

## 一种新的拓扑指数用于饱和链烃类化合物的 结构/性质相关性研究

姚瑜元 许禄\* 袁秀顺

(中国科学院长春应用化学研究所, 长春, 130022)

本文从距离矩阵出发, 结合分子中各原子的支化度, 产生一种新的拓扑指数  $Y_x$ , 并将它用于研究饱和链烃类化合物的理化性质, 获得了比较满意的结果。与其它已有的拓扑指数比较, 该指数具有良好的结构选择性和相关性。

**关键词:** 拓扑指数, 饱和链烃类化合物, QSPR.

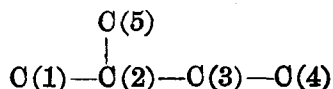
化合物的同分异构现象非常普遍, 利用分子的结构式虽然能直观地区分异构体之间的差异性, 但是, 很难作定量地描述。拓扑指数是以一种数学量来表征化合物的分子结构, 用它来进行定量结构与活性相关性(QSAR)以及定量结构与性质相关性(QSPR)研究是目前非常活跃的领域。迄今为止, 所建议的拓扑指数已经逾百种, 其中, 比较著名的有 Wiener( $W$ )<sup>[1~3]</sup>指数, Randić( $\chi$ )<sup>[4]</sup>指数, Hosoya( $Z$ )<sup>[5]</sup>指数, Balaban( $J$ )<sup>[6]</sup>指数和 Bonchev( $I_D$ )<sup>[7]</sup>指数等。利用这些拓扑指数来讨论饱和链烃的性质已经得到了许多有益的结果<sup>[8~10]</sup>。在作者的实验室中已经建议了广义  $\alpha_N$  指数, 并且将其成功地应用于含有杂原子及多重键的磷类萃取剂<sup>[11~13]</sup>。广义  $\alpha_N$  指数的另一显著特点是可以很好地区分同分异构体<sup>[14]</sup>, 这是迄今诸多的拓扑指数尚未做到的。但是, 广义  $\alpha_N$  指数的计算较为复杂, 因而, 我们试图寻找一个计算更为简单的图的不变量, 去表征分子的结构, 以期在构效关系研究中得到运用。

理想的拓扑指数应同时具有较高的结构选择性和良好的相关性, 所谓结构的选择性, 即不同的分子结构所求得的数学量(即拓扑指数)不一样。选择性好, 相关性不一定好, 如化合物在美国化学文摘服务社(CAS)的登录号 RN, 其唯一性甚好, 但它不具有与化合物性质的任何相关性。一般来说, 同时满足选择性和相关性是比较困难的, 如 Wiener 指数, 相关性在很多情况下较好, 但选择性较差。而 Balaban 指数, 其选择性通常较好, 但相关性有时较差。

为了表征化合物的分子结构, 本文提出一种新的拓扑指数( $Y_x$ ), 以其对化合物的性质进行预测和预报。对于饱和链烃类化合物应用的结果表明, 此种拓扑指数是标记化合物性质的一种可取的指数。

### 方 法

将分子中原子之间的距离和各原子的支化度作为原子的特征, 构成  $N+1$  阶矩阵来描述  $N$  个原子所组成的分子。以 2-甲基丁烷为例, 其隐氢图为:



此分子的距离矩阵( $D$ )和原子的支化度矩阵( $B$ )分别为:

$$D = \begin{bmatrix} 0 & 1 & 2 & 3 & 2 \\ 1 & 0 & 1 & 2 & 1 \\ 2 & 1 & 0 & 1 & 2 \\ 3 & 2 & 1 & 0 & 3 \\ 2 & 1 & 2 & 3 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 1 \\ 1 \end{bmatrix}$$

原子的支化度定义为该原子所连接的边的个数. 将上述两个矩阵合并在一起组成  $N+1$  阶增广矩阵( $X$ ):

$$X = \begin{bmatrix} 1 & 0 & 1 & 2 & 3 & 2 \\ 3 & 1 & 0 & 1 & 2 & 1 \\ 2 & 2 & 1 & 0 & 1 & 2 \\ 1 & 3 & 2 & 1 & 0 & 3 \\ 1 & 2 & 1 & 2 & 3 & 0 \end{bmatrix}$$

此矩阵的协方差矩阵( $V$ )为:  $V = X'X$ . 其中,  $X'$  是  $X$  的转置矩阵, 对本例有:

$$V = \begin{bmatrix} 16 & 12 & 6 & 8 & 14 & 12 \\ 12 & 18 & 10 & 8 & 10 & 14 \\ 6 & 10 & 7 & 6 & 7 & 10 \\ 8 & 8 & 6 & 10 & 14 & 8 \\ 14 & 10 & 7 & 14 & 23 & 10 \\ 12 & 14 & 10 & 8 & 10 & 18 \end{bmatrix}$$

求出矩阵  $V$  的全部本征值. 假如定义新的拓扑指数为  $Y_x$ , 则其表达式为:

$$Y_x = \log(1 + N \cdot \sum_{i=1}^{N+1} \lambda_i)$$

其中  $\lambda_i$  为特征根( $i=1, 2, \dots, N+1$ ),  $N$  为原子个数. 在本例中, 所求得的协方差阵的 6 个本征值分别为:  $\lambda_1=0.0000$ ;  $\lambda_2=66.9964$ ;  $\lambda_3=5.6109$ ;  $\lambda_4=0.4325$ ;  $\lambda_5=14.9602$ ;  $\lambda_6=4.0000$ . 则其  $Y_x$  指数为:

$$Y_x = \log[1 + 5 \cdot (0.0000 + 66.9964 + 5.6109 + 0.4325 + 14.9602 + 4.0000)] = 2.6637$$

## 结果与讨论

**$Y_x$  指数与其它拓扑指数的比较** 对于一种新的拓扑指数, 通常以饱和链烃类化合物为例来评估其性能的优劣. 因为饱和链烃类化合物的异构体可以穷举, 同时很多物理化学性质都已经测定. 我们也沿用这一方案, 为此, 应用如上方法计算了 2~9 个碳的饱和链烃类化合物分子  $Y_x$  指数, 并将它们分别列于表 1 和表 2.

Razinger<sup>[15]</sup>等定义了  $O(s)$  为拓扑指数对结构的选择性系数, 来定量地描述拓扑指数的结构选择性.

$$O(s) = N(val)/N(str)$$

其中:  $N(val)$  为拓扑指数可区分开的异构体数,  $N(str)$  为给定的异构体的总数. 表 3 为相对于 5~9 个碳的饱和链烃类化合物的几种拓扑指数的  $O(s)$  值.

表1 2~8个碳的饱和链烃类化合物的拓扑指数和相应的沸点

链烷烃分子	分子图	$\chi$	$\chi_x$	b.p.(°C)	链烷烃分子	分子图	$\chi$	$\chi_x$	b.p.(°C)
1) 乙烷		1.000	0.9542	-88.630	21) 2, 2, 3-三甲基丁烷		2.943	3.1858	80.882
2) 丙烷		1.414	1.7404	-42.070	22) 辛烷		3.914	3.7471	125.660
3) 丁烷		1.914	3.3032	-0.500	23) 2-甲基庚烷		3.770	3.6885	117.647
4) 2-甲基丙烷		1.732	2.1959	-11.730	24) 3-甲基庚烷		3.808	3.6483	118.925
5) 戊烷		2.414	2.7566	36.074	25) 4-甲基庚烷		3.808	3.6340	117.709
6) 2-甲基丁烷		2.270	2.6637	27.800	26) 3-乙基己烷		3.846	3.6065	118.534
7) 2, 2-二甲基丙烷		2.000	2.5809	9.500	27) 2, 2-二甲基己烷		3.561	5.5898	106.840
8) 己烷		2.914	3.1364	68.740	28) 2, 3-二甲基己烷		3.681	3.5678	115.600
9) 2-甲基戊烷		2.770	3.0527	60.271	29) 2, 4-二甲基己烷		3.664	3.5808	100.429
10) 3-甲基戊烷		2.808	3.011	63.282	30) 2, 5-二甲基己烷		3.626	3.6242	100.103
11) 4, 2-二甲基丁烷		2.561	2.9370	49.741	31) 3, 3-二甲基己烷		3.621	3.5285	111.060
12) 2, 3-二甲基丁烷		2.643	2.9605	57.988	32) 3, 4-二甲基己烷		3.718	3.5367	117.725
13) 庚烷		3.414	3.4622	98.420	33) 2-甲基-3-乙基戊烷		3.718	3.5181	115.650
14) 2-甲基己烷		3.270	3.8918	90.052	34) 3-甲基-3-乙基戊烷		3.682	3.4784	118.259
15) 3-甲基己烷		3.308	3.4111	91.850	35) 2, 2, 3-三甲基戊烷		3.481	3.4691	109.843
16) 3-乙基戊烷		3.346	3.3047	93.475	36) 2, 2, 4-三甲基戊烷		3.417	3.5160	99.238
17) 2, 2-二甲基戊烷		3.061	3.2799	79.197	37) 2, 3, 3-三甲基戊烷		3.504	3.4522	114.760
18) 2, 3-二甲基戊烷		3.181	3.2702	89.784	38) 2, 3, 4-三甲基戊烷		3.553	3.4943	113.460
19) 2, 4-二甲基戊烷		3.126	3.3092	80.500	39) 2, 2, 3, 3-四甲基丁烷		3.250	3.3946	106.470
20) 3, 3-二甲基戊烷		3.121	3.2201	86.064					

\* 表中 $\chi$ 为 Randić 指数。

表 2 9 个碳的饱和链烃类化合物的  $Y_x$  指数

链烷烃分子	分子图	$Y_x$	链烷烃分子	分子图	$Y_x$
1) 壬烷		3.9996	19) 3-甲基-3-乙基己烷		3.7353
2) 2-甲基辛烷		3.9508	20) 3,3-二乙基戊烷		3.6887
3) 3-甲基辛烷		3.9143	21) 2,3,5-三甲基己烷		3.7791
4) 4-甲基辛烷		3.8923	22) 2,3,4-三甲基己烷		3.7354
5) 3-乙基庚烷		3.8530	23) 2,4-二甲基-3-乙基戊烷		3.7087
6) 4-乙基庚烷		3.8305	24) 2,2,5-三甲基己烷		3.8056
7) 2,6-二甲基庚烷		3.8978	25) 2,2,4-三甲基己烷		3.7619
8) 2,5-二甲基庚烷		3.8584	26) 3,3,5-三甲基己烷		3.7397
9) 2,4-二甲基庚烷		3.8385	27) 2,2,3-三甲基己烷		3.7485
10) 2,3-二甲基庚烷		3.8408	28) 2,3,3-三甲基己烷		3.7192
11) 3,5-二甲基庚烷		3.8176	29) 2,2-二甲基-3-乙基戊烷		3.6915
12) 2-甲基-4-乙基己烷		3.7932	30) 3,3,4-三甲基己烷		3.6947
13) 3,4-二甲基庚烷		3.7982	31) 2,3-二甲基-3-乙基戊烷		3.6687
14) 2-甲基-3-乙基己烷		3.7725	32) 2,3,4,4-四甲基戊烷		3.6737
15) 3-甲基-4-乙基己烷		3.7509	33) 2,3,3,4-四甲基戊烷		3.6498
16) 2,2-二甲基庚烷		3.8660	34) 2,2,3,3-四甲基戊烷		3.6301
17) 3,3-二甲基庚烷		3.8043	35) 2,2,4,4-四甲基戊烷		3.7010
18) 4,4-二甲基庚烷		3.7816			

表 3 6种拓扑指数相应的  $C(s)$  值

指 数	饱 和 链 烃 $[N(str)]$									
	戊 烷(3)		己 烷(5)		庚 烷(9)		辛 烷(18)		壬 烷(35)	
	$N(val)$	$C(s)$	$N(val)$	$C(s)$	$N(val)$	$C(s)$	$N(val)$	$C(s)$	$N(val)$	$C(s)$
Wiener( $W$ )	8	1.0	5	1.0	7	0.78	16	0.89	16	0.46
Hosoya( $Z$ )	3	1.0	5	1.0	9	1.0	14	0.78	23	0.66
Randic( $\chi$ )	3	1.0	5	1.0	9	1.0	16	0.89	28	0.80
Bonchev( $I_D$ )	3	1.0	5	1.0	9	1.0	18	1.0	33	0.94
Balaban( $J$ )	3	1.0	5	1.0	9	1.0	18	1.0	35	1.0
$Y_x$	3	1.0	5	1.0	9	1.0	18	1.0	34	0.97

由表 3 可见,选择性最好的是 Balaban 指数,直到 9 个碳(35 种异构体)未发生简并(即不同的异构体为同一指数值)。其次为由本文所建议的  $Y_x$  指数。这说明,  $Y_x$  具有良好的结构选择性。

如前所述,作为一种表征分子结构的拓扑指数,不但应具有很好的结构选择性,还应与化合物的性质之间存在良好的相关性。目前,在众多的拓扑指数中,Randic 指数是应用最为广泛的一种,为了与该指数比较,本文将  $Y_x$  和 Randic 指数( $\chi$ )分别与 1~8 个碳的饱和链烃类化合物的沸点作相关分析,得到如下方程:

$$b.p.(^{\circ}C) = -130.4268 + 67.4850\chi \quad (1)$$

$$r=0.988 \quad s=7.578 \quad n=39$$

和

$$b.p.(^{\circ}C) = -189.1172 + 84.0410Y_x \quad (2)$$

$$r=0.987 \quad s=7.981 \quad n=39$$

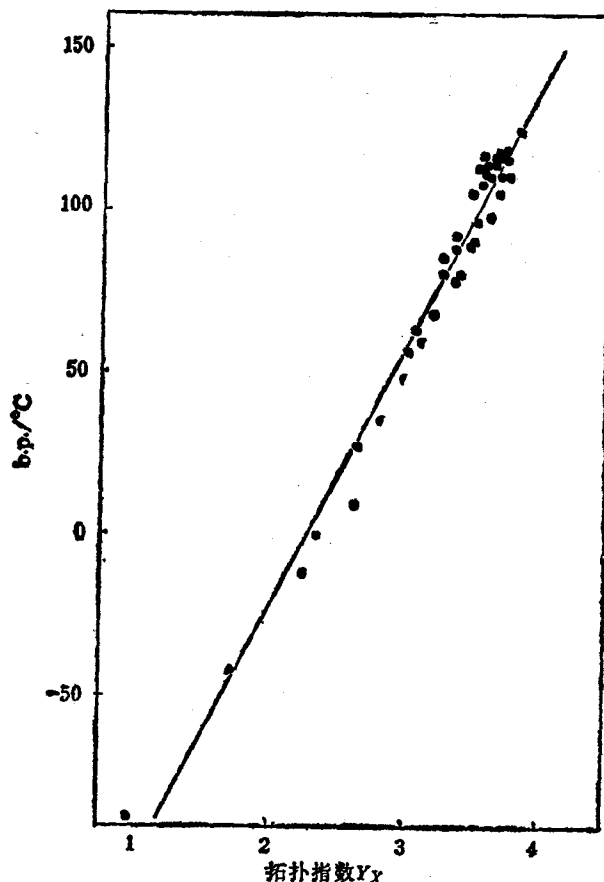
其中, $r$  为相关系数, $s$  为标准偏差, $n$  为回归点数。

与 Randic 指数相比,  $Y_x$  指数具有较好的结构选择性和与其近似的相关性。这说明,  $Y_x$  作为一种新的拓扑指数能很好地描述饱和链烃类分子支化结构特征。图 1 为 1~8 个碳原子的饱和链烃类化合物的沸点与  $Y_x$  指数的相关图。

拓扑指数  $Y_x$  与饱和链烃类分子其它物理化学性质的相关性 将  $Y_x$  指数与饱和链烃类化合物标准生成焓的对数值  $[\log(-\Delta H_f^{\circ})]$  作一元线性相关分析得如下方程:

$$\log(-\Delta H_f^{\circ}) = 1.7674 + 0.1589Y_x \quad (3)$$

$$r=0.9661 \quad s=0.0249 \quad n=39$$

图 1 拓扑指数  $Y_x$  与饱和链烃类化合物沸点的关系

式(3)的预报值和实验值的相对平均偏差为:  $d=1.3132\%$  图2为二者的关系曲线。

此图呈现了一定规律性, 即正烷烃化合物的  $\log(-\Delta H_f^\circ)$  与  $Y_x$  指数具有良好的线性相关性, 随着  $Y_x$  的增大而增大。而每一个相同碳原子数组成的链烃组都具有一个负斜率的相关直线, 并且每一条分立的直线的斜率都近似于相等。这提示我们应用如下的二元回归方程可以

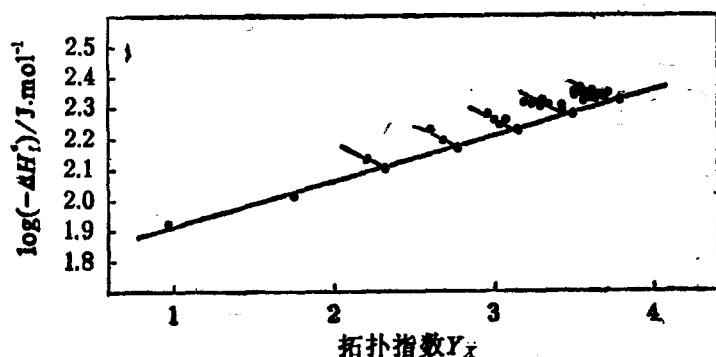


图2 拓扑指数  $Y_x$  与饱和链烃类化合物标准生成焓的关系

更好地描述饱和链烃类化合物的标准生成焓与  $Y_x$  指数的相关性:

$$\begin{aligned} \log(-\Delta H_f^\circ) &= 1.8259 + 0.0604Y_x \\ &\quad + 0.0376N_C \quad (4) \\ r &= 0.9804 \quad F = 445.7688 \\ n &= 39 \end{aligned}$$

其中  $N_C$  为碳原子数,  $F$  为  $F$  检验值。

式(4)的预报值和实验值的相对平均偏差为:  $d=0.6945\%$ 。可见, 方程(4)比方程(3)具有更好的相关性和预报能力。

$Y_x$  与这类化合物的色谱保留指数的对数  $[\log(RI)]$ , 汽化热的对数  $[\log(\Delta H_v)]$ , 热容 ( $C_p$ ), 临界温度 ( $T_c$ ) 以及临界压力 ( $p_c$ ) 作相关分析得到如下回归方程:

色谱保留指数:

$$\begin{aligned} \log(RI) &= 2.1022 + 0.2148Y_x \\ r &= 0.988 \quad s = 0.019 \quad n = 33 \quad d = 0.5662\% \end{aligned}$$

汽化热:

$$\begin{aligned} \log(\Delta H_v) &= 4.0473 + 0.1445Y_x \\ r &= 0.9958 \quad s = 0.0152 \quad n = 18 \quad d = 0.3802\% \end{aligned}$$

热容:

$$\begin{aligned} C_p &= -177.7702 + 120.6693Y_x \\ r &= 0.9867 \quad s = 13.5398 \quad n = 23 \quad d = 4.2066\% \end{aligned}$$

临界温度:

$$\begin{aligned} T_c &= -71.0808 + 98.4375Y_x \\ r &= 0.9973 \quad s = 6.3919 \quad n = 20 \quad d = 3.5457\% \end{aligned}$$

临界压力:

$$\begin{aligned} p_c &= 55.8843 - 8.3724Y_x \\ r &= -0.9917 \quad s = 0.9500 \quad n = 20 \quad d = 2.1835\% \end{aligned}$$

对如上所讨论的几种热力学性质, 拓扑指数  $Y_x$  均随着碳原子个数的增多而增大。当原子个数相同时, 随着支化程度的增大而减小。例如己烷的5种异构体(表1中的第8~12个化合物),  $n$ -己烷无支链, 其  $Y_x$  指数值最大 ( $Y_x=3.1364$ ), 其次为含一个甲基支链的第9和第10个化合物 ( $Y_x=3.0527$  和  $3.0191$ ), 化合物11和12具有两个甲基支链, 它们的  $Y_x$  最小, 其值分别为 2.9370 和 2.9605。

本文利用拓扑指数  $Y_x$  对饱和链烃类化合物的各种物理化学性质进行了分析, 结果表明,  $Y_x$  指数作为一种图的特征量能很好地反映饱和链烃类分子的支化结构, 因而, 它具有良好的结构选择性和相关性。可以期待, 拓扑指数  $Y_x$  将有可能成功地应用于 QSAR/QSPR 的研究中。

## 参 考 文 献

- [1] Wiener, H., *J. Am. Chem. Soc.*, **1947**, *69*, 17.
- [2] Wiener, H., *J. Am. Chem. Soc.*, **1947**, *69*, 2636.
- [3] Wiener, H., *J. Chem. Phys.*, **1947**, *15*, 766.
- [4] Randic, M., *J. Am. Chem. Soc.*, **1975**, *97*, 6609.
- [5] Hosoya, H., *Bull. Chem. Soc. Jap.*, **1971**, *44*, 2332.
- [6] Balaban, A. T., *Theor. Chim. Acta.*, **1971**, *53*, 355.
- [7] Bonchev, D.; Trinajstić, D., *J. Chem. Phys.*, **1977**, *67*, 4517.
- [8] Seybold, P. G.; May, M.; Bagal, U. A., *J. Chem. Educ.*, **1987**, *64*, 575.
- [9] Hosoya, H.; Hosoi, K., *J. Chem. Phys.*, **1976**, *64*, 1065.
- [10] Wilkins, G. L.; Randic, M., *Theor. Chem. Acta.*, **1980**, *53*, 45.
- [11] 王化云, 吕天雄, 许 禄, 汪尔康, 苏 锵, *化学学报*, **1990**, *48*, 1159.
- [12] 王化云, 许 禄, 苏 锵, *化学学报*, **1991**, *49*, 424.
- [13] 王化云, 许 禄, 苏 锵, *应用化学*, **1991**, *8*, 18.
- [14] 王化云, 许 禄, 苏 锵, *化学学报*, **1992**, *50*, 23.
- [15] Razinger, M.; Chretien, J. R.; Dubois, J. E., *J. Chem. Inf. Comput. Sci.*, **1985**, *25*, 23.

## A New Topological Index for Research on Structure-Property Relationships of Alkanes

Yao, Yu-Yuan Xu, Lu\* Yuan, Xiu-Shun

(Changchun Institute of Applied Chemistry, Academia Sinica, Changchun, 130022)

### Abstract

In this paper, based on distance matrix and branch vertex of atoms in a molecule, a new topological index ( $Y_x$ ) has been developed to be used in research on physical and chemical properties of alkanes. It is concluded that this index bears good structure selectivity and relativity when the results from index were compared with that of other ones.