

文章编号:1005-9830(2004)01-0062-05

用量子化学方法预测 PCDFs 的 Ah 受体 结合能力(2):变量的灰色关联分析*

黄 俊^{1**}, 余 刚¹, 王溢磊², 吴国是²

(1. 清华大学 环境科学与工程系; 2. 化学系 北京 100084)

摘要:为进一步阐明 PCDFs 的芳烃(Ah)受体结合能力与分子结构之间的关系,引入灰色关联分析方法,计算出作为描述符的 12 种结构-性质参数及一种衍生参数对于 Ah 受体结合能力的影响程度排序,并结合变量的物理意义进行分析。结果表明,最具影响的 3 种参数依次为分子最低未占据轨道能 (E_{lumo})、亲电超离域度之和 (S_r^E)、分子平均极化率 (α),这说明 PCDFs 分子的得电子能力和分子平均极化率对其 Ah 受体结合能力的影响最大。随着氯原子取代数目的增大, E_{lumo} 和 S_r^E 值变小,分子易得电子而被还原;而 α 值增大则意味着 PCDFs 分子与受体间色散力的增强,这 2 种因素促使 Ah 受体结合能力呈增大趋势。

关键词:多氯联苯并对喹啉;量子化学;灰色理论;灰色关联度分析

中图分类号:X 131, O 641.121 **文献标识码:**A **学科代码:**610.10

文献[1]基于误差反向传播神经网络(BP-ANN),利用半经验 AM1-MO 方法计算得到的电子结构和分子性质参数建立了具有较高精度和较好预测能力的 PCDFs 的 Ah 受体结合能力的定量构效(QSAR)模型,但是这个模型没有给出各个变量参数对于 Ah 受体结合能力的影响程度给出一个直观的量度。为了解决这个问题,本文引入邓聚龙创立的灰色系统理论中的灰色关联分析方法^[1],对用于建模的 12 种量子化学参数进行变量分析,以灰色关联序的形式给出各个变量参数对于 PCDFs 的 Ah 受体结合能力的影响程度大小排序。

1 算法与程序

灰色关联分析是由我国学者邓聚龙教授创立的灰色系统的重要内容之一,其实质是通

收稿日期:2002-09-17

- * 基金项目:国家重点基础研究专项经费资助项目(G1999045711);清华大学环境科学与工程研究院“985”基金项目
- ** 作者简介:黄 俊(1976-),男,江苏江阴人,博士研究生,主要研究方向:持久性有机污染物环境化学行为,E-mail:huangjun00@mails.tsinghua.edu.cn。

通过对序列曲线几何形状的分析比较,来衡量它们之间关联性大小。发展变化状态越接近,则关联度越大。它最早是在时间序列分析中被提出来的,后经改进能够用于非时间性的指标序列的分析,具有计算简便、概念清晰的特点^[2]。

1.1 灰色关联分析算法

设应变量数目为 p , 自变量数目为 m , 样本数目为 n , 则参考数列(也称母数列)记为 $Y_j = \{Y_j(k) | k = 1, 2, \dots, n\}, j = 1, 2, \dots, p$ 。比较数列(也称子数列)为 $X_i = \{X_i(k) | k = 1, 2, \dots, n\}, i = 1, 2, \dots, m$ 。

第 1 步:对原始数据进行无量纲化变换。

根据原始数列的特点,可选用不同的变换方法,较为常用的有初值化变换、均值化变换、极差变换以及效果测度变换,其中极差变换较多用于空间序列或者指标序列中,效果测度变换则适用于存在“望大”或“望小”期望的指标序列。在本研究中根据实际情况,只考虑前面 2 种变换,其公式如下:

(1) 初值化变换后生成的数列为

$$\begin{aligned} X'_i &= \{X'_i(k) = X_i(k)/X_i(1) | k = 1, 2, \dots, n\} & i = 1, 2, \dots, m \\ Y'_j &= \{Y'_j(k) = Y_j(k)/Y_j(1) | k = 1, 2, \dots, n\} & j = 1, 2, \dots, p \end{aligned}$$

(2) 均值化变换后生成的数列为

$$\begin{aligned} X'_i &= \{X'_i(k) = X_i(k)/\bar{X}_i(k) | k = 1, 2, \dots, n\} & i = 1, 2, \dots, m \\ Y'_j &= \{Y'_j(k) = Y_j(k)/\bar{Y}_j(k) | k = 1, 2, \dots, n\} & j = 1, 2, \dots, p \end{aligned}$$

式中, $\bar{X}_i(k)$ 为子序列 i 的算术平均值, $\bar{Y}_j(k)$ 为母数列 j 的算术平均值。

第 2 步:计算 $X'_i(k)$ 与 $Y'_j(k)$ 的关联系数:

$$\xi_{ji}(k) = \frac{\min_i \min_k \Delta i(k) + \rho \max_i \max_k \Delta i(k)}{\Delta i(k) + \rho \max_i \max_k \Delta i(k)}$$

式中, $\Delta i(k) = |Y'_j(k) - X'_i(k)|$, ρ 为分辨系数,一般取 0.5。

第 3 步:计算 Y_j 与 X_i 的灰色关联度:

$$\gamma_{ji} = \gamma(Y_j, X_i) = \frac{1}{n} \sum_{k=1}^n \xi_{ji}(k) \quad j = 1, 2, \dots, p; i = 1, 2, \dots, m$$

这样便得到了灰关联度矩阵 R 。根据 R 中各个行与各个列关联度的大小来判断子因素与母因素的作用,就可以分析出哪些因素具有主要影响,哪些因素具有次要影响,起主要影响的因素称为优势因素。

1.2 灰色关联分析软件

由于上面的计算均为矩阵运算,非常适合于计算机处理。利用 Visual Basic 6.0 软件可视化编程实现了灰色关联分析软件(GRA v1.0),软件内置了 6 种数据预处理方式:初值化变换、均值化变换、极差变换、上限测度变换和下限测度变换。经过对数 10 个经典算例的计算检验^[3],程序运算无误。

2 灰色关联分析步骤

以 33 种文献[4]报道的有 Ah 受体结合能力实验值的 PCDFs 为样本,即 $n = 33$;因变量

为 Ah 受体结合能力(以 pEC_{50} 表示)^[5], 即 $p = 1$; 自变量首先是文献[1]中用于建模的 12 种量子化学参数, 同时考虑到 $E_{\text{homo}} - E_{\text{lumo}}$ 这个差值对于化学物质反应活性的重要性, 将其也作为一个自变量, 因此共有 13 个自变量, 即 $m = 13$ 。这些量子化学参数都是采用下面的过程计算得到的。首先利用 Chem3D 5.0 软件编写所有 135 种 PCDFs 的分子结构输入文件, 然后导入 WinMOPAC V2.0 软件, 再采用 AM1 半经验算法优化分子结构, 计算 12 种量子化学参数, 控制计算所用的关键词为“AM1 EF PI SCFCRT = 1.D - 12 PRECISE GEO - OK LET BONDS LARGE SUPER MULLIK GRAPH POLAR”。

编制原始数据输入文件, 分辨率取方法的默认值 0.5, 然后依次选择无量纲化方法, 进行灰色关联矩阵的计算和排序, GRA 软件将计算结果保存到指定的报告文件中, 同时显示在文本区中。为考察不同预处理方法对结果的影响, 对初值化变换、均值化变换 2 种数据预处理方法分别进行计算和比较。

3 结果与讨论

3.1 灰色关联矩阵与关联度序

由于只有一个应变量, 因此对于每种数据预处理方法分别得到一个维数为 1×13 的灰色关联矩阵, 为表述方便以表格形式列于表 1 中, 下标 a、b 分别对应初值化变换和均值化变换。

表 1 灰色关联分析结果

Table 1 Results of grey relationship analysis

编号	对应的量子化学参数	δ_a	n_a	δ_b	n_b
1	$qH^+ / \text{a. c. u}$	0.662 3	5	0.830 7	5
2	$q^- / \text{a. c. u}$	0.588 4	9	0.778 8	9
3	Q_A^2	0.545 3	12	0.717 8	11
4	Q_M	0.561 2	11	0.756 6	10
5	$E_{\text{homo}} / \text{eV}$	0.640 3	6	0.824 6	6
6	$E_{\text{lumo}} / \text{eV}$	0.815 8	1	0.869 8	1
7	S_r^N	0.675 9	4	0.840 3	4
8	S_r^E	0.753 0	2	0.863 5	2
9	$\Delta H_f^0 / (1/4.18 \text{ kJ/mol})$	0.503 8	13	0.656 5	12
10	E_c / eV	0.640 3	6	0.824 6	6
11	$\alpha (\text{a. u.})$	0.711 2	3	0.856 5	3
12	$\mu (\text{a. u.})$	0.566 1	10	0.652 1	13
13	$E_{\text{homo}} - E_{\text{lumo}} / \text{eV}$	0.620 1	8	0.813 5	8

表 1 中, qH^+ 、 q^- 、 Q_A^2 、 Q_M 分别表示分子中最正氢原子净电荷、最负的原子净电荷、全部原子净电荷的平方和和原子的绝对电荷。

3.2 优势因素分析

从表 1 中的关联度序可以看出, 无论是采用初值化变换还是均值化变换, 对 Ah 受体结合能力影响最大的前 3 种结构 - 性能参数都是相同的, 依次为(影响程度由高到低): E_{lumo} 、 S_r^E 、 α 。对于其它由量子化学计算出的参数, 2 种数据预处理方法得出的关联度排序基本上是

相近的,差别只是发生在关联度较小的次要因素(即关联度排序为 10 及以后的参数),因此可以认为数据预处理方法没有引起灰色关联分析结果的变化。

3.3 几点讨论

(1) 按照前线轨道理论,化学反应多与分子的最高占据分子轨道(HOMO)和最低未占据分子轨道(LUMO)之间电子的跃迁或转移有关。 E_{lumo} 与分子的电子亲和能直接相关,其值越小,电子进入该轨道后体系能量降低得就越多,即该分子接受电子的能力就越强。而 S_r^E 也是一种用于比较分子内反应和分子间反应难易程度的重要理论参数,它同样代表了分子得电子能力的强弱。 E_{lumo} 和 S_r^E 的灰色关联度最大,表明 PCDFs 分子的得电子能力对于其 Ah 受体结合能力的大小具有重要影响。

(2) α 是表示分子大小和相互作用的物理量,分子间色散力的大小主要受分子极化率的影响^[6]。从这个角度来看,PCDFs 的平均极化率影响了其与 Ah 受体之间的色散力的大小,从而也直接影响到了 Ah 受体结合能力的大小。

(3) 将样本集中 33 个分子按氯原子取代数目(N)对 E_{lumo} 、 S_r^E 、 α 和 $pEC50$ (图 1 中以 Y 表示)作散点图,如图 1 所示。

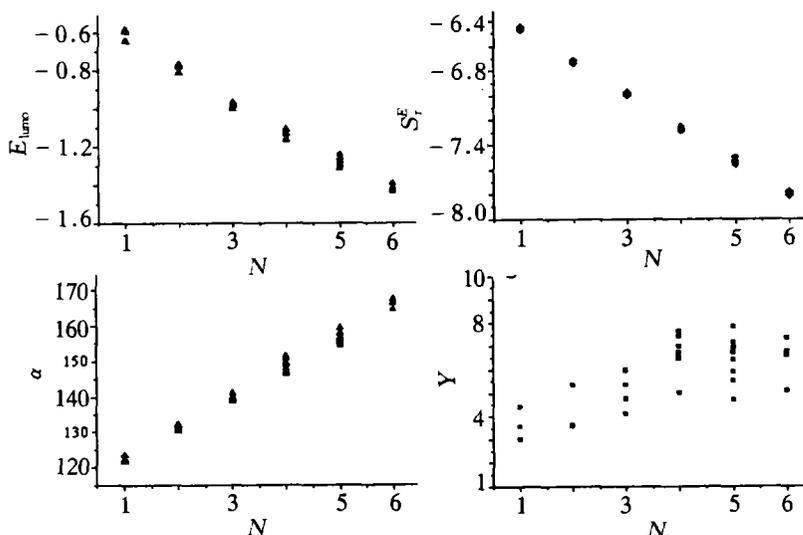


图 1 变量与分子中氯原子取代数目的关系

Fig. 1 Relationship between variables and the number of Cl atoms in the molecular

由图 1 可以清楚地看出,氯原子取代数目与 3 个主要影响因素之间存在着很好的相关性。随着氯原子数目的增加, E_{lumo} 、 S_r^E 逐渐变小,而 α 则逐渐增大。这意味着 PCDFs 分子得电子发生还原的能力增强,并且与其受体之间的分子间色散力增大。 $pEC50$ 主要受到这 2 个因素的控制,总体上呈现增大趋势。

本文在用量子化学计算所得参数结合人工神经网络预测 PCDFs 的 Ah 受体结合能力的工作基础上,将灰色系统理论中的灰色关联分析引入对自变量分析中,揭示了 PCDFs 分子的得电子能力和分子平均极化率对于 Ah 受体结合能力的影响最大。2 种不同的数据预处理方法得出的灰色关联度排序基本一致,尤其是对于主要影响因素的分析结果完全一致。这种

方法具有计算简便、结果直观的特点,同样可以用于其它物质的定量构效关系研究中。

参考文献

- [1] 黄俊,余刚,王溢磊,等.用量子化学方法预测 PCDFs 的 Ah 受体结合能力(1):神经网络模型.南京理工大学学报[J],2003,27(6):714~720.
- [2] 邓聚龙.灰色系统基本方法[M].武汉:华中理工大学出版社,1987.
- [3] 王学萌,张继忠,王荣.灰色系统分析及实用计算程序[M].武汉:华中理工大学出版社,2001.
- [4] Kari T, Juhani R. Electronic eigenvalue (EEVA): A new QSAR/QSPR descriptor for electronic substituent effects based on molecular orbital energies—A QSAR approach to the Ah receptor binding affinity of polychlorinated biphenyls (PCBs), dibenzo-p-dioxins (PCDDs) and dibenzofurans (PCDFs) [J]. Chemosphere, 2000, 41: 843~848.
- [5] So S S, Karplus M. Three-dimensional quantitative structure-activity relationships from molecular similarity matrices and genetic neural networks (2): Applications [J]. Journal of Medicinal Chemistry, 1997, 40(26): 4 360~4 371.
- [6] 邓聚龙.灰色系统预测与决策[M].武汉:华中理工大学出版社,1986.
- [7] 国永敏,李宝宗.链烷烃计算极化率与沸点的关系[J].苏州大学学报(自然科学),2000,16(3):82~85.

Predicting Ah Receptor Binding Data of PCDFs Using Quantum Chemical Method (2): Grey Relationship Analysis of Invariables

HUANG Jun¹, YU Gang¹, WANG Yi-lei², WU Guo-shi²

(1. Department of Environmental Science and Engineering;

2. Department of Chemistry, Tsinghua University, Beijing 100084, China)

Abstract: To elucidate the relationship between the Ah receptor binding affinity and molecular structures of PCDFs, grey relationship analysis approach was introduced. The magnitudes of twelve quantum chemical parameters as well as a derived parameter is calculated from the view of their effects on the Ah receptor binding affinity. The physical meanings of variables were considered for the result analysis. Three parameters of largest magnitudes were obtained, namely, the energy of lowest unoccupied molecular orbit (E_{lumo}), the electrophilic superdelocalizability (S_r^E) and the average molecular polarizability (α). It means the potential of electron reception has the most important effect on the Ah receptor binding affinity. With the increase of chlorine atom number in the molecular, E_{lumo} and S_r^E decreases, which means the molecular becomes more intensive to get electron and to be reduced. At the same time, α increases, which means the molecular has larger affinity for electrons and stronger force on the receptor. Those factors caused the increasing trend of Ah receptor binding affinity.

Key words: PCDFs; quantum chemistry; grey system theory; grey relationship analysis