very easy, especially for LEI, which only takes the heteroatoms into account. Four data sets on diverse physicochemical properties and biological activities were selected as examples to evaluate the utility of these descriptors. One or two-variable multiple linear regression (MLR) models against the four data sets obtained such correlation results as follows: correlation coefficient $R = 0.9998$ and standard deviation $s = 0.095$ for the partition coefficient of halide benzenes; $R = 0.9689$ and $s = 0.248$ for the toxicity of heterocyclic nitrogen-containing compounds to miracidium; $R = 0.9645$ and $s = 0.209$ for the antifungal activity of benzyl alcohols to aspergillus niger; $R = 0.9530$ and $s = 0.206$ for the toxicity of substituted phenols against tetrahymena pyriformis. Leave-one-out cross validation was also carried out to evaluate the stability of each model and good results were obtained.

Keywords: Electrostatic effect; Steric effect; Lone-pair electrons index (LEI); Molecular volume index (MVI); QSPR/QSAR models

1. Introduction

Electrostatic and steric effects are regarded as two important factors involved in drug–target interactions, hydrogen bond formation, stereoselection reactions, and so on. Many chemists have paid attention to them and attempted to describe them quantitatively. By now, various quantum-chemical descriptors, topological indices and experimentally based parameters have been developed to characterize the electrostatic and steric structural information. For example, Hammett constants [1] (σ, σ^-, σ^+, σ^*) were the classical descriptors accounting for electrostatic effect of substituent groups, by which many phenomena in organic chemistry have been rationalized. Cao et al. [2–4] proposed polarizability effect index (PEI) and modeled various properties of diverse compounds. Recently, the molar refractivity (MR and CMR for the experimental and calculated values, respectively) and the number of valence electrons (NVE) were also used to express the polarizability effects in ligand–substrate interactions [5]. For the steric effect, the first and widely used parameter is Taft's E_s [6] proposed in 1952, which is based on the relative rates of the acid-catalyzed hydrolysis of esters. Topological indices S_{ij} [7] and TSEI [8] accounting for steric structural information were also developed to correlate with the bond dissociation energies of C–C bonds and the dihedral angles of 1,2-disubstituted benzene compounds. In addition, some well-known professional molecular modeling programs like CoMFA [9] are also based on the electrostatic and steric fields.

This paper investigates the electrostatic effect from a new point of view, that is, it takes only heteroatoms into consideration rather than all atoms of the molecule. It is observed that the compounds with biological activities generally contain heteroatoms besides carbon and hydrogen and the substitution site of heteroatoms has significant influence on the biological activities of compounds, suggesting that the specific structural information of heteroatoms is worthy of consideration. As we know, the valence electrons of carbon have all bonded to form sigma (σ) or pi (π) bonds, while the heteroatoms often have unshared or lone-pair electrons. Since the lone-pair electrons are subject to the attraction of the central

* Corresponding author. Tel.: +86 571 879 52509; fax: +86 571 879 51138.
E-mail addresses: chengyy@zju.edu.cn (Y.-Y. Cheng), yh_cathy@zju.edu.cn (H. Yuan).

1093-3263/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2005.08.005

atom nucleus alone rather than be shared between the bonded atoms, the electronic density of lone-pair electrons is concentrated close to the central atom [10], which may result in the different contribution to the behavior of compounds between carbon atoms and heteroatoms. By now, few structural descriptors have described this specific electrostatic information. Therefore, this paper intends to quantify the lone-pair electrons effect with a simple formulation.

Besides the electrostatic effect on molecular properties, steric effect caused by molecular bulk is also an important factor to be considered. For example, the toxicity (log $1/C$) [11] values of *p*-methyl phenol and *p*-(*t*)-butyl phenol against tetrahymena pyriformis are −0.192 and 0.913, respectively. The large discrepancy of toxicities may result from the different volumes of substitution groups, methyl and (*t*)-butyl. This paper tries to scale the steric effect based on the van der Waals volumes of compositive atoms. Four data sets concerning different physicochemical properties and biological activities are employed to validate the utility of these two descriptors.

2. Methodology

Many topological indices [12–15] have been developed based on the molecular graph theory and have proven useful in quantitative structure–property/activity relationship (QSPR/QSAR) studies. Due to the simplicity and efficiency of graph theoretical approaches, this paper also developed two topological indices, lone-pair electrons index (LEI) and molecular volume index (MVI) to quantify the molecular electrostatic and steric effects, respectively.

2.1. Lone-pair electrons index

The electrostatic interactions caused by lone-pair electrons depend on both the intrinsic atomic properties and the distance between heteroatoms with lone-pair electrons and other atoms. First, we define LE_i of each atom as follows:

$$LE_i = \frac{\sqrt{n}(N - N_b)}{\chi} \quad (1)$$

where n is the principal quantum number, N and N_b are the numbers of valence electrons and bonding electrons, respectively, and χ is the Pauling electronegativity. There are three inducements to define LE_i as Eq. (1): (a) the more nonbonding electrons ($N - N_b$) of an atom may increase the electronic density around it and further strengthen the electrostatic interactions; (b) the smaller electronegativity and larger atomic radius will weaken the attraction of atomic nucleus for lone-pair electrons, which may cause these electrons to be polarized easily; (c) the atomic radiuses of elements in different period are roughly proportional to their principal quantum number (n), so n is used as a simple relative scale to indicate the atomic radius. Based on Eq. (1), LE_i is designed to characterize the polarity and polarizability resulted from lone-pair electrons. The LE_i values of ordinary atoms such as carbon, nitrogen, oxygen and halogens were listed as Table 1.

Table 1
The LE_i values of atoms

Atoms	LE_i
C	0
N	0.9304
O	1.6444
F	2.1320
Cl	3.2887
Br	4.0541
I	5.0438

Known from Eq. (1) and Table 1, halogens have larger LE_i values than N and O due to more lone-pair electrons ($N - N_b$) on halogens. For atoms with the same $N - N_b$, the larger of n and the smaller of χ will increase the value of LE_i. Thus, iodine has the largest LE_i. In addition, $LE_{(C)} = 0$, which indicates that carbon atoms can be ignored when computing lone-pair electrons effect of molecule as a whole. Therefore, the calculation can be simplified greatly.

Because the interactions between two atoms will decrease with the broadening of distance between them, lone-pair electrons index is defined with the following formulation:

$$LEI = \frac{1}{k}\sum_{i=1}^{P}\sum_{j=1}^{K}\frac{LE_i}{D_{ij}^2} \quad (2)$$

where k is the number of vertices in hydrogen-suppressed molecular graph, p is the number of heteroatoms and D_{ij} is the topological distance between vertex i and j. If $D_{ij} = 0$, then order $D_{ij}^{-2} = 0$. Eq. (2) indicates that LEI equals the sum of interactions between each heteroatom and other atoms except hydrogens.

2.2. Molecular volume index

For the Main Group elements from IVA to VIIA of the Periodic Table, the variation in atomic size may be generalized as a decrease on going from left to right across a Period and an increase on going down a Group [10]. The different atomic size will exhibit different steric effect on molecular behavior. van der Waals volume of atom is usually employed to scale the atomic size, so we describe the steric effect with molecular volume index.

$$MVI = \sum_{i=1}^{k}\sum_{j=i+1}^{k}\frac{V_iV_j}{D_{ij}^2} \quad (3)$$

where V_i and V_j are van der Waals volumes [16] of groups i and j, respectively. Group i is composed of vertex i and the adjacent hydrogen atoms. Because the number of hydrogens in molecule is considerable, the volume of hydrogens cannot be ignored though it is small for only one hydrogen. For example, the volume of CH_3 is 0.374×10^{-2} and 0.206×10^{-2} Å^3 when the hydrogens are and are not considered, respectively. The van der Waals volumes of some groups are shown in Table 2.

Take 2-methyl-6-chlorine-4-bromine phenol as an example to illustrate the calculation of LEI and MVI. Fig. 1 shows its

Table 2
The van der Waals volumes (V_i, 10^{-2} Å^3) of some groups

Groups	V_i	Groups	V_i	Groups	V_i	Groups	V_i
H	0.056	S	0.244	I	0.388	NH	0.197
C	0.206	F	0.115	CH	0.262	NH_2	0.253
N	0.141	Cl	0.244	CH_2	0.318	NH_3	0.309
O	0.115	Br	0.287	CH_3	0.374	OH	0.171

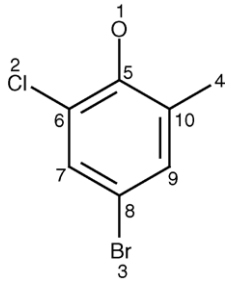

Fig. 1. The hydrogen-suppressed molecular graph of 2-methyl-6-chlorine-4-bromine phenol.

hydrogen-suppressed molecular graph, where the numbers are the random numberings of each vertex.

First, we use matrices LE, V and D to represent LE_i, V_i and D_{ij} of the molecule, respectively.

$$\mathrm{LE} = [\mathrm{LE}_1 \quad \mathrm{LE}_2 \quad \mathrm{LE}_3] = [\mathrm{LE}_\mathrm{O} \quad \mathrm{LE}_\mathrm{Cl} \quad \mathrm{LE}_\mathrm{Br}] = [1.6444 \quad 3.2887 \quad 4.0541]$$

$$V = [V_1 \quad V_2 \quad \cdots \quad V_{10}] = [0.171 \quad 0.244 \quad 0.287 \quad 0.374 \quad 0.206 \quad 0.206 \quad 0.262 \quad 0.206 \quad 0.262 \quad 0.206]$$

$$D = [D_{ij}] = \begin{bmatrix} 0 & 3 & 5 & 3 & 1 & 2 & 3 & 4 & 3 & 2 \\ 3 & 0 & 4 & 4 & 2 & 1 & 2 & 3 & 4 & 3 \\ 5 & 4 & 0 & 4 & 4 & 3 & 2 & 1 & 2 & 3 \\ 3 & 4 & 4 & 0 & 2 & 3 & 4 & 3 & 2 & 1 \\ 1 & 2 & 4 & 2 & 0 & 1 & 2 & 3 & 2 & 1 \\ 2 & 1 & 3 & 3 & 1 & 0 & 1 & 2 & 3 & 2 \\ 3 & 2 & 2 & 4 & 2 & 1 & 0 & 1 & 2 & 3 \\ 4 & 3 & 1 & 3 & 3 & 2 & 1 & 0 & 1 & 2 \\ 3 & 4 & 2 & 2 & 2 & 3 & 2 & 1 & 0 & 1 \\ 2 & 3 & 3 & 1 & 1 & 2 & 3 & 2 & 1 & 0 \end{bmatrix}$$

According to Eqs. (2) and (3), LEI and MVI are computed as follows:

$$\begin{aligned} \mathrm{LEI} &= \frac{1}{10}\times\left(\sum_{j=1}^{10}\frac{\mathrm{LE}_1}{D_{1j}^2}+\sum_{j=1}^{10}\frac{\mathrm{LE}_2}{D_{2j}^2}+\sum_{j=1}^{10}\frac{\mathrm{LE}_3}{D_{3j}^2}\right) \\ &= \frac{1}{10}\times\left[1.6444\times\left(0+\frac{1}{3^2}+\frac{1}{5^2}+\frac{1}{3^2}+\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\frac{1}{3^2}+\frac{1}{2^2}\right)\right. \\ &\quad + 3.2887\times\left(\frac{1}{3^2}+0+\frac{1}{4^2}+\frac{1}{4^2}+\frac{1}{2^2}+\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\frac{1}{3^2}\right)+ \\ &\quad \left. 4.0541\times\left(\frac{1}{5^2}+\frac{1}{4^2}+0+\frac{1}{4^2}+\frac{1}{4^2}+\frac{1}{3^2}+\frac{1}{2^2}+\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}\right)\right] = \frac{1}{10}\times(3.3660+6.6459+7.9044) = 1.7916 \end{aligned}$$

$$\begin{aligned} \mathrm{MVI} &= \sum_{i=1}^{10}\sum_{j=i+1}^{10}\frac{V_iV_j}{D_{ij}^2} \\ &= 0.0787+0.1040+0.1203+0.1440+0.1166+0.0812+0.0771+0.0646+0.0540+0 \\ &= 0.8405 \end{aligned}$$

2.3. Statistical analysis method

Multiple linear regression (MLR) is used through the paper to build QSPR/QSAR models, the intercepts and coefficients of which are reported with their 95% confidence intervals. To test the stability of models, leave-one-out (LOO) cross validation is carried out against each model. The statistical quality is judged by the correlation coefficient (R) and standard deviation (s) for MLR model, R_{cv} and s_{cv} for LOO cross validation.

3. Results and discussion

In order to evaluate the rationality and utility of these two descriptors, LEI and MVI, four data sets served as examples. They are partition coefficient (log P) of 36 halide benzenes [17], toxicity (log 1/C) of 24 heterocyclic nitrogen-containing compounds to miracidium in liquid [18], antifungal activity (log 1/C) of 19 benzyl alcohols to aspergillus niger [19] and toxicity (log 1/C) of 50 substituted phenols against tetrahymena pyriformis [11]. It should be mentioned that the log P values of halide benzenes taken from Ref. [17] are contrary in sign to those from Ref. [20]. For example, log P of PhBr is −2.99 in Ref. [17] and 2.99 in Ref. [20], as can be seen for other compounds. Known from the theory of organic chemistry, log P in the latter seems more reasonable. Because more experimental log P data are available in Ref. [17] than in Ref. [20], this paper cites the log P values from Ref. [17] in the negative form of the original data.

3.1. Partition coefficient (log P) of 36 halide benzenes

Octanol/water partition coefficient (log P) of organic compounds or medicaments is a very important physicochemical parameter relevant to biological properties such as drug action, cellular uptake, metabolism, bioavailability and toxicity. The determination of log P by experiments is not fit for the high-throughput screening of drug-like molecules. Hence, the fast and reliable structure-based modeling methods for the estimation of log P are preferred.

This paper investigated the octanol/water partition coefficient of 36 halide benzenes with lone-pair electrons index and obtained Eq. (4).

$$\log P = (1.4347 \pm 0.0550) + (1.3680 \pm 0.0250)\mathrm{LEI},$$
$$R = 0.9944, s = 0.115, F = 3002.49, R_{cv} = 0.9935,$$
$$s_{cv} = 0.124, n = 36 \quad (4)$$

The high correlation coefficient and low standard deviation of one-variable Eq. (4) indicate that the electrostatic effect resulted from heteroatoms is responsible for the solubility of halide benzenes. A previous paper [17] modeled the log P of the same data set with a modified connectivity index ${}^{1}\chi^{Y}$ and obtained $R = 0.9927$, $s = 0.1313$. The result of this paper is better than that of the literature.

If the molecular volume index was also included, then a two-variable equation was obtained as follows:

$$\log P = (-0.1171 \pm 0.4000) + (0.9032 \pm 0.1208)\mathrm{LEI} + (3.4277 \pm 0.8775)\mathrm{MVI},$$
$$R = 0.9962, s = 0.097, F = 2138.46, R_{cv} = 0.9953,$$
$$s_{cv} = 0.107, n = 36 \quad (5)$$

Seen from the regression coefficients in Eq. (5), the standard deviation (0.4000) of intercept (−0.1171) is very large, which indicates the intercept is not significant in this equation.

Table 3
The molecular descriptors and log P of 36 halide benzenes

No.	Compounds	LEI	MVI	log P	
				Experimental[a]	Calculated[b]
1	$PhCl_6$	3.5714	0.9398	6.42	6.33
2	$PhCl_5$	3.1310	0.8679	5.69	5.69
3	1,2,4,5-$PhCl_4$	2.6288	0.7970	5.05	4.99
4	1,2,3,5-$PhCl_4$	2.6436	0.7975	5.05	5.01
5	1,2,4-$PhCl_3$	2.1126	0.7295	4.27	4.29
6	1,2,3-$PhCl_3$	2.1646	0.7314	4.27	4.35
7	1,3,5-$PhCl_3$	2.0935	0.7285	4.27	4.27
8	1,2-$PhCl_2$	1.5587	0.6649	3.55	3.57
9	1,3-$PhCl_2$	1.5187	0.6634	3.38	3.53
10	1,4-$PhCl_2$	1.5002	0.6630	3.59	3.51
11	PhCl	0.8385	0.5998	2.81	2.69
12	1,2,4,5-$PhBr_4$	3.2406	0.8735	6.01	5.81
13	1,2,3-$PhBr_3$	2.6683	0.7880	4.98	5.00
14	1,2,4-$PhBr_3$	2.6043	0.7854	4.98	4.93
15	1,3,5-$PhBr_3$	2.5807	0.7842	4.98	4.91
16	1,2-$PhBr_2$	1.9215	0.7016	4.07	4.03
17	1,3-$PhBr_2$	1.8722	0.6997	3.75	3.97
18	1,4-$PhBr_2$	1.8494	0.6990	4.07	3.95
19	PhBr	1.0336	0.6175	2.99	2.93
20	1,2-PhF_2	1.0105	0.5571	2.59	2.71
21	1,3-PhF_2	0.9846	0.5559	2.58	2.69
22	1,4-PhF_2	0.9726	0.5558	2.58	2.67
23	PhF	0.5436	0.5467	2.27	2.24
24	1,2,3-PhI_3	3.3197	0.9252	5.86	6.05
25	1,2,4-PhI_3	3.2400	0.9199	5.85	5.95
26	1,3,5-PhI_3	3.2108	0.9180	5.85	5.92
27	1,2-PhI_2	2.3906	0.7895	4.65	4.74
28	1,3-PhI_2	2.3293	0.7859	4.64	4.68
29	1,4-PhI_2	2.3009	0.7841	4.64	4.64
30	PhI	1.2860	0.6591	3.36	3.30
31	1-Cl-2-Br-Ph	1.7401	0.6831	3.83	3.80
32	1-Cl-3-Br-Ph	1.6955	0.6815	3.83	3.75
33	1-Cl-4-Br-Ph	1.6748	0.6809	3.83	3.73
34	1-Cl-4-I-Ph	1.9006	0.7231	4.12	4.08
35	1-Br-2-I-Ph	2.1560	0.7450	4.36	4.38
36	1-Br-4-I-Ph	2.0751	0.7414	4.36	4.30

[a] Taken from Ref. [17].
[b] Calculated by Eq. (6).

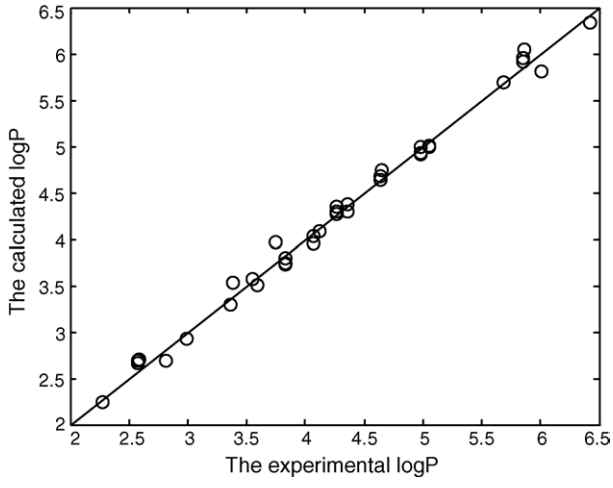

Fig. 2. The plot of calculated vs. experimental log P of halide benzenes.

Therefore, another regression analysis with zero-intercept was carried out and formulated as Eq. (6):

$$\log P = (0.9371 \pm 0.0336)\mathrm{LEI} + (3.1724 \pm 0.1000)\mathrm{MVI},$$
$$R = 0.9998,\ s = 0.095,\ F = 38189.75,\ R_{cv} = 0.9997,$$
$$s_{cv} = 0.102,\ n = 36 \quad (6)$$

The high correlation coefficients ($R = 0.9998$, $R_{cv} = 0.9997$) show good statistical quality of Eq. (6). Compared to Eq. (4), some improvement was achieved by the consideration of MVI. The calculated log P with Eq. (6) are listed in Table 3.

In order to view the relationship between the calculated log P ($\log P_{calc}$) and the experimental ones ($\log P_{exp}$) more intuitively, the plot of $\log P_{calc}$ versus $\log P_{exp}$ is shown in Fig. 2. It is obvious that the calculated and the experimental values are in good agreement.

3.2. Toxicity of 24 heterocyclic nitrogen-containing compounds to miracidium

The toxicity values (log(1/C), where C represents the depressed concentration) of 24 heterocyclic nitrogen-containing compounds to miracidium were obtained from Ref. [18]. The quantitative correlation of log(1/C) with MVI generates the following equation:

$$\log\left(\frac{1}{C}\right) = (-2.4910 \pm 0.1268) + (2.7935 \pm 0.1521)\mathrm{MVI},$$
$$R = 0.9689,\ s = 0.248,\ F = 337.36,\ R_{cv} = 0.9617,$$
$$s_{cv} = 0.275, n = 24 \quad (7)$$

The integration of LEI into Eq. (7) does not improve the correlation quality. The reference study on this data set was carried out by Jiang et al. [17], in which the obtained statistic quantities were $R = 0.9483$ and $s = 0.319$ with one descriptor. Eq. (7) shows that large molecular size decreases the depressed concentration C, that is, a bulky molecule has stronger toxicity.

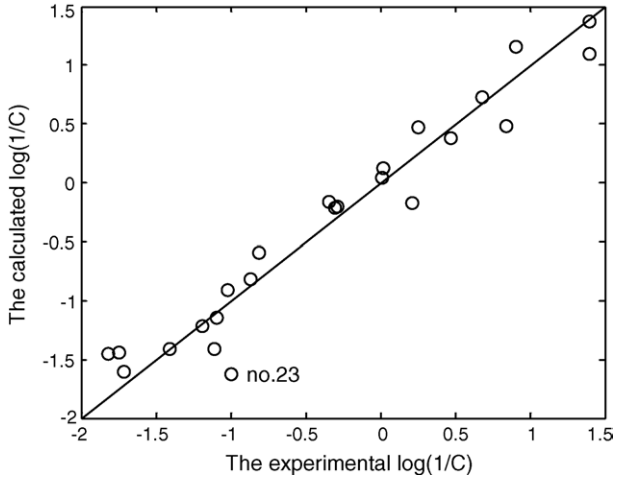

Fig. 3. The plot of calculated vs. experimental toxicity (log(1/C)) of heterocyclic nitrogen-containing compounds.

Table 4
The molecular descriptor and toxicity (log(1/C)) of 24 heterocyclic nitrogen-containing compounds

No.	Compounds	MVI	log(1/C)	
			Experimental[a]	Calculated[b]
1	Phthalizine	0.8338	−0.34	−0.16
2	3-Methyl pyridine	0.5672	−1.02	−0.91
3	2,6-Dimethyl pyridine	0.6820	−0.81	−0.59
4	Quinazine	0.8207	−0.29	−0.20
5	2-Methyl pyrazine	0.4815	−1.09	−1.15
6	2,3-Dimethyl pyrazine	0.5975	−0.87	−0.82
7	1,2-Dimethyl indole	1.0619	0.84	0.48
8	2-Methyl quinoline	1.0273	0.47	0.38
9	2,6-Dimethyl quinoline	1.1535	0.68	0.73
10	Quinoxaline	0.8164	−0.30	−0.21
11	2-Methylquinoxaline	0.9343	0.02	0.12
12	2,3-Dimethylquinoxaline	1.0604	0.25	0.47
13	Acridine	1.3826	1.40	1.37
14	Phenazine	1.2845	1.40	1.10
15	Pyrimidine	0.3758	−1.75	−1.44
16	Pyridine	0.4549	−1.19	−1.22
17	Pyrazine	0.3738	−1.82	−1.45
18	Pyrrole	0.3864	−1.11	−1.41
19	Indole	0.8307	0.21	−0.17
20	Quinoline	0.9061	0.01	0.04
21	Carbazole	1.3086	0.91	1.16
22	Pyrazole	0.3151	−1.71	−1.61
23	Imidazole	0.3092	−1.00	−1.63
24	Pyridazine	0.3868	−1.41	−1.41

[a] Taken from Ref. [18].
[b] Calculated by Eq. (7).

The values of calculated toxicity were listed in Table 4 and shown in Fig. 3.

Seen from Fig. 3, the largest deviation of the calculated from the experimental value is observed for imidazole (no. **23**, labeled in Fig. 3) with the residual of 0.63. While for its isomeric compound pyrazole (no. **22**), the calculated log(1/C) is very close to the experimental one. It is wondering that the experimental log(1/C) of imidazole (−1.00) is quite different from that of pyrazole (−1.71) though they are isomeric compounds.

3.3. Antifungal activity of 19 benzyl alcohols for aspergillus niger

A series of 19 benzyl alcohols with antifungal activity (log(1/C)) measured by the inhibitory power against aspergillus niger were selected as another data set to test the QSAR ability of MVI and LEI. The linear regression model was expressed as Eq. (8):

$$\log\left(\frac{1}{C}\right) = (-1.8368 \pm 0.4125) + (3.9127 \pm 0.5093)\mathrm{MVI} + (0.8791 \pm 0.0935)\mathrm{LEI},$$
$$R = 0.9645,\ s = 0.209,\ F = 106.58,\ R_{cv} = 0.9519,$$
$$s_{cv} = 0.243,\ n = 19 \quad (8)$$

Table 5
The molecular descriptors and antifungal activity (log(1/*C*)) of 19 benzyl alcohols

No.	Compounds	MVI	LEI	log(1/*C*)	
				Experimental[a]	Calculated[b]
1	Benzyl alcohol	0.7108	0.3365	1.51	1.24
2	4-Chlorobenzyl alcohol	0.7752	0.9811	2.07	2.06
3	2,4-Dichlorobenzyl alcohol	0.8442	1.5784	3.07	2.85
4	3,4-Dichlorobenzyl alcohol	0.8434	1.5833	3.07	2.86
5	2,4,5-Trichlorobenzyl alcohol	0.9135	2.0955	3.32	3.58
6	3,4,5-Trichlorobenzyl alcohol	0.9132	2.1134	3.63	3.59
7	2-Bromobenzyl alcohol	0.7975	1.1927	2.15	2.33
8	4-Bromobenzyl alcohol	0.7935	1.1387	2.27	2.27
9	4-Iodobenzyl alcohol	0.8365	1.3424	2.75	2.62
10	4-Methylbenzyl alcohol	0.8305	0.3042	1.79	1.68
11	2,4-Dimethylbenzyl alcohol	0.9606	0.2841	2.14	2.17
12	3,5-Dimethyl-4-chlorobenzyl alcohol	1.0323	0.8811	3.05	2.98
13	3,5-Dimethyl-4-iodobenzyl alcohol	1.1015	1.2122	3.42	3.54
14	2-Nitrobenzyl alcohol	0.8083	0.8609	2.49	2.08
15	4-Nitrobenzyl alcohol	0.8046	0.8337	2.00	2.04
16	4-nitrilebenzyl alcohol	0.8111	0.4340	1.67	1.72
17	2-Hydroxylbenzyl alcohol	0.7460	0.6684	1.39	1.67
18	3-Hydroxylbenzyl alcohol	0.7446	0.6513	1.39	1.65
19	4-Hydroxylbenzyl alcohol	0.7441	0.6427	1.39	1.64

[a] Taken from Ref. [19].
[b] Calculated by Eq. (8).

Estrada [19] also investigated the same data set with three descriptors and obtained a model with $R = 0.967$ and $s = 0.201$.

The molecular descriptors MVI and LEI together with the values of log(1/*C*) were listed in Table 5.

The relationship between the calculated and the experimental log(1/*C*) was represented intuitively in Fig. 4.

3.4. Toxicity of 50 substituted phenols against tetrahymena pyriformis

The QSAR study was extended to the toxicity (log(1/*C*)) of 50 substituted phenols against tetrahymena pyriformis [11] with two descriptors MVI and LEI. In the literature, compound no. **38** (2,4,6-trimethylphenol) may have been spelled incorrectly and we corrected it as 2,4,6-trichlorophenol judging by its experimental log(1/*C*) and the calculated structural descriptors. The correlation result was formulated by Eq. (9):

$$\log\left(\frac{1}{C}\right) = (-1.6010 \pm 0.1206) + (1.8828 \pm 0.1053)\mathrm{MVI} + (0.8249 \pm 0.0455)\mathrm{LEI},$$
$$R = 0.9530,\ s = 0.206,\ F = 232.29,\ R_{cv} = 0.9449,\ s_{cv} = 0.223,\ n = 50 \quad (9)$$

In Roy and Ghosh's work [11], a complicated discussion on the extended topochemical atom (ETA) indices was performed and a five-parameter correlation model was built up ($R = 0.977$ and

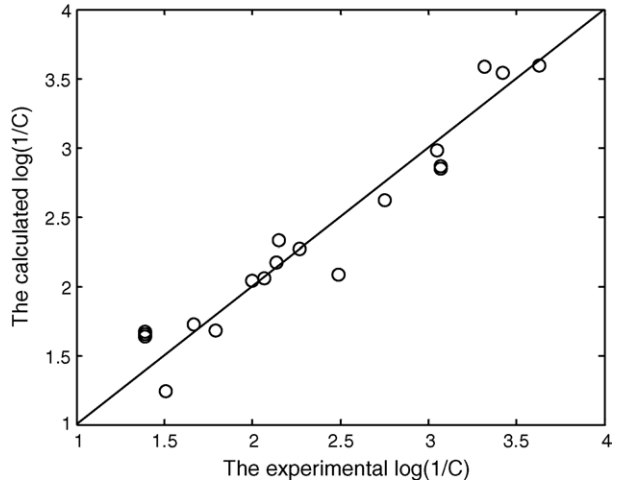

Fig. 4. The plot of calculated vs. experimental antifungal activity (log(1/*C*)) of benzyl alcohols.

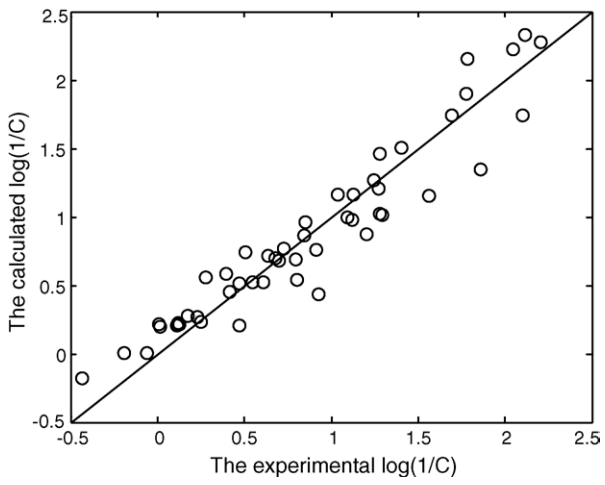

Fig. 5. The plot of calculated vs. experimental toxicity (log(1/*C*)) of tetrahymena pyriformis.

Table 6
The molecular descriptors and toxicity (log(1/*C*)) of 50 substituted phenols

No.[a]	Compounds	MVI	LEI	log(1/*C*)	
				Experimental[b]	Calculated[c]
1	Phenol	0.5698	0.4193	−0.431	−0.182
2	2,6-Difluorophenol	0.5906	1.2945	0.396	0.579
3	2-Fluorophenol	0.5801	0.8949	0.248	0.229
4	4-Fluorophenol	0.5789	0.8614	0.017	0.200
5	3-Fluorophenol	0.5790	0.8720	0.473	0.208
6	4-Methylphenol	0.6864	0.3751	−0.192	0.001
7	3-Methylphenol	0.6868	0.3797	−0.062	0.005
8	2-Chlorophenol	0.6339	1.169	0.277	0.557
9	2-Bromophenol	0.6518	1.3504	0.504	0.740
10	4-Chlorophenol	0.6325	1.1252	0.545	0.518
11	3-Ethylphenol	0.8409	0.3448	0.229	0.267
12	2-Ethylphenol	0.8425	0.3578	0.176	0.280
13	4-Bromophenol	0.6503	1.2998	0.681	0.696
14	2,3-Dimethylphenol	0.8132	0.3578	0.122	0.225
15	2,4-Dimethylphenol	0.8095	0.3537	0.128	0.215
16	2,5-Dimethylphenol	0.8084	0.3578	0.009	0.216
17	3,4-Dimethylphenol	0.8116	0.3448	0.122	0.212
18	3,5-Dimethylphenol	0.8087	0.3489	0.113	0.209
19	3-Chloro-4-fluorophenol	0.6431	1.519	0.842	0.863
20	2-Chloro-5-methylphenol	0.7547	1.0823	0.640	0.713
21	4-Iodophenol	0.6922	1.5255	0.854	0.961
22	3-Iodophenol	0.6926	1.5443	1.118	0.977
23	2-Isopropylphenol	0.9901	0.3323	0.803	0.537
24	3-Isopropylphenol	0.9881	0.3169	0.609	0.521
25	4-Isopropylphenol	0.9872	0.3092	0.473	0.513
26	2,5-Dichlorophenol	0.6978	1.7548	1.128	1.160
27	2,3-Dichlorophenol	0.6997	1.8068	1.271	1.207
28	4-Chloro-2-methylphenol	0.7533	1.0433	0.700	0.678
29	4-Chloro-3-methylphenol	0.7541	1.0522	0.795	0.687
30	2,4-Dichlorophenol	0.698	1.7589	1.036	1.164
31	3-*tert*-Butylphenol	1.1283	0.2941	0.730	0.766
32	4-*tert*-Butylphenol	1.1272	0.2852	0.913	0.757
33	3,5-Dichlorophenol	0.6971	1.7446	1.562	1.151
34	2-Phenylphenol	1.2604	0.2693	1.094	0.994
35	2,4-Dibromophenol	0.7347	2.086	1.403	1.503
36	2,3,6-Trimethylphenol	0.9397	0.3403	0.418	0.449
37	3,4,5-Trimethylphenol	0.9417	0.3206	0.930	0.437
38	2,4,6-Trichlorophenol	0.7650	2.3070	1.695	1.742
39	4-Chloro-3,5-dimethylphenol	0.8806	0.9938	1.203	0.877
40	4-Bromo-2,6-dichlorophenol	0.7837	2.4562	1.779	1.901
41	2,4,5-Trichlorophenol	0.7649	2.3002	2.100	1.737
42	4-Bromo-6-chloro-2-methylphenol	0.8405	1.7916	1.277	1.459
43	4-Bromo-2,6-dimethylphenol	0.8984	1.1270	1.278	1.020
44	2,4,6-Tribromophenol	0.8214	2.7656	2.050	2.227
45	2-*tert*-Butyl-4-methylphenol	1.2622	0.2910	1.297	1.016
46	4-Chloro-2-isopropyl-5-methylphenol	1.1888	0.8606	1.862	1.347
47	6-*tert*-Butyl-2,4-dimethylphenol	1.4001	0.2826	1.245	1.268
48	2,6-Diphenylphenol	1.9933	0.214	2.113	2.329
49	2,4-Dibromo-6-phenylphenol	1.4447	1.4074	2.207	2.280
50	2,6-Di-*tert*-butyl-4-methylphenol	1.8844	0.2489	1.788	2.152

[a] Compound name of no. **38** was corrected from the original Ref. [11].
[b] Taken from Ref. [11].
[c] Calculated by Eq. (9).

$s = 0.149$). Though the standard deviation (s) of Eq. (9) is a little larger than that of the literature, the number of descriptors employed in Eq. (9) is much fewer and the calculation is much easier. Table 6 shows the molecular descriptors and toxicity of substituted phenols.

The calculated toxicity ($\log(1/C)_{calc}$) by Eq. (9) was plotted against the experimental one ($\log(1/C)_{exp}$) as shown in Fig. 5.

4. Conclusion

Two novel structural descriptors, LEI and MVI, were developed based on graph theory. Lone-pair electrons index reflects the polarity and polarizability mainly brought about by heteroatoms. The consideration of heteroatoms alone rather than all the atoms of molecule not only simplifies the

calculation but also describes the crucial structural information. This approach provides new insight into the construction of structural descriptors. The LEI values of hydrocarbons are zero due to no heteroatoms there, so LEI is not applicable to the QSPR/QSAR modeling of hydrocarbons. Molecular volume index characterizes the steric interactions among the groups of molecule in a simple way. With these two structural descriptors, good QSPR/QSAR models were built for various physico-chemical properties and biological activities of several classes of compounds. As the electrostatic and steric effects are two major substituent effects affecting molecular behavior, these two descriptors are promising to be used to model more extensive properties and activities.

Acknowledgments

This work was financially supported by a key grant from the National Natural Science Foundation of China (No. 90209005). The authors also thank Dr. Le-Ming Shi (National Center for Toxicological Research, the US Food and Drug Administration (FDA), Jefferson, AR 72079, USA) for his constructive suggestions and helpful discussions.

References

[1] L.P. Hammett, The effect of structure upon the reactions of organic compounds. Benzene derivatives, J. Am. Chem. Soc. 59 (1937) 96–103.
[2] C.Z. Cao, Z.L. Li, Molecular polarizability. 1. Relationship to water solubility of alkanes and alcohols, J. Chem. Inf. Comput. Sci. 38 (1998) 1–7.
[3] C.Z. Cao, S.S. Liu, Z.L. Li, On molecular polarizability: 2. Relationship to the boiling point of alkanes and alcohols, J. Chem. Inf. Comput. Sci. 39 (1999) 1105–1111.
[4] C.Z. Cao, H. Yuan, Sh.L. Liu, R.J. Zeng, On molecular polarizability: 3. Relationship to the ionization potential of haloalkanes, amines, alcohols, and ethers, J. Chem. Inf. Comput. Sci. 40 (2000) 1010–1014.
[5] C. Hansch, W.E. Steinmetz, A.J. Leo, S.B. Mekapati, A. Kurup, D. Hoekman, On the role of polarizability in chemical–biological interactions, J. Chem. Inf. Comput. Sci. 43 (2003) 120–125.
[6] R.W. Taft Jr., Polar and steric substituent constants for aliphatic and *o*-benzoate groups from rates of esterification and hydrolysis of esters, J. Am. Chem. Soc. 74 (1952) 3120–3128.
[7] C.Z. Cao, H. Yuan, A new approach of evaluating bond dissociation energy from eigenvalue of bonding orbital-connection matrix for C–C and C–H bonds in alkane, J. Chem. Inf. Comput. Sci. 43 (2003) 600–608.
[8] C.Z. Cao, L. Liu, Topological steric effect index and its application, J. Chem. Inf. Comput. Sci. 44 (2004) 678–687.
[9] R.D. Cramer III, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, J. Am. Chem. Soc. 110 (1988) 5959–5967.
[10] K.M. Mackay, R.A. Mackay, Introduction to Modern Inorganic Chemistry, fourth ed., Prentice Hall, New Jersey, 1989, p. 128.
[11] K. Roy, G. Ghosh, Introduction of extended topochemical atom (ETA) indices in the valence electron mobile (VEM) environment as tools for QSAR/QSPR studies, Internet Electron. J. Mol. Des. 2 (2003) 599–620. http://www.biochempress.com.
[12] M. Randić, Wiener-Hosoya index—a novel graph theoretical molecular descriptor, J. Chem. Inf. Comput. Sci. 44 (2004) 373–377.
[13] E. Estrada, Generalized spectral moments of the iterated line graphs sequence. A novel approach to QSPR studies, J. Chem. Inf. Comput. Sci. 39 (1999) 90–95.
[14] S.S. Liu, C.Z. Cao, Z.L. Li, Approach to estimation and prediction for normal boiling point (NBP) of alkanes based on a novel molecular distance-edge (MDE) vector, λ, J. Chem. Inf. Comput. Sci. 38 (1998) 387–394.
[15] H. Yuan, C.Z. Cao, Topological indices based on vertex, edge, ring and distance: application to various physico-chemical properties of diverse hydrocarbons, J. Chem. Inf. Comput. Sci. 43 (2003) 501–512.
[16] C.Z. Cao, The Substituent Effect in Organic Chemistry (Youji Huaxue zhong de Qudaiji Xiaoying), Kexue Chubanshe, Beijing, 2003, p. 118 (in Chinese, http://www.sciencep.com/).
[17] Y.R. Jiang, J.Y. Liu, Y.H. Hu, T. Fujita, Novel topological index for research on structure–property relationships of complex organic compounds, J. Comput. Chem. 24 (2003) 842–849.
[18] L.B. Kier, L.H. Hall, Molecular Connectivity in Chemistry and Drug Research, Academic, New York, 1976, p. 64.
[19] E. Estrada, Spectral moments of the edge-adjacency matrix of molecular graphs. 2. Molecules containing heteroatoms and QSAR applications, J. Chem. Inf. Comput. Sci. 37 (1997) 320–328.
[20] D.R. Lide (Ed.), CRC Handbook of Chemistry and Physics, Internet Version 2005, 85th ed., CRC Press, Boca Raton, FL, 2005, http://www.hbcpnetbase.com.