

文章编号: 1000-4556(2007)02--

原小檗属类生物碱¹³C 核磁共振 化学位移的模拟

邓 婕^{1,2}, 梅 虎², 周 鹏^{1,2}, 孙家英¹, 吴世容¹, 李志良^{1,2*}

(1. 重庆大学化学化工学院, 重庆 400044; 2. 生物力学与组织工程教育部重点实验室, 重庆 400044)

摘 要: 将表征分子局部化学微环境的原子电性作用矢量(AEIV)和原子所处状态的原子杂化状态指数(AHSI)用于原小檗属类生物碱¹³C 核磁共振(NMR)波谱模拟中, 所得 5 参数回归模型的 r , q , RMSEE 和 RMSCV 分别为 0.982 9, 0.982 1, 7.732 9 和 7.884 3. 采用留分法(LMO)和交叉测试(cross test, CT)相结合的方式对模型稳定性和预测能力的检验结果较为满意, LMO 及 CT 对样本原子化学位移预测值与实验值的相关系数 r_{pred} 分别为 0.982 9, 0.986 5 和 0.982 1. 进一步采用该模型对¹³C NMR 未知的 4 个小檗碱类化合物共计 58 个等价共振碳原子进行预测, 所得结果与谱学模拟软件 gNMR 计算值较为接近.

关键词: 原子电性作用矢量(AEIV); 原子杂化状态指数(AHSI); ¹³C NMR 化学位移模拟; 原小檗属类生物碱; 定量构谱关系

中图分类号: O461 **文献标识码:** A

原小檗(藜)属类生物碱是一类具有多种药理作用的天然植物活性物质^[1-5], 一直以来都倍受关注. 核磁共振(NMR)与紫外、红外、质谱技术并称为四大谱学, 在有机物结构鉴定中起到极为重要的作用. 然而天然植物活性成分分子结构的多样性和复杂性给其核磁共振碳谱实验鉴定工作造成了一定难度. 为此, 现代核磁共振波谱模拟技术利用化学、数学、计算机等多学科知识, 将有机分子中处于不同化学环境的¹³C 原子进行参数化表征, 并与其化学位移(chemical shift)建立起定量构谱关系(QSSR), 该法有助于简化

收稿日期: 2006-07-31; **收修改稿日期:** 2006-10-10

基金项目: 重庆直辖市应用基础基金(No. 01-3-6); 国家“春晖计划”教育部启动基金(No. 99-4-4+37); 重庆大学自主创新科技攻关重点项目(No. 03-5-6+04-10-10).

作者简介: 邓婕(1982-), 女, 汉族, 重庆市, 硕士研究生, 主要从事药物分析与化学计量学研究. 电话: 13594007280, E-mail: dengjie18@163.com. * 通讯联系人: 李志良, 电话: 023-65106677, E-mail: zlli-cqu@163.com; zlli2662@163.com.

实验鉴定工作. Grant 和 Paul^[6] 及 Lindeman 和 Adams^[7] 最早使用 QSSR 技术对烷烃¹³C NMR 进行模拟, 之后又有系列相关工作报道^[8~10]. 本文使用笔者提出的原子电性作用矢量(AEIV)^[11]和原子杂化状态指数(ASHI)^[11]成功对 36 个天然原小檗属类生物碱 687 个共振碳原子的¹³C NMR 的化学位移大小进行了模拟和预测. 同时采用留分法(LMO)和交叉测试(CT)相结合的方式对模型稳定性和预测能力进行了严格检验, 结果较为满意. 在此基础上对 4 个未知样本分子的¹³C 原子的核磁共振化学位移值加以预测, 所得结果与谱学模拟软件 *g*NMR 计算值较为接近. 由此表明, AEIV 和 ASHI 能够有效表达原子的微观结构特征, 并与之化学位移存在较优的线性关系. 该法对建立原小檗属类生物碱的结构与其波谱特性间的构谱关系具有一定的理论和实际意义.

1 原理及方法

1.1 原子电性作用矢量(AEIV)和原子杂化状态指数(ASHI)

首先, 我们研究了化学环境影响的表达方式. 由于各类原子受周围环境的键合原子电负性与距离的 6 次方有着密切的关系^[11], 并且具有不同化学性质的原子对目标原子的作用效果不尽相同, 因此原子电性作用矢量(AEIV)将分子中的原子按元素周期表的族分成了 5 类, 并将整个的原子电负性和键长统一规划起来. 得到不同类别原子对目标原子的作用效果计算公式:

$$\nu_{it} = \sum_{j \in t, j \neq i}^{\text{all}(j)} \frac{\chi_j}{d_{i,j}^6} \quad (1 \leq t \leq 5) \quad (1)$$

式中, i 为目标原子, j 为分子中属于第 t 种类型的所有原子 ($j \neq i$); χ 取原子相对电负性大小; d_{ij} 为 i, j 两原子的相对键距, 即为连接它们的最短路径的相对键长之和.

另外, 对于原子自身杂化状态指数(ASHI), 计算式如下:

$$ASI = \sqrt{\nu/4} \cdot ((2/n)^2 \delta_{\sigma+\pi} + 1) / \delta_{\sigma} \quad (2)$$

其中, ν 是原子价电层的电子数; n 为该原子价电层所属的主量子数; $\delta_{\sigma+\pi}$ 是原子参与成 σ 和 π 键的总电子数; δ_{σ} 为成 σ 键的电子数. 详细的理论原理及常见原子的不同杂化状态 ASHI 值请参见文献[11].

1.2 计算实例

小檗碱分子仅含 sp^3 和 sp^2 杂化的碳原子(图 1). 对于 sp^2 杂化的 1 号碳原子, 氧原子对其作用项计算如下:

$$\nu_4 = \frac{\chi_{O7}}{d_{14}^6} + \frac{\chi_{O9}}{d_{15}^6} = \frac{1.3490}{(1.0000 + 0.9286)^6} + \frac{1.3490}{(1.0000 + 0.9286)^6} = 2.1398 \quad (3)$$

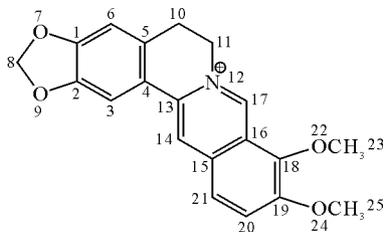


图 1 小檗碱的分子结构图

Fig. 1 Molecular structure of a protoberberine alkaloid

上式中两部分分别是 7, 9 号氧原子的作用效果, 而对于 22 和 24 号氧原子的作用效果由于距离过远而忽略不计. 同理可以分别计算出 H、C、N 原子对该碳原子的作用项: $\nu_1=0.058\ 8$, $\nu_2=3.789\ 1$, $\nu_3=0.000\ 1$. 并可从文献[11]中查出其 ASHI 值为 1.666 7. 由于原小檗属类生物碱分子中不含有 VII 族卤素原子, 因而所有 ν_5 项的值为零. 同理可计算出其它杂化态碳原子的 AEIV 描述子及 ASHI 值.

2 原小檗属类生物碱的¹³C NMR 化学位移波谱模拟

2.1 定量构谱关系建模

从文献[12]中取得了 36 个天然原小檗属类生物碱共计 687 个等价共振碳原子的¹³C NMR 化学位移数据(见附表 2*). 将这 36 个分子进行统计建模. 首先对结构图中原子进行编号, 并将每个分子的原子数、原子类型及连接关系输入到利用 C 语言自编制的应用程序 GETAEIV 进行识别、找寻最短路径并计算表征原子环境描述子 AEIV 的值. 利用多元线性回归技术(MLR)将 687 个等价共振碳原子的 AEIV 值与化学位移建立起 4 参数回归模型 M1(未包含 ASHI), 并采用留一法交互校验(LOOCV)进行验证, 所得模型如下:

$$CS = -229.691 + 9.430 \times \nu_1 + 73.302 \times \nu_2 + 49.888 \times \nu_3 + 46.212 \times \nu_4 \quad (4)$$

回归模型拟合: $n=687$, $m=4$, $r=0.961\ 9$, $RMSEE=11.470\ 2$, $F=2110.282\ 5$

交互校验预测: $n=687$, $m=4$, $q=0.960\ 2$, $RMSCV=11.715\ 2$, $F_{CV}=2\ 015.902\ 7$

可以看出, 仅用原子微环境描述子 AEIV 就可以较好模拟该组样本的化学位移情况. 进一步将表征原子自身状态的 ASHI 描述子加入到模型当中, 得到如下 5 参数回归模型 M2:

$$CS = -168.859 - 0.637 \times \nu_1 + 3.070 \times \nu_2 + 11.827 \times \nu_3 + 13.105 \times \nu_4 + 166.640 \times ASHI \quad (5)$$

回归模型拟合: $n=687$, $m=5$, $r=0.982\ 9$, $RMSEE=7.732\ 9$, $F=3878.353\ 8$

交互校验预测: $n=687$, $m=5$, $q=0.982\ 1$, $RMSCV=7.884\ 3$, $F_{CV}=3\ 725.611\ 6$

可以看到 M2 的统计质量在 M1 的基础上有了显著改进, 拟合和预测相关系数 r 和 q 分别由 0.961 9 和 0.960 2 提高到 0.982 9 和 0.982 1, 而均方根误差 RMSEE 和 RMSCV 则分别由 11.470 2 和 11.715 2 降至 7.732 9 和 7.884 3. 将加入 ASHI 前后所建模型 M1 和 M2 进行比较可以发现样本残差的绝对值显著减低且离群样本减少、计算值与观测值间的相关性增大; 另外模型 M2 中 AHSI 项回归系数显著大于前 4 项化学环境描述子系数. 这些迹象说明杂化状态对¹³C NMR 化学位移影响显著, 加入 ASHI 能有效提高模型的稳定性及预测能力. 实验数据(见附表 1*).

图 2 和 3 分别为模型 M2 对样本的计算值、LOOCV 预测值与实验观测值相关性, 可以看到几乎所有的样本点都均匀分布在过原点 45°斜线附近, 由此说明模型 M2 能够较好模拟该组样本的¹³C NMR 化学位移. 但由图可见, 模型亦出现了 4 个具有较大正向偏差的样本点, 它们分别为 584、607、621 和 637 号碳原子, 其文献[12]对应的实验核磁共振数据分别为 $\delta\ 168.8$ 、 $\delta\ 168.8$ 、 $\delta\ 166.7$ 和 $\delta\ 161.7$, 而模型 M2 预测结果分别为 δ

* 附表 1~3 不发表, 如读者需要请直接与作者交流或索取

211.086, δ 211.086, δ 211.080 和 δ 204.050(附表 1*), 偏差分别达到 +42.286、+42.286、+44.380 和 +42.350. 它们分别属于 32、33 和 34 号分子(图 4). 对这 4 个原子所处化学环境进行分析表明, 它们具有共同微观结构特征:(1)自身为 sp^2 杂化;(2)以双键的形式连接一个羰基;(3)以单键的形式连接一个电负性较大的原子(O 或 N). 我们发现在所有样本中仅有这 4 个原子同时具备上述三个特征. 据此可以认为, AEIV 描述子可能存在以下两点不合理的方面:(1) AEIV 按距离 6 次方递减处理, 该类作用衰减方式与实际可能有一定的差异;(2) AEIV 将周围原子对目标原子作用效果进行线性加和, 未考虑原子作用间的交互效应等非线性因素的存在. 由于上述两点问题在多数情况下并不显著, 但当原子所处结构特征较为特殊时则表现出较大影响, 故对上述 4 个非常见类型原子的计算结果偏差较大. 本文考虑到分子内部原子之间电性影响的复杂性, 并且异常原子数目较少, 因此未对模型做进一步改进.

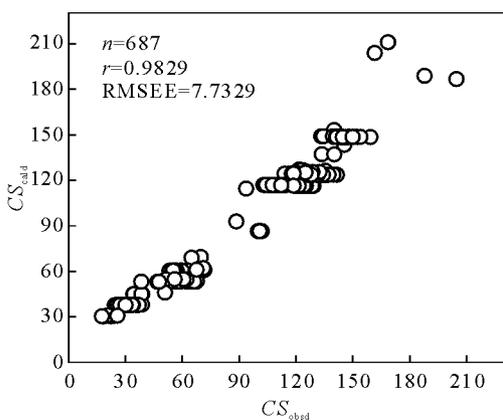


图 2 模型 M2 的估计值与观测值的相关情况

Fig. 2 Plot for M2-estimated vs. observed ^{13}C NMR chemical shifts of protoberberine alkaloids

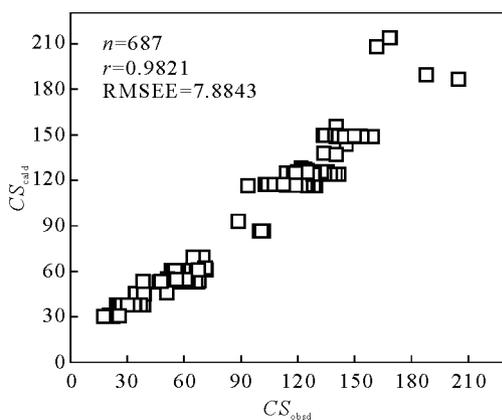


图 3 模型 M2 的 LOOCV 预测值与观测值相关情况

Fig. 3 Plot for M2-predicted vs. observed ^{13}C NMR chemical shifts of protoberberine alkaloids

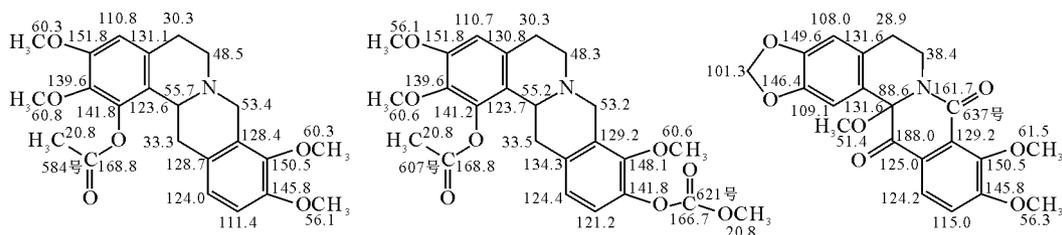


图 4 有特殊结构的分子 32、33 和 34

Fig. 4 Molecular structures of some particular protoberberine alkaloids

2.2 模型验证

留分法(LMO)检验:模型验证是统计建模中非常重要的一个环节, 其结果好坏直接决定了模型的外部预测能力和真实有效性. 为了进一步考察模型的稳定性, 我们采用留分法(LMO)对模型进行检验, 即每次留下一个分子中所含的碳原子不参与建模, 再用剩下的分子中的碳原子样本所建回归模型对该分子中碳原子进行预测, 重复该过程直到所

有碳原子都被预测一遍。对于 36 个天然原小檗属类生物碱, LMO 法交互校验为 36 次, 所得预测结果与实验值相关情况参见图 5。由图可见 LMO 总体预测能力较好, 但也出现了部分离群点, 这些点出现异常的情况与上文类似, 亦是由于结构特殊所致。

交叉测试(CT): 对于一个优秀的统计模型, 不但要对内部样本有好的拟合, 更应当对外部测试集具有较高的预测能力。Golbraikh 等^[13]最近研究结果显示: 交互检验 q^2 值的大小与模型预测能力并没有明显直接关系, 对模型预测能力的评价只能通过外部样本集即测试集来进行。

鉴于此, 作者将 36 个天然原小檗属生物碱分子均分为两部分各 18 个分子, 并互为训练/测试集进行交叉测试, 所得模型 M3 和 M4 对外部测试集样本化学位移预测值与实验值的相关系数 r_{pred} 和均方根误差 RMSEP 分别为 0.986 5, 6.973 9 和 0.982 1, 7.836 9。该结果与训练集拟合值较为接近, 表明模型具有良好的稳定性和泛化能力。

2.3 未知样本预测

采用模型 M2 对 2, 3, 8 和 9 号分子中核磁共振数据未有文献报道的 58 个碳原子进行预测, 并将预测结果与用谱学模拟软件 $g\text{NMR}$ 的计算值进行比较, 二者较为接近(附表 3*)。

3 结论

使用用于描述分子微观环境及原子状态的 AEIV 和 AHSI 指数成功对 36 个天然原小檗类生物碱¹³C 位移值进行了模拟, 同时采用 LMO 和 CT 相结合的方式对模型稳定性和预测能力进行了严格检验, 结果较为满意。在此基础上对 4 个未知样本分子的¹³C 原子的核磁共振化学位移值加以预测, 所得结果与谱学模拟软件 $g\text{NMR}$ 计算值较为接近。由此表明, AEIV 和 ASHI 具有良好的化学结构选择性和优良的谱学性质相关性。

参考文献:

- [1] Kupeli E, Kosar M, Yesilada E, *et al.* A comparative study on the anti-inflammatory, antinociceptive and antipyretic effects of isoquinoline alkaloids from the roots of Turkish Berberis species[J]. *Life Sci*, 2002, 72: 645-657.
- [2] ZHU Zhao-jing(朱照静), CHEN Shi-lin(陈士林), XIAO Ting-chao(肖廷超), *et al.* Studies on the Active-Constituents of Medical Plants(具有开发前景的药用植物活性成分)[J]. *Resource Development & Market(资源开发与市场)*, 1995, 11(1):19-21.
- [3] Wing H K, Xiao Q Y, Chi W L, *et al.* Vasorelaxant and antiproliferative effects of berberine[J]. *Eur J Pharmacol*, 2000, 399: 187-196.

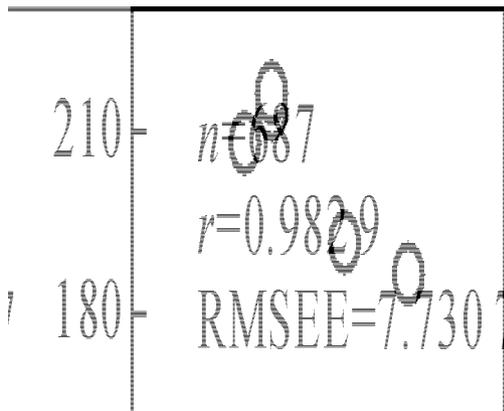


图 5 留分法的预测值与观测值相关情况

Fig. 5 Plot for LMO-predicted residual vs. observed ¹³C NMR chemical shifts of protoberberine alkaloids

- [4] Iwasa K, Kamigauchi M, Ueki M, *et al.* Antibacterial activity and structure-activity relationships of berberine analogs[J]. *Eur J Med Chem*, 1996, 31: 469–478.
- [5] Iwasa K, Kim H S, Wataya Y, *et al.* Antimalarial activity and structure-activity relationships of protoberberine alkaloids[J]. *Eur J Med Chem*, 1998, 33: 65–69.
- [6] Grant D M, Paul E G. Carbon-13 Magnetic Resonance. II. Chemical Shift Data for the Alkanes[J]. *J Am Chem Soc*, 1964, 86: 2 984–2 990.
- [7] Lindeman L P, Adams J Q. Carbon-13 Nuclear Magnetic Resonance Spectrometry Chemical Shifts for the Paraffins through C9[J]. *Anal Chem*, 1971, 43: 1 245–1 252.
- [8] Clerc J T, Sommerauer H A. Minicomputer program based on additivity rules for the estimation of ^{13}C NMR chemical shifts[J]. *Anal Chim Acta*, 1977, 95: 33–40.
- [9] Jensen K L, Barber A S, Small G W. Simulation of carbon-13 nuclear magnetic resonance spectra of polycyclic aromatic compounds[J]. *Anal Chem*, 1991, 63: 1 082–1 090.
- [10] Jurs P C, Ball J W, Anker L S, *et al.* Carbon-13 nuclear magnetic resonance spectrum simulation[J]. *J Chem Inf Comp Sci*, 1992, 32: 272–278.
- [11] Mei H(梅虎), Zhou P(周鹏), Qin R H(覃仁辉), *et al.* Calculation of ^{13}C chemical shifts of amino acids using atomic electronegativity interaction vector(原子电性作用矢量用于氨基酸化学位移计算)[J]. *Chinese J Magn Reson(波谱学杂志)*, 2005, 22(2): 163–172.
- [12] Yu De-quan(于德泉), Yang Jun-shan(杨峻山), Xie Jing-xi(谢晶曦). *Handbook of Analytical Chemistry: Vol. 5. Spectroscopy of nuclear magnetic resonance(分析化学手册第五分册:核磁共振波谱分析[M]. Beijing(北京):Chemical Industry Press(化学工业出版社), 1989. 609–611.*
- [13] Golbraikh A, Tropsha A. Beware of q^2 [J]. *Mol Graphics Mod*, 2002, 20: 269–276.

Modeling Quantitative Structure-Spectrum Relationship of ^{13}C NMR Chemical Shifts of Protoberberine Alkaloids

DENG Jie^{1,2}, MEI Hu², ZHOU Peng^{1,2}, SUN Jia-ying¹,
WU Shirong¹, LI Zhi-liang^{1,2*}

(1. College of Chemistry and Chemical Engineering, Chongqing University, Chongqing 400044, China;

2. Key Laboratory of Biomedical Engineering of Educational Ministry and Chongqing Municipality, Chongqing 400044, China)

Abstract: Atomic electronegativity interaction vector (AEIV) and atomic hybridization state index (AHSI) were employed for quantitative structure spectroscopy relationship modeling of ^{13}C NMR chemical shifts of the carbon atoms in protoberberine alkaloids. The five-parameter model yielded r , q , RMSEE and RMSCV of 0.982 9, 0.982 1, 7.732 9 and 7.884 3, respectively. Leave-one-molecule-out (LMO) method and cross test (CT) were combined to test the reliability of this model, and the results appeared satisfactory. Predicted chemical shifts of the samples were in good linear relationship with the experimental data, with r_{pred} of 0.982 9, 0.986 5 and 0.982 1 respectively. The model was also used to predict unknown ^{13}C NMR chemical shifts of 58 carbon atoms in four protoberberine compounds, and the results obtained were in agreement with calculated results using the gNMR software.

Key words: atomic electronegativity interaction vector, atomic hybridization state index, NMR, chemical shift, protoberberine alkaloids