

• 研究论文 •

DMSO 溶剂中有机化合物 pK_a 的定量构效关系研究

谢 昆* 乔 澍 程 聪

(重庆三峡学院化学与环境工程学院 重庆 万州 404000)

摘要 应用物质结构-性质关系(QSPR, quantitative structure-property relationships)方法研究了 DMSO 溶剂中 45 个有机化合物的 pK_a 数值与电拓扑状态指数之间的定量关系, 并建立如下模型: $pK_a = -1.545 \times E_2 + 0.633 \times E_4 - 0.737 \times E_5 + 0.973 \times E_6 - 4.048 \times E_7 - 1.456 \times E_8 + 0.628 \times E_9 + 28.650$. $R=0.991$, $S=0.828$. “逐一剔除”交叉验证的结果证明模型具有良好的稳定性和较强的预测能力. 研究表明电拓扑状态指数能够有效地预测有机化合物的 pK_a 值.

关键词 电拓扑状态指数; QSPR; pK_a 值

QSPR Studies on pK_a Values of Organic Compounds in DMSO

Xie, Kun* Qiao, Shu Cheng, Cong

(College of Chemistry & Environmental Engineering, Chongqing Three Gorges University, Wanzhou, Chongqing 404000)

Abstract The quantitative structure-property relationship (QSPR) method was performed to study the relationship between pK_a values and electrotopological state indices of 45 organic compounds in DMSO. The relationship was expressed as: $pK_a = -1.545 \times E_2 + 0.633 \times E_4 - 0.737 \times E_5 + 0.973 \times E_6 - 4.048 \times E_7 - 1.456 \times E_8 + 0.628 \times E_9 + 28.650$ with $R=0.991$, and $S=0.828$. Furthermore, leave-one-out cross validation was performed, which confirmed the good stability and powerful prediction of this model. The result indicated that the electrotopological state indices could be effectively used for the prediction of pK_a values of organic compounds.

Keywords electrotopological state index; QSPR; pK_a value

物质定量结构-性质/活性关系(QSPR/QSAR)是化学计量学的一个十分重要的分支, 化学计量学的发展, 将 QSPR/QSAR 的研究又提高到了一个新的水平^[1], 并且在预测和评估实际物质的活性和性质方面有着较为广泛的应用, 比如通过 QSPR 方法预测阴离子表面活性剂亲水亲油平衡值^[2], 再如通过 QSPR 方法研究电离能和生产焓的相关性^[3]. 利用拓扑指数与化合物性质、活性参数建立 QSPR/QSAR 模型来对化合物的物理化学性质参数进行评估和预测是目前使用的一种较为简便的研究方法^[4-6].

一个化合物给予或接受一个质子的能力大小, 一般

是通过该化合物的酸碱性, 即 pK_a 值的大小来衡量^[7]. 在过去的十几年中, 以 Bordwell 小组为主的科研工作者们测量了大量有机化合物在 DMSO 溶剂中的 pK_a 值^[8-13], 这为我们的研究提供了可靠的实验基础.

本文以 Hall 和 Kier 等^[14]提出的电拓扑状态指数 (electrotopological state indices) 作为参数, 建立了 DMSO 溶剂中 45 种有机化合物 pK_a 值的数学模型. 由于电拓扑状态指数包含分子的原子类型、成键和连接度等信息, 故可以较好地表征分子结构, 并揭示影响有机分子 pK_a 值的一些因素.

* E-mail: xiekun_79@163.com; Tel.: 023-58960027

Received August 5, 2008; revised August 26, 2008; accepted October 30, 2008.

教育部“春晖计划”(No. Z2007-1-630011)、重庆市教委科学技术研究基金(No. KJ081113)、重庆三峡学院引进人才科研基金(No. 2007-SXXYRC-005)资助项目.

1 理论方法

1.1 电拓扑状态指数

电拓扑状态指数(electrotopological state indices, 简称 E -状态指数, E_n)是对分子中每一个非氢原子在分子中的拓扑环境与成键电子信息进行计算得到的一组数据, 多原子分子一般具有多个电拓扑指数. 计算过程如下:

$$I_i = \frac{\left(\frac{2}{N_i}\right)^2 \times \delta_i^* + 1}{\delta_i}, \quad \Delta I_i = \sum_{j \neq i} \frac{I_i - I_j}{(r_{ij})^2}$$

式中: I_i 为非氢原子的固有状态值, ΔI_i 为其他非氢原子对固有状态 I_i 的影响值.

成键原子 i 的电拓扑状态指数, 用符号 E_i 表示, 定义为某原子 i 的固有状态值 I_i 与其他非氢原子 j 引起的增量 ΔI_j 之和. 即

$$E_i = I_i + \Delta I_j$$

环境中结构复杂的有机污染物或复杂药物分子可能含有多个原子类型相同(固有状态值相同)的非氢原子. 处理这种情况时, 需要对每一种原子类型计算电拓扑指数而不是仅对各非氢原子进行计算. 每一种原子类型的电拓扑指数 E_n 为同类原子的 E_i 之和:

$$E_n = \sum_k E_i$$

1.2 变量选择

使用 QSPR 方法建模, 要求使用尽量少的变量尽量全面地表征化合物结构方面的信息, 这样一方面保证了选入变量的有效性和重要性, 另一方面可以提高样本容量与变量个数之间的比率, 有效地降低了随机相关的可能性, 使所建模型的质量可以得到提高, 保证了可信度和稳定性.

本研究中, 采用了快速多模型(Leaps and Bounds)回归分析的方法^[15], 根据此种方法可以得到一组含有不同变量个数的最优方程. 该算法采用剩余平方和(RSS)作为从相同数目的自变量组成的子集集合中识别出最优子集的判据. 即相同自变量个数子集的 RSS 越小, 则由这些变量回归所得到的数学模型最优. 该法将寻找最优数学模型问题转化成寻找具有最小剩余平方和值子集的问题.

2 结果与讨论










2.1 DMSO 溶剂中 45 种有机化合物 QSPR 建模

首先, 在表 1 中列出了 DMSO 溶剂中被研究的 45

种化合物所包含的所有局部结构和它们所对应的非氢原子固有状态值.

表 1 45 种有机化合物包含的 34 种非氢原子类型及相应的固有状态值(I_i)

Table 1 The intrinsic values of atomic types of 45 organic molecules in DMSO

Structure formula	I_i	E_n	Structure formula	I_i	E_n
—CH ₃	2.000	E_1		3.625	E_{18}
—CH ₂ —	1.5	E_2	=NH	4.000	E_{19}
	1.333	E_3	=N—	2.500	E_{20}
	1.250	E_4	—O—	3.500	E_{21}
—CH=	2.000	E_5	—NH—	2.500	E_{22}
	1.667	E_6	—C≡	2.500	E_{23}
—OH	6.000	E_7		4.833	E_{24}
=O	7.000	E_8	—SH	3.222	E_{25}
—NH ₂	4.000	E_9	=S	3.667	E_{26}
—F	8.000	E_{10}	—S—	1.833	E_{27}
—Cl	4.111	E_{11}	= ⁺ NH—	2.000	E_{28}
—Br	2.750	E_{12}	=CH ₂	3.000	E_{29}
—I	2.120	E_{13}	≡CH	4.000	E_{30}
	1.667	E_{14}	— ⁺ NH ₂ —	1.500	E_{31}
—O ⁻	8.000	E_{15}		1.333	E_{32}
≡N	6.000	E_{16}	—SH=	2.667	E_{33}
	2.333	E_{17}		0.472	E_{34}

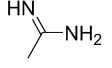
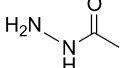
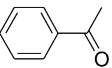
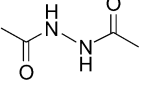
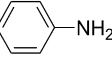
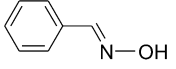
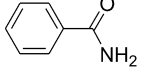
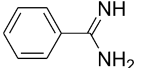
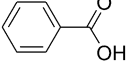
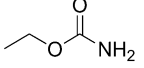
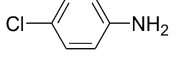
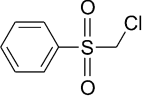
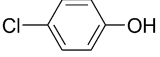
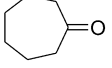
分别计算选取的 45 种化合物的所有的非氢原子的电拓扑状态指数值, 并使用快速多模型回归分析的方法, 在上述的 34 个变量当中选取了 $E_2, E_4, E_5, E_6, E_7, E_8, E_9$ 7 个变量建立七元回归模型. 通过程序拟合进行计算, 得到最终的七元回归模型如下:

$$\begin{aligned} \text{p}K_a = & -1.545 \times E_2 + 0.633 \times E_4 - 0.737 \times E_5 + \\ & 0.973 \times E_6 - 4.048 \times E_7 - 1.456 \times E_8 + \\ & 0.628 \times E_9 + 28.650 \end{aligned}$$




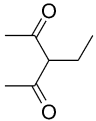
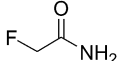
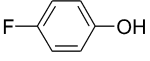


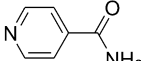



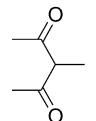

通过该模型公式计算得到这 45 种化合物的 $\text{p}K_a$ 值, 并与它们的电拓扑状态指数值一起列于表 2, 表格中的 $\text{p}K_a$ 实验值取自文献[16~36].

用实验 $\text{p}K_a$ 值对计算 $\text{p}K_a$ 值作图, 得到图 1, 并得到相关系数 $R=0.991$, 标准偏差 $S=0.828$.


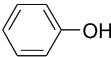
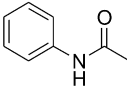
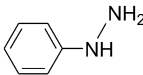

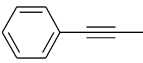
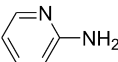
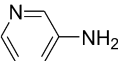




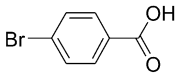
表 2 DMSO 溶剂中 45 种化合物的电拓扑指数及 pK_a 值
 Table 2 E -state Indices and pK_a values of 45 compounds in DMSO

Structure formula	E_2	E_4	E_5	E_6	E_7	E_8	E_9	pK_a	
								Exp.	Cal.
	0.000	0.000	0.000	-1.250	0.000	0.000	0.583	27.1	27.8
	0.000	0.000	0.000	-1.884	0.000	2.576	-0.401	21.8	22.8
	0.000	0.000	3.645	-2.436	0.000	-0.858	0.000	24.7	24.8
	0.000	0.000	0.000	-3.888	0.000	6.405	0.000	16.7	15.5
	0.000	0.000	-0.513	-0.845	0.000	0.000	1.358	30.6	29.1
	0.000	0.000	-1.013	-0.930	1.922	0.000	0.000	20.2	20.7
	0.000	0.000	-1.238	-3.283	0.000	3.423	1.308	23.3	22.2
	0.000	0.000	-0.541	-2.451	0.000	0.000	1.308	26.7	27.5
	0.000	0.000	-2.268	-3.880	2.385	3.201	0.000	11.1	12.2
	-1.167	0.000	0.000	-2.377	0.000	2.599	0.875	24.2	24.9
	0.000	0.000	-0.946	-1.867	0.000	0.000	1.368	29.4	28.4
	-2.875	0.000	-2.907	-2.070	0.000	6.784	0.000	23.8	23.3
	0.000	0.000	-1.640	-2.447	2.698	0.000	0.000	16.8	16.6
	-2.490	0.000	0.000	-1.178	0.000	2.188	0.000	27.8	28.2

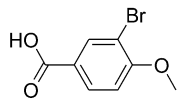
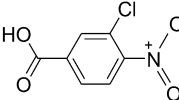
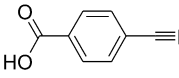
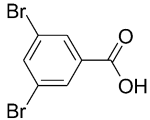
续表

Structure formula	E_2	E_4	E_5	E_6	E_7	E_8	E_9	pK_a	
								Exp.	Cal.
	-2.260	0.000	0.000	-1.202	0.000	3.463	0.000	26.4	25.9
	-2.030	0.000	0.000	-1.213	0.000	3.243	0.000	25.8	25.9
	-0.523	0.000	0.000	0.000	0.000	0.000	0.000	29.0	29.5
	-1.058	0.000	0.000	-3.379	0.000	7.191	0.000	15.3	16.5
	-2.556	0.000	0.000	-2.578	2.215	0.278	2.642	22.3	22.4
	0.000	0.000	-2.990	-3.574	2.590	0.000	0.000	18.0	16.9
	0.000	0.000	-1.185	-1.494	0.000	3.413	0.000	23.9	23.1
	0.000	0.000	0.000	-1.750	0.000	0.000	1.166	28.5	27.7
	0.000	0.000	-1.440	-3.189	0.000	3.409	1.294	21.6	22.5
	0.000	0.000	0.000	0.000	0.000	3.440	0.000	23.8	23.6
	0.000	0.000	0.000	0.000	0.000	4.572	0.000	23.4	22.0
	-1.639	0.000	0.000	0.000	0.000	2.231	0.000	29.0	27.9
	0.000	0.000	0.000	-3.448	0.000	6.729	0.000	15.1	15.5
	0.000	0.000	-1.767	0.000	0.000	4.410	0.000	22.5	23.5

续表

Structure formula	E_2	E_4	E_5	E_6	E_7	E_8	E_9	pK_a	
								Exp.	Cal.
	0.000	0.000	0.000	0.000	0.000	4.299	0.000	22.1	22.4
	0.000	0.000	-1.287	-1.345	2.632	0.000	0.000	18.0	17.6
	0.000	0.000	-0.462	-3.023	0.000	3.502	0.000	21.5	20.9
	0.000	0.000	-0.397	-0.687	0.000	0.000	1.100	28.8	29.0
	-2.143	0.000	0.000	-1.452	0.000	3.352	0.000	26.6	25.7
	0.000	0.000	-0.049	-0.458	0.000	0.000	0.000	28.0	28.2
	0.000	0.000	-0.642	-0.970	0.000	0.000	1.302	27.7	29.0
	0.000	0.000	-0.762	-0.900	0.000	0.000	1.451	28.5	29.2
	-2.677	0.000	0.000	-1.702	0.000	3.221	0.000	25.5	26.4
	0.000	0.000	-0.713	0.000	0.000	0.000	0.000	29.4	29.2
	0.000	-6.111	0.000	-3.923	0.000	2.118	-1.007	17.2	17.2
	0.000	-7.107	0.000	0.000	0.000	3.701	0.000	18.8	18.8
	0.000	0.000	-1.508	-4.713	2.460	3.297	0.000	10.2	10.4

续表

Structure formula	E_2	E_4	E_5	E_6	E_7	E_8	E_9	pK_a	
								Exp.	Cal.
	0.000	0.000	-1.403	-6.104	2.438	3.278	0.000	10.61	9.1
	0.000	0.000	-2.239	-4.791	2.501	3.397	0.000	9.68	10.6
	0.000	0.000	-2.242	-5.052	2.465	3.318	0.000	9.27	10.6
	0.000	0.000	-0.999	-5.745	2.592	3.464	0.000	8.88	8.3

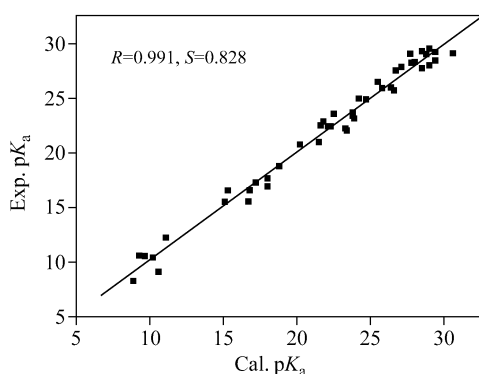


图1 pK_a 实验值和计算值的线性拟合
Figure 1 Predicted vs. experimental pK_a

为了避免随机相关, 需要对该模型进行稳定性校验. 由于研究的样本较少(45 个), 所以采用“逐一剔除”的方法, 即每次剔除一个化合物, 剩余的 44 个化合物建模, 求其相关系数, 结果为小于 0.991 的有 3 个, 在 0.991~0.992 之间有 41 个, 大于 0.992 的有 1 个. 而平均值为 0.9914, 与 0.991 也相当地接近, 所以该模型中不存在有显著影响的异常值, 该模型中不存在随机相关, 具有可以被接受的总体稳定性.

2.2 电拓扑状态指数的意义

从所建的 QSPR 模型中可以看出, DMSO 溶剂中有机化合物的 pK_a 值更多地取决于静电作用. 电拓扑状态指数中包含了较多分子中非氢原子的固有特性、成键价态信息以及拓扑环境等分子结构信息. 由固有状态值 (I_i) 的定义式可知: (1) 主量子数 N 可以反映原子的大小;

(2) δ_i^v 可以用来表示非氢原子的成键情况, 其中 $\sigma_i + \pi_i$ 反映了该原子的价键信息; (3) δ_i 则呈现了各原子局部的拓扑环境. 由 ΔI_i 的定义式可知, 它可以较好地反映分子中各非氢原子之间的电性相互作用和拓扑相互作用: (1) 电性相互作用: 以非氢原子 i 与分子中所有其它非氢原子 j 的固有状态值之差表示; (2) 拓扑相互作用: 以非氢原子 i 与其它非氢原子 j 之间的图论距离即相隔的拓扑距离(未考虑实际键长与键角, 而是基于两者之间的化学键数)表示^[37].

3 结论

通过研究 45 个有机化合物在 DMSO 溶剂中的 pK_a 值与电拓扑状态指数之间的定量关系, 我们建立了七元线性回归模型, 相关系数为 0.991, 且无随机相关.

$$pK_a = -1.545 \times E_2 + 0.633 \times E_4 - 0.737 \times E_5 + 0.973 \times E_6 - 4.048 \times E_7 - 1.456 \times E_8 + 0.628 \times E_9 + 28.650$$

从上述模型可以看出, 许多有机化合物在 DMSO 溶剂中的酸碱性和 pK_a 值与自身的电拓扑指数之间存在着定量关系. 在电拓扑状态指数中包含了大量的关于有机分子中非氢原子的固有信息和成键信息, 所以它可以很好地表征复杂的有机分子. 另一方面, 电拓扑指数能够较好地反映有机分子中的各种电子效应(如共轭效应、诱导效应、场效应等), 可见有机化合物的 pK_a 与静电作用有着密切的关系. 通过电拓扑指数预测有机物 pK_a 是一种切实可行的方法.

References

- 1 Liang, Y.-Z.; Yu, R.-Q. *Chemometrics*, Higher Education Press, Beijing, **2003**, p. 250 (in Chinese). (梁逸曾, 俞汝勤, 化学计量学, 高等教育出版社, 北京, **2003**, p. 250.)
- 2 Chen, M.-L.; Wang, Z.-W.; Zhang, G.-X.; Wang, W.-D.; Wang, Z.-N. *Acta Chim. Sinica* **2007**, *65*, 1265 (in Chinese). (陈美玲, 王正武, 张革新, 王文德, 王仲妮, 化学学报, **2007**, *65*, 1265.)
- 3 Cao, C.-Z.; Liu, J.-L. *Acta Phys.-Chim. Sinica* **2007**, *23*, 955 (in Chinese). (曹晨忠, 刘金玲, 物理化学学报, **2007**, *23*, 955.)
- 4 Zhang, Y.-T.; Li, L. *Chem. J. Chin. Univ.* **2008**, *29*, 374 (in Chinese). (张运陶, 李莉, 高等学校化学学报, **2008**, *29*, 374.)
- 5 Cao, C.-Z.; Gao, S. *Acta Chim. Sinica* **2007**, *65*, 2898 (in Chinese). (曹晨忠, 高硕, 化学学报, **2007**, *65*, 2898.)
- 6 Mu, L.-L.; He, H.-M.; Feng, C.-J. *Chin. J. Chem.* **2007**, *25*, 743 (in Chinese). (沐来龙, 何红梅, 冯长君, 中国化学, **2007**, *25*, 743.)
- 7 Fu, Y.; Liu, L.; Li, R.-Q.; Liu, R.; Guo, Q.-X. *J. Am. Chem. Soc.* **2004**, *126*, 814.
- 8 Fu, Y.; Shen, K.; Liu, L.; Liu, R.; Guo, Q.-X. *J. Am. Chem. Soc.* **2007**, *129*, 13510.
- 9 Fu, Y.; Liu, L.; Wang, Y.-M.; Li, J.-N.; Yu, T.-Q.; Guo, Q.-X. *J. Phys. Chem. A* **2006**, *110*, 5874.
- 10 Li, J.-N.; Liu, L.; Fu, Y.; Guo, Q.-X. *Tetrahedron* **2006**, *62*, 4453.
- 11 Qi, X.-J.; Liu, L.; Fu, Y.; Guo, Q.-X. *Organometallics* **2006**, *25*, 5879.
- 12 Li, J.-N.; Liu, L.; Fu, Y.; Guo, Q.-X. *Tetrahedron* **2006**, *62*, 11808.
- 13 Shen, K.; Fu, Y.; Li, J.-N.; Liu, L.; Guo, Q.-X. *Tetrahedron* **2007**, *63*, 1568.
- 14 Hall, L. H.; Kier, L. B. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1039.
- 15 Shen, Q.; Xu, L.; Li, H. *Chem. J. Chin. Univ.* **1997**, *18*, 544 (in Chinese). (申琦, 许禄, 李华, 高等学校化学学报, **1997**, *18*, 544.)
- 16 Bordwell, F. G.; Ji, G. Z. *J. Am. Chem. Soc.* **1991**, *113*, 8398.
- 17 Bordwell, F. G. *Acc. Chem. Res.* **1988**, *21*, 456.
- 18 Zhao, Y.-Y.; Bordwell, F. G.; Cheng, J.-P.; Wang, D. F. *J. Am. Chem. Soc.* **1997**, *119*, 9125.
- 19 Bordwell, F. G.; Algrim, D. J.; Vanier, N. R. *J. Org. Chem.* **1977**, *42*, 1817.
- 20 Bordwell, F. G.; Algrim, D. J. *J. Am. Chem. Soc.* **1988**, *110*, 2964.
- 21 Bordwell, F. G.; Ji, G. Z. *J. Org. Chem.* **1992**, *57*, 3019.
- 22 Bordwell, F. G.; Bares, J. E.; Bartmess, J. E. *J. Org. Chem.* **1977**, *42*, 326.
- 23 Olmstead, W. N.; Bordwell, F. G. *J. Org. Chem.* **1980**, *45*, 3299.
- 24 Bordwell, F. G.; Fried, H. E. *J. Org. Chem.* **1991**, *56*, 4218.
- 25 Bordwell, F. G.; Drucker, G. E.; Andersen, N. H. *J. Am. Chem. Soc.* **1986**, *108*, 7310.
- 26 Arnett, E. M.; Maroldo, S. G.; Schilling, S. L. *J. Am. Chem. Soc.* **1984**, *106*, 6759.
- 27 Bordwell, F. G.; Fried, H. E.; Hughes, D. L. *J. Org. Chem.* **1990**, *55*, 3330.
- 28 Bordwell, F. G.; Puy, M. V.; Vanier, N. R. *J. Org. Chem.* **1976**, *41*, 1885.
- 29 Bordwell, F. G.; McCallum, R. J.; Olmstead, W. N. *J. Org. Chem.* **1984**, *49*, 1424.
- 30 Matthews, W. S.; Bares, J. E.; Bartmess, J. E.; Bordwell, F. G. *J. Am. Chem. Soc.* **1975**, *97*, 7006.
- 31 Arnett, E. M.; Harrelson, J. A. *J. Am. Chem. Soc.* **1987**, *109*, 809.
- 32 Bordwell, F. G.; Vanier, N. R.; Matthews, W. S. *J. Am. Chem. Soc.* **1975**, *97*, 7160.
- 33 Jover, J.; Bosque, R.; Sales, J. *QSAR Comb. Sci.* **2007**, *26*, 385.
- 34 Taft, R. W.; Bordwell, F. G. *Acc. Chem. Res.* **1988**, *21*, 463.
- 35 Chong, S. S.; Fu, Y.; Liu, L.; Guo, Q. X. *J. Phys. Chem. A* **2007**, *111*, 13112.
- 36 Qi, X. J.; Fu, Y.; Liu, L.; Guo, Q. X. *Organometallics* **2007**, *26*, 4197.
- 37 Yang, W.-H.; Feng, C.-J. *J. Jishou Univ. (Nat. Sci. Ed.)* **2006**, *27*, 84 (in Chinese). (杨伟华, 冯长君, 吉首大学学报(自然科学版), **2006**, *27*, 84.)

(A0808053 Sun, H.; Zheng, G.)