

手性分子与手性拓扑指数

王 俊, 许 禄

(中国科学院长春应用化学研究所, 长春 130022)

摘要 应用自制的有机化合物结构解析专家系统 ESESOC, 对与 28 个外消旋类化合物(特殊的氨基酸)分子手性中心有关的对称性问题进行了判别, 如对称面、对称点和对称轴等. 同时用的开发的程序对这些化合物的构型进行了自动识别和标示. 在此基础上, 以手性拓扑指数为参数, 利用 Fisher 意义下的分析和人工神经网络法, 对这些化合物的对映体进行了高效液相色谱分离中先后淋洗出的的对映体进行分类判别. 结果表明, 人工神经网络法优于 Fisher 意义下的判别分析.

关键词 手性分子; 手性拓扑指数; 构型; 手性中心; Fisher 意义下的分析; 人工神经网络法

中图分类号 O652.7 **文献标识码** A **文章编号** 0251-0790(2007)01-0046-03

许多有机化合物, 特别是天然化合物分子具有手性. 含有手性的药物分子, 两个对映体在活性及代谢等方面往往存在显著差异, 从而使得其中一个异构体可能是有效的, 而另一个异构体可能是无效甚至是有毒的. 近年来, 已对手性分子的合成、分离、生物实验以及手性分子的药理实验以及对定量结构-活性/性质相关性(QSAR/QSPR)开展了研究, 但大多侧重于新的药物方面, 而有关于手性化合物预测预报的研究则研究得较少^[1]. 在研究构效关系的预测数学模型的方法中, 分子拓扑指数法运用尤为广泛. 然而, 已有的拓扑指数大部分未考虑分子的立体因素, 因而其对于手性分子的适应性较差. 本文对本实验室所建议的拓扑指数 A_m 及目前应用最为广泛的分子连接性指数进行了改进, 进而用于 28 个外消旋类化合物, 获得了比较满意的结果.

1 实验部分

1.1 研究对象

取 28 个外消旋类化合物(特殊的氨基酸)作为模式识别中的样本集. 由高效液相色谱进行分离^[2], 即由一个化合物的两个对映体先后分离的次序进行对映体的识别. 本文将先分离出来的对映体归为类 1, 后分离出来的对映体归为类 0. 后淋洗出的 28 个对映体的编号为 1L, 2L, ..., 28L; 相应的先流出来的对映体的编号为 1F, 2F, ..., 28F. 由所得判别结果可以帮助确定测试化合物的构型.

1.2 实验过程

为了对手性分子进行有效表征, 在该类拓扑指数的计算中将涉及如下问题: (1) 自动地正确识别分子的“R”或“S”构型; (2) 自动地正确识别手性原子即不对称中心; (3) 制定出有效的表征方案.

1.2.1 化合物构型的自动识别 我们研制了自动识别程序. 识别过程分为两步: (1) 根据化合物 24F 的原子编号(图 1), 判断出对称原子(在此分子中为原子 3, 4, 5, 6, 7, 8, 9 和 10)及其非对称原子(手性原子的编号为 1 和 2); (2) 程序排除对称的原子, 并对非对称原子的连接进行识别: 1 号原子连接的 4 个基团为—COOH, —C(NH₂)CH₂—, —CH₂CH=和—H, 由大到小按顺时针方向排序,

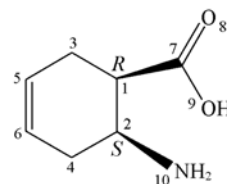


Fig. 1 Configuration of compound 24F

收稿日期: 2006-01-24.

基金项目: 国家自然科学基金(批准号: 20375039)资助.

联系人简介: 许 禄(1941 年出生), 男, 研究员, 博士生导师, 主要从事计算机化学方面的研究. E-mail: luxu@ciac.jl.cn

故程序识别为 *R* 构型; 2 号原子连接的 4 个基团为 $-\text{NH}_2$ 和 $-\text{C}(\text{COOH})\text{CH}_2-$, $-\text{CH}_2-\text{CH}=\text{CH}_2$ 和 $-\text{H}$, 由大到小逆时针方向排序, 故程序识别为 *S* 构型. 用上述步骤对于 28 个化合物(即 56 个结构)一一进行了自动标示.

1.2.2 分子手性中心的判别 手性分子必有互为镜像的构型, 不能与镜像重合是手性分子的特征. 一个分子是否有手性可通过图论和群论方法考察分子的对称性进行判断. 利用自制的有机化合物结构解析专家系统 ESESOC^[3] 可自动辨别分子所具有的对称因素: 对称面、对称中心和对称轴.

1.2.3 手性拓扑指数 我们尝试了多个拓扑指数, 并观察了多种方案, 对于本文的研究体系, 我们发现手性 A_m 指数与手性分子连接性指数的结果较好, 能很好地反映出化合物结构与活性的相关性.

A_m 指数(A_{m1} , A_{m2} , A_{m3})由本实验室所建议, 应用于多类化合物均取得良好的结果^[4~7]. 为了使 A_m 指数适合于手性化合物, 对分子中的手性原子引入适当的校正因子(即权重). 在扩展 A_m 指数时, 在距离矩阵中, 其主对角线元素由手性原子的电负性替代. 其余非手性原子在对角线上的对应位置仍然为零. 由此, 得到新的拓扑指数 eA_{m1} , eA_{m2} , eA_{m3} . 分子连接性指数在 QSAR 研究中得到了广泛的应用^[8~10]. 对该指数进行扩展的方法是对分子中手性原子的支化度引入适当的校正因子(即权重): *R* 构型的手性原子相应的支化度值加 1; *S* 构型的手性原子相应的支化度值减 1^[11].

我们已成功地将上述两个手性拓扑指数应用于 14 个 3-苯基哌啶类化合物的 QSAR 研究^[12].

2 结果与讨论

2.1 Fisher 意义下的判别分析结果

为了构造数学模型, 计算了手性 A_m 指数(eA_{m1} , eA_{m2} , eA_{m3})和手性分子连接性指数. 由于样本集中化合物的结构均不大, 未超出分子连接性指数的阶数, 故手性分子连接性指数仅计算到 6 阶. 选取其中 $e^2\chi_{pv} \sim e^6\chi_{pv}$, 加上 3 个 eA_{mi} 参数, 二者一起构成该研究中的变量集(共 8 个变量). 样本集由 28 个外消旋类化合物及其对映体共计为 56 个结构所构成, 从统计学的角度考虑, 一般认为 N/m 值不能太小(N 为试样总量, m 为所选变量数), 所得的数学模型才比较稳定. 故本文随机选取了 50 个试样作为训练集, 剩余的 6 个试样(3L, 13F, 17L, 21F, 23F, 28F)作为测试集. 在计算中, 由训练集所得结果为识别率, 由测试集所得结果为预测率. 表 1 为 Fisher 意义下判别分析^[13] 所得结果. 其中, “No.” 为序号, “Class” 为原类别, “Result” 为判别分类. 结果表明, 训练集中试样 25F, 26F, 25L, 26L 和 27L 与原归类不一致, 识别率仅为 90%; 测试集中 6 个试样均与原归类相一致, 预测率达 100%.

Table 1 The results with fisher analysis

No.	Class	Result	No.	Class	Result	No.	Class	Result	No.	Class	Result
1F	0	0	15F	0	0	1L	1	1	15L	1	1
2F*	0	0	16F	0	0	2L	1	1	16L	1	1
3F	0	0	17F*	0	0	3L	1	1	17L	1	1
4F	0	0	18F	0	0	4L	1	1	18L	1	1
5F	0	0	19F	0	0	5L	1	1	19L	1	1
6F	0	0	20F	0	0	6L	1	1	20L	1	1
7F	0	0	21F	0	0	7L	1	1	21L*	1	1
8F*	0	0	22F	0	0	8L	1	1	22L	1	1
9F	0	0	23F	0	0	9L	1	1	23L*	1	1
10F	0	0	24F	0	0	10L	1	1	24L	1	1
11F	0	0	25F	0	0	11L	1	1	25L	1	1
12F	0	0	26F	0	0	12L	1	1	26L	1	1
13F	0	0	27F	0	0	13L*	1	1	27L	1	1
14F	0	0	28F	0	0	14L	1	1	28L*	1	1

* Samples of the test set.

2.2 人工神经网络法研究结果

在人工神经网络的诸多算法中, 目前用得最多的是 BP 人工神经网络法, 但是它也存在如训练时间较长, 易于出现局部最优以及需要设置 β 和 η 参数等一些明显的不足. 本文采用贗-Newton(BFGS)

算法的人工神经网络. BFGS 算法的人工神经网络具有训练速度快、不需要设定速率常数和增添附加项等优点^[13~16]. 因此, 常进一步运用神经网络法改进 Fisher 意义下的判别分析^[17]得到的结果.

人工神经网络的 8 个输入结点, 即为 Fisher 意义下的判别分析中所选择的 8 个变量, 如 3 个 eA_{mi} 参数和 6 个 $e^2\chi_{pv} \sim e^6\chi_{pv}$ 参数. 输出结点为 1 个. 为了避免发生过训练现象, 训练时利用测试集监控训练集. 只要测试集的结果有改进, 训练继续进行. 当测试集的结果不再改善时, 训练停止. 这种方法能够有效地避免过训练. 因此, 将 56 个试样随机地分成 2 组, 其中 1 组为训练集, 共 50 个化合物; 另 1 组为测试集, 共 6 个化合物. 本文研究的 56 个样本, 其活性分为两类, 输出用 1 个结点的表示为 1 和 0. 由此, 隐蔽层结点最好不超过 4. 结果表明, 在 1, 2, 3 和 4 个隐蔽层结点中, 最好的结果是 4 个隐蔽层结点. 由此确定人工神经网络的结构为 8:4:1. 人工神经网络的研究结果表明, 所得分类结果与原归类一致. 训练中, 得到了非常理想的预测正确率, 所得结果明显优于 Fisher 意义下的判别分析方法.

参 考 文 献

- [1] Jeffrey J. S., Lee A. O., Donald F. W.. J. Med. Chem. [J], 2004, **47**: 5541—5554
- [2] Aires-de-Sousa J., Gasteige J.. Journal of Molecular Graphics and Modelling[J], 2002, **20**: 373—388
- [3] HAO Jun-Feng(郝军峰), XU Lu(许禄), HU Chang-Yu(胡昌玉). Sciences in China(Series B)(中国科学, B 辑)[J], 2000, **30**: 436—443
- [4] YAO Yu-Yuan(姚瑜元), XU Lu(许禄), YUAN Xiu-Shun(袁秀顺). J. Chem. Inf. Comput. Ci. [J], 1993, **33**: 590—594
- [5] QI Yu-Hua(齐玉华), YANG Jia-An(杨嘉安), XU Lu(许禄). Chin. J. Anal. Chem. [J], 2000, **28**(2): 223—227
- [6] ZHANG Qing-You(张庆友), XU Lu(许禄). Chem. J. Chinese Universities(高等学校化学学报)[J], 2002, **23**(11): 2125—2128
- [7] XU Lu(许禄), HU Chang-Yu(胡昌玉). Apply Chemistry Graph Theory(应用化学图论)[M], Beijing: Science Press, 2000: 152—154
- [8] Golbraikh A., Bonchev D., Tropsha A.. J. Chem. Inf. Comput. Sci. [J], 2001, **41**: 147—158
- [9] WANG Jun(王俊), ZHANG Qing-You(张庆友), XU Lu(许禄), *et al.*. Acta Chimica Sinica(化学学报)[J], 2004, **19**: 1907—1911
- [10] QI Yu-Hua(齐玉华), YANG Jia-An(杨嘉安), XU Lu(许禄). Chin. J. Anal. Chem. (分析化学)[J], 2000, **28**: 223—237
- [11] ZHANG Qing-You(张庆友), XU Lu(许禄). Chem. J. Chinese Universities(高等学校化学学报)[J], 2002, **23**: 2125—2128
- [12] XU Lu(许禄), SHAO Xue-Guang(邵学广). Methods of Chemometrics, 2nd Edition(化学计量学方法, 第二版)[M], Beijing: Science Press, 2004: 324—328
- [13] Broyden C. G.. Inst. Math. Appl. [J], 1970, **6**(6): 76—90; 222—231
- [14] Fletcher R.. Comput. J. [J], 1970, **13**: 317—322
- [15] Goldfarb D.. Math. Comput. [J], 1970, **24**: 23—26
- [16] Shanno D. F.. Math. Comput. [J], 1970, **24**: 647—656
- [17] Xu L., Ball J. W., Dixon S. L., *et al.*. Environ. Tox. Chem. [J], 1994, **13**: 841—851

Chirality Molecules and Chiral Topological Indices

WANG Jun, XU Lu*

(Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun 130022, China)

Abstract In this paper, the symmetry of 28 racemoids(unusual amino acids), such as symmetrical face, symmetrical node and axis, were recognized by using expert system of structure elucidation of organic compound, and the configurations of those compounds were recognized by using the program developed by our laboratory. On the basis, with chiral topological indices, the elution order with HPLC were performed by using Fisher analysis and artificial neural network for enantiomers of these compounds. Artificial neural network gave out a better result than that obtained by using Fisher analysis.

Keywords Chiral molecule; Chiral topological indice; Configuration; Chiral center; Fisher analysis; Artificial neural network

(Ed. : D, I)