

一个用于有机物生物活 (毒) 性研究的新拓扑指数 A Novel Topological Index for the Study of Biological Activity (Toxicity) of Organic Compounds

唐平 余训民*

Tang Ping Yu Xunmin

(湖北江汉大学环保系 武汉 430010)

(Dept. of Environmental Protection, Jiangnan Univ., Wuhan, Hubei, 430010, China)

摘要 提出一个新的拓扑指数 Y_i , 即隐氢图中的所有相邻边的一阶分子连接性指数的乘积之和占分子中有效碳原子的分率. 用 Y_i 研究取代苯、甲基苯、卤代苯、卤代烃、氯苯、含氮杂环化合物和直链伯醇的生物活 (毒) 性, 建立相应的相关方程 ($r = 0.9203 \sim 0.9914$). 计算出部分取代苯类化合物对发光菌 (*Photobacterium phosphoreum*)、呆鲦鱼 (*Pimephales promelas*)、大型蚤 (*Daphnia magna straus*)、酵母菌 (*Saccharomyces cerevisiae*) 和绿藻 Green algae 的毒性数据. 新的拓扑指数易于计算, 有较高的结构区分能力, 新方法计算方便, 物理意义明确.

关键词 拓扑指数 有机物 分子连接性 生物活 (毒) 性

中图法分类号 O 625. 7; O 621. 13

Abstract A novel topological index (Y_i) reflecting the structure of organic compounds are developed and employed in the study of biological activity (toxicity) of organic compounds, such as substituted benzenes, methylbenzene, benzene halide, aromatic halide, chlorobenzene, nitrogen-containing heterocycles, and linear alcohol. The index (Y_i) is the ratio of production sum of connectivity index of all the first order molecules with neighboring sides to effective carbons in molecules in the hydrogen-suppressed graph. The toxicities of some substituted benzenes were calculated for *Photobacterium phosphoreum*, *Pimephales promelas*, *Daphnia magna straus*, *Saccharomyces cerevisiae*, Green algae. The Y_i has a close correlation with the selected biological activities of compounds ($r = 0.9203 \sim 0.9914$). This method is convenient in calculation, and has a clear physical significance and a good ability in structure-distinguishing.

Key words topological index, organic compounds, molecular connectivity, biological activity (toxicity)

用分子拓扑学理论研究分子的结构及其与性质 / 活 (毒) 性 (即 QSPR/QSAR) 的关系被越来越多的化学家所关注^[1-5]. 拓扑指数是用来表征分子拓扑图某种特征的不变量, 以此实现分子结构信息的数值化. 一个好的拓扑指数不仅包含丰富的分子结构信息, 能充分反映分子图的连接信息和化学环境, 而且还具有较低的简并度和有效的结构 - 性质 / 活 (毒) 性相关性. 已发现的拓扑指数中能兼具这两种性质的并不多, 特别是化合物的物性与活 (毒) 性总是以不同的方式依赖于分子的内在结构. 因此, 研究完备的新的拓扑指数不仅具有重要的理论意义, 也具有广泛的应用价值. 本文根据分子拓扑理论, 构建了一个包含

有机物分子结构大部分信息的新拓扑指数 Y_i , 并用 Y_i 研究了它与取代苯、含氮杂环化合物和直链醇对生物活 (毒) 性的相关性.

1 计算方法

1.1 新的拓扑指数 Y_i

在众多的拓扑指数中, 比较完备且应用最广泛的是 Randic-Kier^[1] 提出的分子连接性ⁿ $^n i$ ($n = 0, 1, 2, \dots, n$), 但存在一些指数的物理意义不明确, 低阶的对异构体的区分性较小, 高价的计算又较复杂, 而且对电负性较大原子或原子团的结构信息很难准确反映^[6]. Randic^[4] 一阶分子连接性指数¹ i 的定义为:

$$^1 i = \sum ^1 x_k = \sum (V_i V_j)^{-0.5}, \quad (1)$$

式 (1) 中, k 为分子图中即隐氢图中的边 (共价键) 数, $^1 x_k$ 为第 k 边的一阶连接性分指数, V_i 和 V_j 为第 k 边两顶点原子 i 和 j 的点价 (成键电子数). 为了克服

2000-05-09收稿, 2001-01-04修回.

* 湖北荆州师范学院化学系, 荆州市郢都路 48号, 434100 (Department of Chemistry, Jingzhou Teachers College, Jingzhou, Hubei, 434100).

式(1)的不足,我们对式(1)进行了改进和推广,定义了一个新的拓扑指数 Y_i :

$$Y_i = \sum ({}^1x_k {}^1x_{k'}) h_c^a = \sum [(V_{ki} V_{kj})^{-0.5} (V_{k'i} V_{k'j})^{-0.5}] h_c^a, \quad (2)$$

式(2)中, k' 是与 k 相邻的边, n_c 为含碳数, a 的取值与研究的化合物类型有关,研究同类化合物的物性/活(毒)性时, $a = 0$; 研究非同类化合物的物性/活(毒)性时, $a = (n_c - 3m) / (n_c - 6m)$, 其中 m 表示隐氢图中的环数, b 取值与 m 有关: 当 $m = 1$ 时, 对于卤代苯、苯酚、苯胺, 则 $b = 2$; 对于硝基苯、苯甲酸类, $b = 1.25$; 当 $m > 2$ 时, 萘环 $b = 2.51$; 蒽、菲环 $b = 2.64$; 屈、并四苯环 $b = 2.45$; 联苯类 $b = 2.2$. n_c 的物理意义是表示分子结构中碳原子数发挥的有效作用, 分子结构中含有环会使碳的有效作用减小, 从而导致分子的内吸引力增加. Y_i 的意义是分子图即隐氢图中的所有相邻边的一阶分子连接性指数的乘积之和占分子中有效碳原子的分率. 用式(2)计算分子的 Y_i 时, 碳原子的点价按下式取:

$$V_c = Z_c - h_i, \quad (3)$$

式(3)中 Z_c 为碳原子的价电子数目, h_i 是与碳原子连接的氢原子数目. 对于化合物中的杂原子, 其点价 V_h 按下式计算:

$$V_h = Z_h (Z_h - h_i) [(8 - N_h)(n_h - 1)]^{3h_i} / (2n_h + 1), \quad (4)$$

式(4)中, Z_h , N_h 和 n_h 分别为杂原子 h 的价电子数, 所在的族数和周期数, h_i 为与杂原子连接的氢原子数目. 对于正丙酸值, 其隐氢图如下图1. Y_i 的计算方法如下:

表1 Y_i 指数与取代苯、含氮杂环和直链伯醇生物毒性的回归分析结果

Table 1 Regression analysis of Y_i index to biological toxicities of substituted benzene nitrogen-containing heterocycles and linear alcohol

化合物 Compound	生物毒性 Biological toxicity	相关方程式 Correlation equation	样本数 n	标准偏 差 S	方差比 F	相关系 数 r
取代苯 Substituted benzene	发光菌 <i>Photobacterium phosphoreum</i> (mol/L) ^[6]	$-\lg EC_{30} = 30.1594Y_i - 1.51$	25	0.27	134.24	0.9240
取代苯 Substituted benzene	呆鲱鱼 <i>Pimephales promelas</i> (mol/L) ^[6]	$-\lg EC_{30} = 38.165Y_i - 3.26$	20	0.20	151.01	0.9452
取代苯 Substituted benzene	大型蚤 <i>Daphnia magna straus</i> (mol/L) ^[6]	$-\lg EC_{30} = 30.1040Y_i - 1.89$	25	0.21	179.70	0.9472
甲基苯与卤代苯 Methylbenzene and benzene halide	酵母菌 <i>Saccharom yces cerevisiae</i> (mol/L) ^[6]	$-\lg EC_{30} = 52.6489Y_i - 8.6$	16	0.12	217.62	0.9693
卤代烃 Aromatic halide	绿藻 Green algae (mol/L) ^[6]	$-\lg EC_{50} = 35.3192Y_i - 3.18$	12	0.18	71.77	0.9369
氯苯 Chlorobenzene	$Na^+ - K^+ - ATPase$ activity (mol/L) ^[7]	$A = -81.1159Y_i + 23.92$	10	0.25	318.00	-0.9877
含氮杂环化合物 Nitrogen-containing heterocycles	Ciliate ^[8]	$\lg C = 1.8388Y_i - 2.2$	24	0.39	121.82	0.9203
直链伯醇 Linear alcohol	发光菌 <i>Photobacterium phosphoreum</i> (mol/L) ^[3]	$LC_{50} = -2.0319Y_i - 0.38$	9	0.28	403.10	-0.9914

$$V_{OH} = 6 \times (6 - 1) [(8 - 6)(2 - 1)]^3 / (2 \times 2 + 1) = 48,$$

$$Y_i = \sum ({}^1x_k {}^1x_{k'}) h_c^a = ({}^1x_1 {}^1x_2 + {}^1x_2 {}^1x_3 + {}^1x_3 {}^1x_4 + {}^1x_3 {}^1x_4) h_c^1 = [(1 \times 2)^{-0.5} (2 \times 4)^{-0.5} + (2 \times 4)^{-0.5} (4 \times 6)^{-0.5} + (2 \times 4)^{-0.5} (4 \times 48)^{-0.5} + (4 \times 6)^{-0.5} (4 \times 48)^{-0.5}] / 3 = 0.1208.$$

在化学中, 常用取代苯类、醇的同系物作一类代表性化合物用于 QSAR 研究. 我们利用式(2)~(4)计算了一些有机物 Y_i 的值, 部分在表2中. 结果表明, 新拓扑指数对有机物的结构具有良好的结构区分性和选择性.

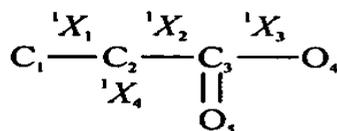


图1 正丙酸隐氢结构

Fig 1 Hydrogen-suppressed graph of n -propionic acid

1.2 相关性研究

一个好的拓扑指数不仅要具有良好的结构区分性和选择性, 更重要的是具有良好的性质/活(毒)性相关性. 为了检验 Y_i 指数与有机化合物生物活(毒)性的相关性, 我们提出如下数学模型:

$$P_i = c + c_1 Y_i, \quad (5)$$

式(5)中, P_i 表示有机物对不同生物的活(毒)性, Y_i 表示相应有机物结构特征的拓扑指数, c 和 c_1 为回归系数. 为了验证式(5)的正确性, 我们对部分取代苯、含氮杂环和醇化合物的一些对生物活(毒)性数据代入(5)式进行了相关性研究, 其结果见表1.

表 2 部分取代苯类化合物的 Y_k 指数及其对某些生物的毒性数据Table 2 The biological toxicity data and Y_k index of some substituted benzenes to some life-forms

化合物 Compound	Y_i	- $\lg EC_{50}$ (mol/L)						- $\lg C_{miz}$	
		大型蚤 <i>Daphnia magna</i> <i>straus</i>		呆鲦鱼 <i>Pimephales promelas</i>		绿藻 Green algae		酵母菌 <i>Saccharom-ycetes cerevisiae</i>	
		实验值 Observed	计算值 Calculated	实验值 Observed	计算值 Calculated	实验值 Observed	计算值 Calculated	实验值 Observed	计算值 Calculated
苯 Benzene	0.1739	3.34	3.35	3.40	3.38				
甲苯 Methylbenzene	0.1893	3.48	3.81	3.92	3.96			1.10	1.28
邻二甲苯 o-Xylene	0.1941	3.64	3.95	3.84	4.15				
间二甲苯 m-Xylene	0.1999	3.58	4.13	3.84	4.37	3.42	2.91	1.70	1.83
对二甲苯 p-Xylene	0.1996	3.69	4.12	4.21	4.36			1.74	1.82
氯苯 Chlorobenzene	0.1873	3.59	3.75	3.77	3.87	3.41	3.47	1.18	1.17
邻二氯苯 1,2-Dichlorobenzene	0.1970	4.28	4.04	4.40	4.26	3.97	3.81	1.96	1.68
间二氯苯 1,3-Dichlorobenzene	0.2013	4.21	4.17	4.41	4.42	4.01	3.96	1.87	1.91
对二氯苯 1,4-Dichlorobenzene	0.2008	4.16	4.15	4.62	4.40	3.75	3.94	1.96	1.88
1,2,3-三氯苯 1,2,3-Trichlorobenzene	0.2109	4.54	4.46	4.89	4.79	4.16	4.30	2.41	2.41
1,2,4-三氯苯 1,2,4-Trichlorobenzene	0.2137	4.70	4.54	5.00	4.90	4.10	4.40	2.45	2.56
1,2,3,4-四氯苯 1,2,3,4-Tetrachlorobenzene	0.2229	4.84	4.82						
五氯苯 Pentachlorobenzene	0.2394	5.49	5.32						
六氯苯 Hexachlorobenzene	0.2457	5.51	5.51						
溴苯 Bromobenzene	0.1911	3.81	3.86	3.89	4.03	3.55	3.60	1.40	1.37
邻二溴苯 1,2-Dibromobenzene	0.2060	4.28	4.31						
间二溴苯 1,3-Dibromobenzene	0.2090	4.35	4.40	4.53	4.23	2.32	2.31		
对二溴苯 1,4-Dibromobenzene	0.2084	4.66	4.38	5.00	4.69	4.42	4.21	2.37	2.28
邻氯溴苯 1-Chloro-2-bromobenzene	0.2025	4.37	4.21						
对氯溴苯 1-Chloro-4-bromobenzene	0.2046	4.04	4.27	4.15	4.08	2.08	2.08		
间氯溴苯 1-Chloro-3-bromobenzene	0.2051	4.39	4.28						
酚 Phenol	0.1697	3.26	3.22	3.26	3.22	0.86	0.24		
邻氯酚 2-Chlorophenol	0.1825	3.75	3.60	3.75	3.71	1.13	0.92		
对氯酚 4-Chlorophenol	0.1832	3.93	3.63	3.93	3.73				
氟苯 Fluorobenzene	0.1829	3.67	3.62						
苯胺 Aniline	0.1605	3.17	2.94	2.84	2.87				
对溴苯胺 4-Bromoaniline	0.1778	3.56	3.53	3.10	3.13				
对氯苯胺 4-Chloroaniline	0.1744	3.56	3.40	3.03	3.01				
3-氯-4-氟苯胺 3-Chloro-4-fluoroaniline	0.1817	0.83	0.88						
对氯甲苯 4-Chloromethylbenzene	0.2002	4.33	4.38	3.77	3.92	1.80	1.85		
邻氯甲苯 2-Chloromethylbenzene	0.1961	3.55	3.78	1.85	1.63				

* 为不在回归集内化合物的预测 Prediction of the compounds excluding in regression set.

从表 1 可知, Y_i 指数与上述化合物对生物的毒性具有较好的相关性. 文献 [6] 用 Y_i 处理 19 种甲苯和卤代苯类化合物对发光菌毒性的相关系数只有 0.8637, 我们达到了 0.9240. 利用表 1 中的有关方程计算了部分取代苯化合物对几种生物的毒性数据列在表 2 中.

2 结果与讨论

通过对表 1 和表 2 的结果进行分析与比较, 得出如下结论

2.1 新方法的计算值与实验值接近程度较高

根据通常评价直线方程关联程度的标准, 从表 1 看出, 本文拟合的 8 个线性回归方程, 达优级的 1 个, 良好的 2 个, 较好的 5 个, 明显高于文献 [3, 6~8] 所给的相关程度. 所建立的回归方程都在 $a = 0.01$ 显著水平下通过 F 检验. 从表 2 的计算结果可以看出, 新的方法的预测值与实验值接近的程度较高. 因此, 在化学物质使用或排污之前, 可用该文方法对其生物的毒性进行预测, 从而实行污染预防和控制.

2.2 新方法使用方便, 物理意义较明确

Y_i 及回归方程不涉及高深复杂的数学知识. 因此, 计算方法非常简单, 建构 Y_i 并不需查找任何化学数据, 直接根据分子的隐氢图及杂原子的价电子结构可获得相应的值, 所以 Y_i 的使用极其方便. 物质结构决定物质的性质/活(毒)性, 是化学中的一条基本规律. Y_i 对有机物的结构实行定量化表征, 它们与有机物的生物活(毒)性呈现比较好的相关性. Y_i 具有明确的物理意义, 回归方程也具有比较明确的物理意义, 量化了有机物的结构与活(毒)性, 从而实现了 QSAP 的研究.

2.3 新方法揭示了不同类型有机化合物的活(毒)性的内在规律

式 (5) 不仅可描述取代苯对生物的活(毒)性, 也可描述含氮杂环和醇类化合物对生物的活(毒)性.

因此, 式 (5) 用共同的因子 Y_i 揭示了不同类型有机化合物的结构与对生物活(毒)性 Y_i 之间的内在规律, 这对探讨和揭示有机化合物的物性/活(毒)性与结构之间的统一规律具有重要的理论意义.

综上所述, 新方法克服了其它方法的不足, 定量地描述了有机化合物对生物活(毒)性与其结构之间的递变规律. 新拓扑指数 Y_i 具有良好的结构选择性和活(毒)性的相关性. 由此可见, 在探讨有机化合物 QSPR/QSAR 的研究中, 分子拓扑方法不失为一种有效的方法.

参考文献

- 1 Kier L B, Hall L H. Molecular Connectivity in Structure-Activity Analysis. Ltd Letchworth, England: Research Studies Press, 1986.
- 2 Wang Liansheng. Progress of Environmental Chemistry. Beijing: Chemical Industry Press, 1995 (in Chinese). 127, 133~151, 262~272.
- 3 Ren Biye, Xu You, Chen Guobin. A novel topological index for QSPR/QSAR study of organic compounds. Acta Chimica Sinica, 1999, 57: 563.
- 4 Randic M. On characterization of molecular branching. J Amer Chem Soc, 1975, 97: 6609.
- 5 Hansch C, Leo A. Substituent constants for correlations, analysis in chemistry and biology. New York: Wiley-Interscience, 1979.
- 6 Wang Liansheng, Han Shuakui. Progress of organic contaminated chemistry. Beijing: Chemical Industry Press, 1995 (in Chinese). 1, 70, 84, 97, 134, 139, 159, 182, 195.
- 7 Yu Hongxia, Liu Zhengtao, Li Wei et al. Aromatic relationship studies between Na^+/K^+ -ATPase activity and structure of substituted compounds. Environmental Chemistry, 1997, 16 (2): 146.
- 8 Yao Yuyuan, Xu Lu, Yuan Xiushun. Study on structure-activity/property relationships of organic compounds. Acta Chimica Sinica, 1993, 51: 1041.

(责任编辑: 黎贞崇)