The Topogical Model of the Chromatographic Retention Index of Nitrogen-Containing Polycyclic Aromatic Hydrocarbons

Yan Chen[a] Changjun Feng[b] Xiaotao Ding[c]

School of Chemistry and Chemical Engineering, Xuzhou Institute of Technology, Xuzhou, Jiangsu, 221111, China

[a]chenyan681110@126.com, [b]fengcj@xzit.edu.cn [c]1254533304@qq.com

Keywords: PANHs, topological index, chromatographic retention index, QSRR

Abstract: Based on the topological theory and MATLAB program, molecular connectivity index (X_i), molecular shape index (K_i), electrotopological state index (E_i) and electro-negativity distance vector (M_i) were calculated for 117 nitrogen-containing polycyclic aromatic hydrocarbons. A eight-element regression model of quantitative structure-retention relationship (QSRR) was constructed using leaps-and-bounds regression (LBR). The traditional correlation coefficient (R), determination coefficient (R^2) and the cross-validation correlation coefficient (Q^2) were 0.992, 0.985 and 0.981 respectively. The robustness of the regression model was validated by Jackknife method, the correlation coefficient R was between 0.992 and 0.994. Meanwhile, the model was further tested by external validation procedure, the calculated values of the compound in validation set were in good agreement with experimental data, the average relative error was 3.04%. The regression results indicate that the model is highly reliable and has favorable predictive ability, and can better elucidate the change rule of retention indexes for nitrogen-containing polycyclic aromatic hydrocarbons.

Introduction

Polycyclic aromatic hydrocarbon (PAH) is a typical persistent organic pollutant, which is usually found in the petrochemical products, rubber, plastic, lubricating oil, antirust oil, incomplete combustion of organic compounds and other substances. The nitrogen-containing polycyclic aromatic hydrocarbons (PANHs) such as indol, quinoline, isoquinoline and their derivative exist widely in industrial waste came from petroleum industry, food industry, pesticide and pharmaceutical companies[1]. Because they are not easy to be degraded by organisms in the natural environment, They will do potential damage to the environment and human health[2]

In recent years, the quantitative structure-property/activity/retention relationship QSPR/QSAR /QSRR method was widely used in the prediction of physical and chemical properties, bioactivity and chromatographic properties of organic pollutants[3-6]. Among them, QSRR research has become a simple and effective method for the chromatographic research field, the results of the research work have also been fully recognized. On the basis of previous work[7, 8], molecular connectivity index (X_i), molecular shape index (K_i), electrotopological state index (E_i) and electro-negativity distance vector (M_i) were calculated using the MATLAB programs[9,10], and through quantitative structure-retention relationship analysis between these indexes and the chromatographic retention index of 117 PANHs, A topological model with good stability and predictive ability was established.

Materials and Methods

Chromatographic retention index of PANHs

117 PANHs were selected as research objects. The experimental values of their chromatographic retention index (RI) were obtained from the literature[11]. The RI values of 117 kinds of PANHs were shown in Table 1.

© 2015. The authors - Published by Atlantis Press

Table 1 The chromatographic retention index (*RI*) of 117 PANHs compounds

No.	PANHs	*RI*		No.	PANHs	*RI*	
		Exp.	Cal.			Exp.	Cal.
1	1-Aminoindan	207.63	223.95	60	1,4-Dimethylcarbazole	343.16	340.65
2	Quinolie	210.26	212.61	61	2-Phenylindole	346.18	344.26
3	Isoquinoline	214.14	216.13	*62	1,2-Dimethylcarbazole	347.31	353.00
4	1-Methylindole	216.90	227.47	63	2-Azafluoranthene	347.39	362.30
*5	Indole	222.66	220.09	64	1-Azafluoranthene	348.17	361.59
6	7-Azaindole	223.70	201.68	65	1,3-Dimethylcarbazole	348.45	350.40
7	2-Methylquinoline	224.13	225.78	*66	9-Cyanoanthracene	350.46	359.94
8	8-Methylquinoline	225.18	228.13	67	7-Azafluoranthene	350.50	362.31
9	1-Methylisoquinoline	229.21	228.38	68	9-Cyanophenanthrene	351.84	356.76
*10	7-Methylquinoline	231.37	232.68	69	2-Nitrofluorene	353.06	354.35
11	5-Aminoindan	232.12	243.62	70	4-Aminophenanthrene	353.97	345.14
12	3-Methylquinoline	232.47	230.37	71	9-Nitroanthracene	357.42	364.13
13	7-Methylindole	235.49	240.97	*72	1-Azapyrene	357.73	369.83
*14	4-Methylquinoline	235.77	229.69	73	4-Azapyrene	357.94	372.19
15	3-Methylindole	239.20	239.72	74	2-Azapyrene	362.43	367.48
16	2-Methylindole	240.10	234.29	75	1-Aminophenanthrene	362.62	347.13
17	2,7-Dimethylquinoline	244.04	245.37	76	1-Aminoanthracene	362.83	355.81
18	2,6-Dimethylquinoline	244.19	245.95	77	9-Aminophenanthrene	362.83	347.65
19	1,2-Dimethylindol	244.42	249.90	78	9-Aminoanthracene	363.91	364.13
*20	2, 2-Bipyridyl	247.15	221.46	79	Benzo[def]carbazole	363.92	368.72
21	2,4-Dimethylquinoline	247.96	245.00	*80	3-Aminophenanthrene	365.60	352.94
*22	4-Azabiphenyl	252.35	246.99	81	2-Aminophenanthrene	365.80	353.30
23	2,5-Dimethylindole	256.65	256.61	82	2-Aminoanthracene	367.45	361.41
24	1-Cyanonapythalene	256.75	264.03	83	3,5-Diphenylpyridine	372.84	372.59
25	2,3-Dimethylindole	257.32	259.81	*84	9-Phenylcarbazole	381.51	400.80
*26	2-Cyanonapythalene	260.88	266.37	85	Benz[c]acridine	392.60	403.99
27	5-Nitroindan	261.55	266.24	86	Benz[a]acridine	398.65	405.08
28	1-Aminopaphthalene	262.98	250.37	87	1-Azabenz[a]anthracene	400.00	411.20
29	2-Aminoaphthalene	265.53	255.79	*88	4-Azachrysene	401.16	403.02
*30	2,3,5-Trimethylindole	273.61	283.18	89	Benzo[a]carbazole	402.22	405.71
31	2-Aminobiphenyl	237.63	279.16	90	1-Azachrysene	407.18	404.45
32	1-Nitronaphthalene	274.95	271.19	91	Benzo[b]carbazole	409.63	414.20
33	4-Azafluorene	279.85	283.14	92	3-Aminofluoranthene	409.97	403.33
*34	2-Nitronaphthalene	280.63	278.97	93	2-Azachrysene	411.49	408.69
35	3-Methyl-2-aminonaphthalene	283.73	281.91	94	Benzo[c]carbazole	411.89	401.26
36	2-Nitrobiphenyl	290.25	300.46	95	4-Aminopyrene	412.31	410.78
*37	Phenazine	294.37	282.80	*96	2-Aminopyrene	413.83	417.92
38	4-Aminobiphenyl	298.05	283.74	97	1-Aminopyrene	415.39	410.65
39	Benzo[h]quinoline	301.94	306.68	98	1-Nitropyrene	421.48	429.66
40	Acridine	304.04	309.11	99	2,2-Biquinoline	422.56	411.15
*41	Acridan(9,10-dihydroacridine)	304.11	318.07	100	7,9-Dimethylbenz[c]acridine	438.32	435.75
42	Benzo[f]quinoline	307.94	307.76	101	5,7-Dimethylbenz[a]acridine	438.38	434.78
43	Phenanthridine	307.94	305.12	*102	7,10-Dimethylbenz[a]acridine	439.46	433.30
44	3-Nitrobiphenyl	310.09	307.61	103	2-Aminobenzo[c]phenanthrene	450.10	448.66
45	Carbazole	311.71	305.56	104	4-Aminobenzo[c]phenanthrene	451.51	442.69
*46	4-Nitrobiphenyl	314.59	308.52	105	10-Azabenzo[a]pyrene	455.40	466.33
47	3-Methylbenzo[f]quinoline	320.26	320.37	106	6-Aminochrysene	463.19	444.95
48	2-Methylbenzo[f]quinoline	320.50	324.76	107	9,10,12-Trimethylbenz[a]acridine	466.79	463.30
*49	2-Methylacridine	324.34	329.56	108	Dibenz[a,c]phenazine	474.08	461.44
50	1-Methylcarbazole	324.45	325.68	*109	5-Aminochrysene	487.88	442.58
51	4-Aminofluorene	325.11	320.66	110	Dibenz[a,h]acridine	488.55	500.40
52	1-Aminofluorene	327.21	324.90	111	Dibenzo[a,i]carbazole	490.57	506.27
53	3-Methylcarbazole	328.81	327.46	112	Dibenz[a,j]acridine	490.66	501.42
*54	3-Aminofluorene	329.08	330.92	113	6-Nitrobenzo[a]pyrene	501.71	514.46
55	2-Methylcarbazole	329.61	328.33	114	Dibenzo[a,g]carbazole	502.30	501.80
56	9-Methylacridine	331.15	321.50	*115	Dibenzo[c,g]carbazole	502.92	496.75
57	4-Methylcarbazole	331.88	320.93	116	7-Aminobenzo[a]pyrene	511.98	509.09
*58	2-Aminofluorene	331.91	332.11	117	6-Aminobenzo[a]pyrene	515.66	500.05
59	6-Phenylquinoline	340.84	344.28				

Construction method of topological index

Topological index method is one of the most convenient methods for QSPR/QSRR/QSAR research. At present, the more commonly used topological index are molecular connectivity index, molecular shape index, electrotopological state index and molecular electronegativity distance vector, which are constructed based on the method of graph theory.

Using Chem3D Ultra 9.0, molecule structures of 117 PANHs were built, and saved as .mol file, then called the above files, in the MATLAB environment, 10 molecular connectivity indexes, 4 molecular shape indexes, 46 electrotopological state indexes and 91 molecular electronegativity distance vector were calculated by MATLAB programs. We got 151 topological indexes as molecular descriptors.

Screening of molecular descriptors

According to the principle of statistics, the number of variables is less than 5% of the independent variable, and its contribution to the dependent variable can be neglected. Therefore, the independent variables which number was less than 6(117×5%)were removed, and the remaining 26 topological indexes were used to characterize the molecular structure of 117 PANHs.

Multiple linear regression method

Using the remaining 26 topological indexes of 117 PANHs as independent variables, and the chromatographic retention index as dependent variables, We had chosen the best variables of relativity with Chromatographic properties by MINITAB 14 software, and then built up the quantitative structure- retention relationship (QSRR) mathematical model between these topological index and RI. Meanwhile, Kubinyi (function FIT, Kubinyi) [12-13] was introduced to judge the stability and prediction ability of the model, that the calculation formula is:

$$\mathrm{FIT} = \frac{R^2(y-b-1)}{(y+b^2)(1-R^2)} \quad (1)$$

In the formula, y is the sample size of the compounds, b is the number of variables. The bigger is *FIT*, the more stable is the model, and the better is the ability of prediction.

Results and discussion

QSRR equation of the *RI*

By employing MINITAB 14.0 program, leaps-and-bounds regression method was carried out, with the results between *RI* and topological indexes of 117 PANHs presented in Table 2.

Table 2 Results of the topological index and RI with the leaps-and-bounds regression

No.	R	R^2	R_{adj}^2	S	F	FIT	Variables
1	0.955	0.913	0.912	24.298	1203.295	9.173	X_1
2	0.983	0.966	0.965	15.267	1613.367	26.768	X_1, M_{15}
3	0.986	0.971	0.971	14.018	1283.190	30.028	X_1, M_{15}, M_6
4	0.988	0.976	0.975	12.887	1144.112	34.246	X_1, M_{15}, M_6, M_{18}
5	0.989	0.978	0.977	12.337	1000.929	34.750	X_1, M_{15}, M_6, M_{18}, K_1
6	0.990	0.980	0.979	11.820	910.557	35.229	X_1, M_{15}, M_6, M_{18}, K_1, E_{19}
7	0.991	0.982	0.980	11.464	830.847	35.932	X_1, M_{15}, M_6, M_{18}, K_1, E_{19}, M_2
8	0.992	0.984	0.983	10.601	852.633	36.696	X_1, M_{15}, M_6, M_{18}, K_1, E_{19}, M_2, M_{14}
9	0.992	0.985	0.984	10.462	778.573	35.486	X_1, M_{15}, M_6, M_{18}, K_1, E_{19}, M_2, M_{14}, X_{12}

where R is the traditional correlation coefficient, R^2 is the determination coefficient, R_{adj}^2 is the square of adjusted correlation coefficient, S is the standard deviation of the regression and F is the Fisher ratio. From Table 2, The *FIT* value gradually increased, and the turning point was 36.696, which showed that the eight element model had the best stability and prediction ability. Corresponding multiple regression equations were shown as follows:

$$RI = 27.535 + 48.059\,X_1 - 24.068\,K_1 + 4.343\,E_{19} - 4.759\,M_2 - 7.641\,M_6 - 0.908\,M_{14} + 1.703\,M_{15} - 1.459\,M_{18} \quad (2)$$

n =117, R = 0.992, R^2 = 0.984, S = 10.601, F = 778.573

The calculated values of *RI* given by the formula (2) were listed in Table 1, which were in agreement with the corresponding experimental values. The average relative error was 2.22%.

Stability test of the model

LOO cross validation of the model

By LOO cross validation correlation coefficient (Q^2) was 0.982 and slightly smaller than0.985; cross validation standard deviation was 10.682, slightly larger than 10.601 This shows that the stability and prediction ability of the model are ideal.

Jackknifed test of the model

To test whether there was any "abnormal value" in model (2), we carried through the stability test based on Jackknifed method[14]. Concerning the researched was big samples (the capability was more than 30), we applied the elimination way group by group. Namely, every time we eliminated the compounds whose serial number contains 1, 2, 3 … 0 on unit order in the sample, and then established the model with the rest compounds' RI. The average value of these 10 correlation coefficients was 0.992, which was consistent with the model (2). and R's fluctuating extension was very small. Fig 1 was control graph of 10 correlation coefficients. From Figure 1, all values were in the control area (between 0.9905 and 0.9940.)it suggest that the eight element model has good robustness.

External validation of the model

In order to further verify the robustness of the model, we texted the model by an external validation. The 117 samples were divided into the training set and the test set. The training set was used to set up the model, and the test set was used to be predicted. We selected 93 compounds randomly as training set samples, the remaining 24 compounds (with * mark in Table 1) as the external test set sample. The predicted values of the retention index of the test set was close to the experiment values.

The model of the training set was used to estimate the chromatographic retention index, the average relative error was 3.04%. Fig 2 was the plot of predicted against experimental values of chromato- graphic retention indexes

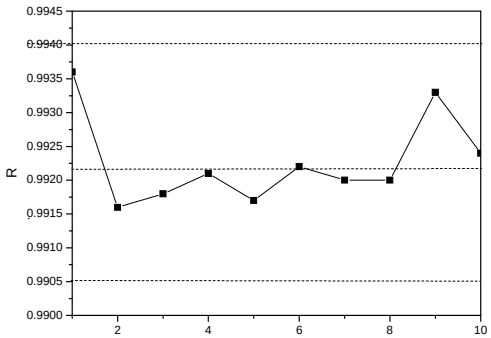

Fig.1 The control chart of Jackknifed correlation coefficient

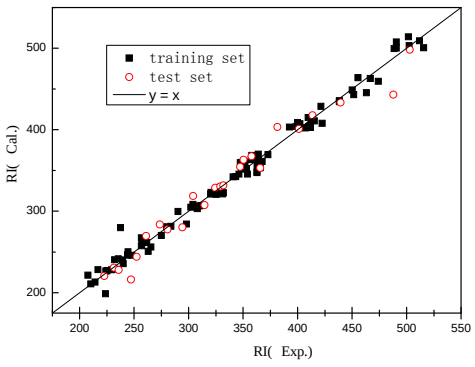

Fig.2 Plot of calculated against experimental values of retention indexes

Conclusion

There are a lot of factors affecting gas chromatography retention index, but the molecular interaction including dispersion force, induction force, orientation force and hydrogen bond between the components and the stationary phase is the main factor. Nitrogen containing polycyclic aromatic hydrocarbons are weakly polar molecules, so the molecular interaction between molecules is mainly based on the dispersion forces. The X_1 and K_1 in the model reflect the size and shape of the molecules, so the value can be used to characterize the dispersion force. E_{19} corresponding to the structure of aNa (a is the conjugated bonds in the aromatic ring), that is, the nitrogen atoms in the ring, the polar

groups, can be used to characterize the induction force and orientation force, At the same time, the electronegativity of the nitrogen atom is relatively large, so when the stationary phase has active hydrogen atom, it can form hydrogen bonds, which also can be used to characterize the hydrogen bonds. M_2, M_6, M_{14}, M_{15}, M_{18} are the interaction between the first class atom (CH_3-) and second class atom (-CH_2-), the first class atom and the sixth class atom(-N-), the second class atom and the second class atom, the second class atom and the fourth class atom(＞C＜), the second class atom and the sixth class atom, respectively. M_{14}, M_{15} and M_2, the interaction between polar groups, can be used to characterize the dispersion force, and M_{18}, M_6, the interaction between polar groups and non polar groups, can be used to characterize the induction force and orientation force. So the molecular connectivity index, molecular shape index, electrotopological state index and molecular electronegativity distance vector reveal the factors affecting gas chromatography retention index of PANHs. The cut error ratio (i.e., R^2) of the model (2) is 98.4%, and only 1.6% of the other factors that affect the RI are not disclosed.

In summary, the model has good correlation, robustness and disclose the essential factors that affect the gas chromatographic retention index of the compounds. It is reasonable that four kinds of topological indices are used to characterize the molecular structures of the PANHs.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (No. 21272095)

References

[1] Z D Xu, R H Shi and M Geng. J China Agric. Univ. Vol. 13 (2008), 57-62 (in Chinese) .

[2] M Jiang, Y M Li, G W Gu. Acta Sci. Circumst. Vol. 25(2005), 1253-1258 (in Chinese).

[3] C J FENG, W H Yang. Chinese J. Struct. Chem., Vol. 33(2014):830-834 (in Chinese).

[4] C J FENG, W H Yang, L L Mu. Chinese J. Struct. Chem., Vol. 27(2008):575-587 (in Chinese).

[5] C Wang, W Wang, L Q Pan, et al. Chem. Vol. 76 (2013): 929- 934 (in Chinese).

[6] Y Chen, W Yue, B Wang. J. Wuhan Univ.(Nat. Sci.Ed.), Vol. 60(2014): 52-56 (in Chinese).

[7] Y Chen. Bull Sci Technol, Vol. 29 (2013):16-19 (in Chinese).

[8] Y Chen. Food Sci.Vol. 32(2011):274-277 (in Chinese).

[9] Q. N. Hu, Y. Z. Liang, Y. L.Wang. Comp. App. Chem. Vol.20 (2003):386-390 (in Chinese).

[10] T. Zhang, Y.Z. Liang, C.X. Zhao. Chin. J. Anal. Chem. Vol. 34(2006):1607-1610 (in Chinese).

[11] C Wang, W Wei, L Q Pan, et, al. Chem. Vol.76(2013): 929-934 (in Chinese).

[12] L. H. Hall, L. B.Kier. J. Chem. Inform. Mod. Vol. 35(1995):1039-1045.

[13] L. S.Urra, M. P.Gonzalez, M.Teijeira. Bio. Med. Chem., Vol. 15(2007): 3565-3571.

[14] W S Dietrich, N D Dreyer, C Hansch. J Med Chem, Vol. 23(1980): 1201- 1205