Structural Vector Description and Estimation of Normal Boiling Points for 66 Aromatic Hydrocarbons

Shu Shen LIU[1]*, Chun Sheng YIN[2], Shao Xi CAI[1], Zhi Liang LI[1]

[1]College of Biomedical Engineering, Chongqing University, Chongqing 400044
[2]Department of Applied Chemistry, University of Science and Technology of China, Hefei 230026
*Department of Applied Chemistry, Guilin Institute of Technology, Guilin 541004

Abstract: A molecular vector-type descriptor containing 6 variables is used to describe the structure of aromatic hydrocarbons (AHs) and relate to normal boiling points (*bp*) of AHs. The correlation coefficient (*R*) between the estimated *bp* and experimental *bp* is 0.9988 and the root mean square error (*RMS*) is 7.907°C for 66 AHs. The *RMS* obtained by cross-validation is 9.131°C, which implies the relationship model having good prediction ability.

**Keywords:** Molecular vector-type descriptor, aromatic hydrocarbon, normal boiling point.

It is necessary to describe the chemical structure of the examined compound as one or more parameters such as various topological indices in quantitative structure-property relationship (QSPR) studies. Many standard topological indices, including Wiener index[1~2], Hosoya index[3], Randic indices[4~6], Balaban index[7~8], and so on, have been developed. Recently, a vector-type descriptor, molecular distance-edge (MDE) vector containing 10 variables have been used to describe the structure of alkanes and relate to many physical properties such as boiling point[9] and thermodynamic functions of alkanes[10] with good results. In order to avoid computing a geometric distance and to describe the structure of a molecule containing one heteroatom such as oxygen atom, a modified MDE vector was developed and related the boiling points of 106 alcohols with satisfactory results[11]. In this paper, the vector-type descriptor is extended to express the structure of 66 aromatic hydrocarbons (AHs) and relate to normal boiling points (*bp*) with the correlation coefficient (*R*) between the estimated *bp* and experimental *bp* being 0.9988 and the root mean square error (*RMS*) being 7.907°C.

In the conjugated ring system for 66 AH compounds, C-C single bond tends to be shorter and C=C double bond tends to be longer, which differ from pure C-C bond and pure C=C bond. For the convenience of calculation, the bond-length (*nm*) C-C (0.154) and C=C (0.134) in this conjugated ring have an average length of 0.144. In 66 AHs, there is no heteroatom and the relative electronegatives of all non-hydrogen atoms are 1. So, the MDE vector[11] can be rewritten as follows:

$$m_{kl} = \sum_{i \in k, j \in l} \frac{1}{d_{ij}^2} \qquad (k, l = 1, 2, 3, 4; l \geq k) \qquad (1)$$

where k or l is an atomic type of an atom, and i or j is a coding number or series number of an atom in the molecular skeleton graph and i belongs to the kth atomic type and j to the lth atomic type. An atom belonging to the kth atomic type atom is such a non-hydrogen atom which binds to other k non-hydrogen atom/atoms. The d_{ij} expresses the relative distance between the ith atom and jth atom. From equation (1), there are 10 variables, *i.e.*, m_{11}, m_{12}, m_{13}, m_{14}, m_{22}, m_{23}, m_{24}, m_{33}, m_{34} and m_{44}, in the MDE vector for the organic or biological molecules. Because of the absence of the fourth atomic type atom in the examined 66 AHs, there are in fact only 6 variables in the MDE vector, *i.e.*, m_{11}, m_{12}, m_{13}, m_{22}, m_{23}, and m_{33}.

Figure 1. The skeleton structure of acenaphthene with relative bond-length (RBL)
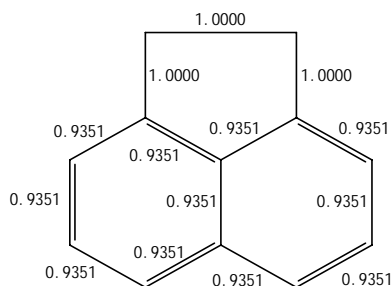
To illustrate a procedure of calculating the MDE vector, the MDE vector of acenaphthene (see skeleton structure in **Figure 1**) is obtained by the following calculation. The relative distance d_{ij} used in calculation is a sum of the relative bond-lengths of various chemical bonds passed from atom i to j and the relative bond-lengths (RBL) of a bond is defined as its bond-length dividing into the length of C-C bond. For example, the RBL of the conjugated C-C or C=C bond is 0.144/0.154=0.9351. So,

$$m_{22} = 2 \cdot \frac{1}{1.9351^2} + 2 \cdot \frac{1}{(1+2\times0.9351)^2} + 4 \cdot \frac{1}{(1+3\times0.9351)^2} + 2 \cdot \frac{1}{(2+2\times0.9351)^2}$$

$$+ \frac{1}{1^2} + 4 \cdot \frac{1}{0.9351^2} + 3 \cdot \frac{1}{(2\times0.9351)^2} + 4 \cdot \frac{1}{(4\times0.9351)^2} + 2 \cdot \frac{1}{(2+0.9351)^2}$$

$$+ 2 \cdot \frac{1}{(3\times0.9351)^2} + 2 \cdot \frac{1}{(5\times0.9351)^2} = 8.4831$$

In the same way, the other 9 descriptors can be obtained as follows:

$m_{23} = 11.5529$, $m_{33} = 4.2889$, $m_{11} = m_{12} = m_{13} = m_{14} = m_{24} = m_{34} = m_{44} = 0$

Molecular structure determines the properties of the molecule. The MDE vector should be related to properties such as boiling point. Applying multiple linear regression (MLR), a quantitative structure-property relationship (QSPR) model between the MDE vector and the *bp*s of 66 AHs is developed as follows:

$$bp = 12.9598 + 169.5369m_{11} + 7.9964m_{12} + 2.2342m_{13} + 10.8194m_{22} + 12.0227m_{23}$$

$$+ 5.4712m_{33} \quad (n = 66, R = 0.9988, RMS = 7.907^{o}C, F = 4049.56) \quad (2)$$

where n is the number of samples and R, RMS and F are the correlation coefficient, the root mean square error and F statistic, respectively. The experimental boiling points (see the column "bp_{EXP}" in **Table 1**) used in developing the model are taken from the reference 12. The estimated boiling points of 66 AHs by substituting the variable's values in the MDE vector, m_{ij} (i,j=1,2,3;$j \geq i$), into Eq.2 are listed in **Table 1** (see column "bp_{M1}"). From **Table 1**, there is a large difference value being Δbp=29.79°C for benzene, which is because only element m_{22} is non-zero and the other 5 variables are zero in the MDE vector of benzene. The relationship profile is easily seen from **Figure 2** with bp_{M1} vs bp_{EXP}. These results show that Eq.2 have a good estimation ability.

Table 1 The MEDV and boiling points estimated and observed for 66 PAHs

No	Compound	bp_{EXP}	bp_{M1}	No	Compound	bp_{EXP}	bp_{M1}
1	benzene	80.10	109.89	34	phenanthrene	338.00	321.72
2	toluene	110.60	122.50	35	anthracene	340.00	326.91
3	ethylbenzene	136.20	146.51	36	4H-cyclopenta[def]phenanthrene	359.00	349.17
4	o-xylene	144.40	142.83	37	fluoranthene	383.00	371.87
5	m-xylene	139.10	142.81	38	pyrene	393.00	381.36
6	p-xylene	138.40	140.66	39	benzo[a]fluorene	403.00	396.10
7	n-propylbenzene	159.20	165.12	40	benzo[b]fluorene	398.00	401.69
8	1-methyl-2-ethylbenzene	165.20	160.23	41	benzo[c]fluorene	406.00	396.03
9	1-methyl-3-ethylbenzene	161.30	163.65	42	benzo[ghi]fluoranthene	422.00	430.09
10	1-methyl-4-ethylbenzene	162.00	162.94	43	cyclopenta[cd]pyrene	439.00	437.88
11	1,2,3-trimethylbenzene	176.10	170.85	44	chrysene	431.00	428.78
12	1,2,4-trimethylbenzene	169.40	168.70	45	benz[a]anthracene	425.00	433.98
13	1,3,5-trimethylbenzene	164.70	170.81	46	triphenylene	429.00	423.13
14	n-butylbenzene	183.30	182.74	47	naphthacene	440.00	439.01
15	1,2-diethylbenzene	183.40	182.80	48	benzo[b]fluoranthene	481.00	479.10
16	1,3-diethylbenzene	181.10	186.74	49	benzo[j]fluoranthene	480.00	479.00
17	1,4-diethylbenzene	183.80	186.41	50	benzo[k]fluoranthene	481.00	484.73
18	1-methyl-2-propylbenzene	184.80	175.81	51	benzo[a]pyrene	496.00	488.77
19	1-methyl-3-propylbenzene	181.80	180.58	52	benzo[e]pyrene	493.00	483.31
20	1-methyl-4-propylbenzene	183.80	180.55	53	perylene	497.00	483.60
21	1,2-dimethyl-3-ethylbenzene	193.90	185.07	54	anthanthrene	547.00	548.21
22	1,2-dimethyl-4-ethylbenzene	189.80	187.79	55	benzo[ghi]perylene	542.00	542.94
23	1,3-dimethyl-2-ethylbenzene	190.00	181.64	56	indeno[1,2,3-cd]fluoranthene	531.00	528.96
24	1,3-dimethyl-4-ethylbenzene	188.40	184.36	57	indeno[1,2,3-cd]pyrene	534.00	538.67
25	1,3-dimethyl-5-ethylbenzene	183.40	188.47	58	dibenz[a,c]anthracene	535.00	538.04
26	1,4-dimethyl-2-ethylbenzene	186.10	182.92	59	dibenz[a,h]anthracene	535.00	542.96
27	1,2,3,4-tetramethylbenzene	205.00	204.45	60	dibenz[a,j]anthracene	531.00	543.33
28	1,2,3,5-tetramethylbenzene	198.20	204.43	61	picene	519.00	537.76
29	1,2,4,5-tetramethylbenzene	196.80	202.28	62	coronene	590.00	602.03
30	naphthalene	218.00	217.31	63	dibenz[a,e]pyrene	592.00	593.37
31	acenaphthylene	270.00	274.60	64	dibenz[a,h]pyrene	596.00	598.10
32	acenaphthene	279.00	267.16	65	dibenz[a,i]pyrene	594.00	598.10
33	fluorene	294.00	290.76	66	dibenz[a,l]pyrene	595.00	593.22

To validate the prediction ability of model (Eq.2), a cross-validation (CV) procedure is also performed by using leave-one-out method. The average correlation coefficient (R_{ave}) for 66 predictions is 0.9983 and the root mean square error between predicted bps and experimental bps is RMS= 9.193°C. The prediction results state that

the Δbp for benzene is 39.87°C, which is obviously different from all other compounds. Delete the benzene from the whole sample set and recreate a QSPR model by using MLR and obtain Eq.3.

$$bp = 20.4966 + 174.2943m_{11} + 6.4509m_{12} + 1.3918m_{13} + 11.1033m_{22} + 11.2996m_{23}$$
$$+ 5.8424m_{33} \quad (n = 65, R = 0.9991, RMS = 6.742^o\,C, F = 5385.90) \quad (3)$$

Plot of estimated *bp*s (bp_{M2}) *vs* experimental *bp*s is shown in **Figure 3** and the results are better than ones obtained by Eq.2. The R_{ave} and *RMS* are 0.9988 and 7.542°C, respectively.

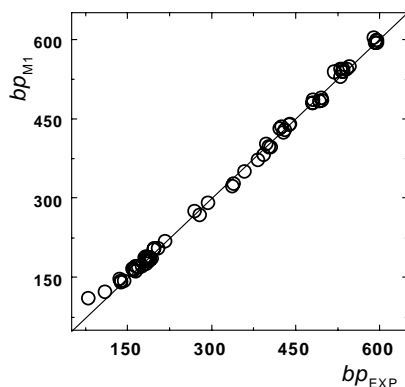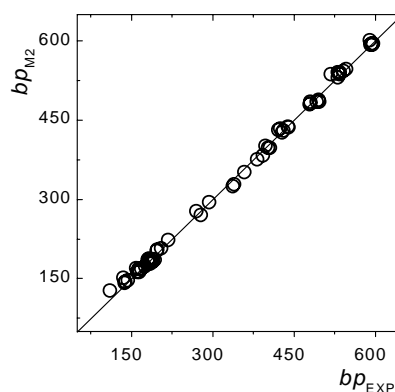**Figure 2.** Plot of bp_{M1} by Eq.2 *vs* bp_{EXP} **Figure 3.** Plot of bp_{M2} by Eq.3 *vs* bp_{EXP}

References

1. H. Wiener, *J. Am. Chem. Soc.*, **1947**, *69*, 17.
2. H. Wiener, *J. Phys. Chem.*, **1948**, *52*, 2636.
3. H. Hosoya, *Bull. Chem. Soc. Japan*, **1971**, *44*, 2332.
4. M. Randic, *J. Am. Chem. Soc.*, **1975**, *97*, 6609.
5. M. Randic, *J. Math. Chem.*, **1991**, *7*, 155.
6. M. Randic, *Chem. Phys. Lett.*, **1993**, *211*, 478.
7. A. T. Balaban, *J. Chem. Inf. Comput. Sci.*, **1985**, *25*, 334.
8. A. T. Balban, *Chem. Phys. Lett.*, **1982**, *89*, 399.
9. S. S. Liu, C. Z. Cao, Z. L. Li, *J. Chem. Inf. Comput. Sci.*, **1998**, *38*, 378.
10. S. Liu, C. Z. Cao, Z. L. Li, *Chem. in Hong Kong*, **1998**, (2), 113.
11. S. Liu, H. L. Liu, Z. N. Xia, C. Z. Cao, Z. L. Li, *J. Chem. Inf. Comput. Sci.*, **1999**, *39*, 951.
12. C. Basak, G. D. Grunwald, *J. Chem. Inf. Comput. Sci.*, **1995**, *35*, 366.