

环氧化合物分子结构与致变活性的关系

白乃彬¹, 顾玛力², 谢飞², 王兴平² (1. 中国科学院生态环境研究中心, 北京 100085; 2. 北京联合大学化学工程学院, 北京 100023)

摘要:由化学物质毒性效应登录 (RTECS) 1998 年光盘系统检索获得 86 种环氧化合物对鼠沙门氏菌致变活性数据. 应用主成分分析法从 11 种分子描述符中选择出 5 种对致变活性有明显影响的描述符: 碳原子数, 苯环数, 氢原子数, 烷基数和氧原子数. 通过样本学习集训练并优化神经网络结构, 建模分类预报. 对于 86 个样本, 低和高 2 类活性, 正确分类率达到 92%. 结果表明: 3~4 个拼接苯环的环氧化物的特殊结构具有很强的致变活性.

关键词:神经网络; 环氧化物; 致变活性; 构效关系

中图分类号: R994.6 文献标识码: A 文章编号: 0250-3301(2001)06-05-0112

Quantitative Structure and Activity Relationship for Epoxides

Bai Naibin¹, Gu Mali², Xie Fei², Wang Xingping² (1. Research Center for Eco-Environmental Sciences, Chinese Academy Sciences, Beijing 100085, China; 2. The Chemical Engineering College of Beijing Union University, Beijing 100023, China)

Abstract: Mutagenic activity data of 86 epoxides were collected from a CD system of Registry of Toxic Effects of Chemical Substances (RTECS 1998 Version). By using principal component analysis, the 5 kinds from 11 kinds of molecular structure descriptors, such as the number of carbon atom, the number of benzene-cycle, the number of hydrogen atom, the number of substituted alkyl group and the number of oxygen atom which can have influence on their mutagenic activity very much were chosen. After learning of above samples, The Back-Propagation Method (BP) network structure was optimized, the learning sets of sample were classified, then unknown samples were predicted one by one; for 86 samples, the correctly classified rate can reach to 92% between the low class and the high class. Results shown that the special molecular structure, the epoxides of polycyclic aromatic hydrocarbons contained 3~4 cycles, can be responsible for their very high mutagenic activities.

Key words: Back-Propagation Method (BP); neural networks; epoxides; mutagenic activity; QSAR

环境污染多环芳烃 (Polycyclic Aromatic Hydrocarbons, PAHs) 具有较强的致变和致癌作用. 近期研究表明: 多环芳烃进入人体后经过代谢, 形成环氧化物, 为致癌的前兆体 (见图 1). 通过定量构效关系 (Quantitative Structure and Activity Relationship, QSAR) 研究以预报未经测试的环氧化物的致变性能, 具有理论和实际价值^[1~4]. 鉴于以往环氧化物 QSAR 研究所用模式限制, 样本数较少, 本研究则应用神经网络模式继续扩大样本数目; 预测值与实验值比较, 正确分类率达到 92%.

1 样本结构描述符与其致变活性数据矩阵

脂肪烃和芳香烃环氧系列化合物致变活性数据来自 RTECS 光盘系统 1998 年版, 共检索得样本 86 组:

其中, 对鼠沙门氏菌致变活性大于等于 $1 \mu\text{mol}/\text{皿}$ 的为 24 组; 而小于 $1 \mu\text{mol}/\text{皿}$ 的为 62 组. 然而, 目前尚无专家公认的致变活性的分级标准, 就本体系而言, 可相对依 $1 \mu\text{mol}/\text{皿}$ 为分界线划为低 (+) 和高 (++) 2 类活性, 列入表 1.

2 主成分分析选择有影响的分子描述符

样本矩阵做主成分分析得到 11 个主成分; 它们对应的特征值经大小排队后得到前 3 个主成分. 第 1 个主成分所含信息量占总信息量的 71%, 其表达式所含 11 个加合项中第 1 项和第 5 项权重为大, 因此, 第 1

基金项目: 国家自然科学基金项目 (29877032)

作者简介: 白乃彬 (1940~), 男, 研究员, 主要从事环境污染物质构效关系研究.

收稿日期: 2001-02-20

续表 1

化合物名称	分子结构与致变活性描述符												毒性分级	
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]
5,6-环氧-5,6-二氢-7-甲基苯并(a)蒽	19	14	1	0	3	1	0	0	0	0	0	1.9	+	++?
(-)-(1S,2R)-环氧-1,2,3,4-四氢苯并(a)蒽	18	14	1	1	3	0	0	0	0	1	0	0.001	++	++
(+)-(1S,2R)-环氧-1,2,3,4-四氢苯并(a)蒽	18	14	1	1	3	0	0	0	1	0	0	0.001	++	++
3,4-环氧-1,2,3,4-四氢苯并(a)蒽	18	14	1	1	3	0	0	0	0	0	0	0.0001	++	++
(+)-(3S,4R)-环氧-1,2,3,4-四氢苯并(a)蒽	18	14	1	1	3	0	0	0	1	0	0	0.001	++	++
(-)-(3S,4R)-环氧-1,2,3,4-四氢苯并(a)蒽	18	14	1	1	3	0	0	0	0	1	0	0.001	++	++
3,4-环氧-1,2,3,4-四氢二苯并(a,i)芘	24	16	1	1	5	0	0	0	0	0	0	0.001	++	++
4,5-二氢-4,5-环氧苯并(e)芘	20	12	1	0	4	0	0	0	0	0	0	0.003	++	++
9,10-环氧-7,8,9,10-四氢苯并(a)芘	20	14	1	0	4	0	0	0	0	0	0	0.0001	++	++
9,10-环氧-9,10,11,12-四氢苯并(a)芘	20	14	1	1	4	0	0	0	0	0	0	0.001	++	++
7,8-环氧-7,8,9,10-四氢苯并(a)芘	20	14	1	1	4	0	0	0	0	0	0	0.0003	++	++
11,12-环氧-11,12-二氢-3-甲基胆蒽	21	16	1	0	3	0	0	0	0	0	0	0.001	++	++
11,12-环氧-3-甲基胆蒽	21	14	1	0	3	0	0	0	0	0	0	0.177	++	++
1,2-环氧-1,2,3,4-四氢蒎	18	14	1	1	3	0	0	0	0	0	0	0.001	++	++
3,4-环氧-1,2,3,4-四氢蒎	18	14	1	1	3	0	0	0	0	0	0	0.001	++	++
(-)-3,4-环氧-1,2,3,4-四氢蒎	18	14	1	1	3	0	0	0	0	1	0	0.00005	++	++
(+)-3,4-环氧-1,2,3,4-四氢蒎	18	14	1	1	3	0	0	0	1	0	0	0.0005	++	++
7,8,9-环氧-8,9,10,11-四氢-7H-环庚基(a)芘	21	16	1	1	3	0	0	0	0	0	0	0.010	++	++
5,6-环氧-5,6-二氢苯并(a,h)蒽	22	14	1	0	4	0	0	0	0	0	0	0.001	++	++
1,2-环氧-1,2,3,4-四氢二苯并(a,h)芘	24	16	1	1	5	0	0	0	0	0	0	0.0001	++	++
1,2-二氢-1,2-环氧茚并(1,2,3-cd)芘	22	12	1	0	5	0	0	0	0	0	0	0.003	++	++
1,2-环氧-7,8-环氧辛烷	8	14	2	2	0	0	0	0	0	0	0	0.433	++	++?
1,2-环氧环己烷	6	10	1	2	0	0	0	0	0	0	0	10	+	+
1,2-环氧-4-(环氧乙基)环己烷	8	12	2	3	0	0	0	0	0	0	0	0.356	++	++?
3,4-环氧环己烯	6	8	1	1	0	1	0	0	0	0	0	34.8	+	+
1,2-环氧环戊烷	5	8	1	2	0	0	0	0	0	0	0	15	+	+
1,2-环氧-1,2,3,4-四氢菲	14	12	1	1	2	0	0	0	0	0	0	0.001	++	++
1,2-环氧-1,2,3,4-四氢三苯	18	14	1	1	3	0	0	0	0	0	0	0.001	++	++
[(4-甲基)-甲基]环氧乙烷	10	12	1	1	0	0	0	0	0	0	0	100	+	+

1)[1]碳原子数

[2]氢原子数

[3]氧原子数

[4]烷基数

[5]芳香环数

[6]烯键

[7]顺结构

[8]反结构

[9]外消旋

[10]内消旋

[11] $\alpha\beta$ 构象

[12]致变活性($\mu\text{mol}/\text{皿}$)

[13]毒性试验分级

[14]毒性计算分级

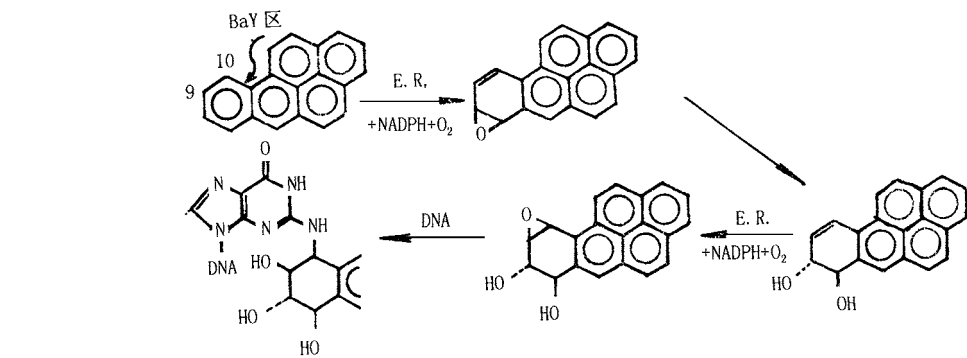


图 1 多环芳烃致癌反应历程

Fig.1 Mechanism for carcinogenic reaction of PAHs

主成分代表碳原子数和与环氧化物拼接的苯环数.如此,第2和第3个主成分其所含信息量分别为总信息量的16%和6%;分别代表氢原子数,取代环氧化物氢的烷基数及氧原子数.前3个主成分所含信息量之和已占总信息量的93%,足以保持原样本矩阵所包含的

QSAR规律,且使问题约化,变得清晰.主成分分析后筛选变量得到5个对毒性影响较大的特征分子描述符,1个致变活性描述符及86个样本所组成的特征样本矩阵.含碳原子和氢原子数较多,且带有3~4个与环氧化物相拼接的苯环的环氧化物其致变活性较高.

其它 6 个分子描述符未能进入特征样本矩阵的原因可能是在本体系中样本数较少,难以显示其对致变活性的影响。

3 BP 神经网络结构优化

多维化学数据本身有结构,不同学习集结构不同。BP 神经网络结构应适应于学习集,进行优化,才能获得最佳预报结果,以避免所谓的过拟合现象^[5]。所谓结构优化包括隐含层数和各隐含层的结点数的优化选择;包括学习参数 α 和动量因子 η 适宜数值选取。本文以特征样本矩阵为学习集;取计算值与实验值之差平方和再除样本数 (D) 为优化判据, D 值愈小,预报功能愈高。

3.1 隐含层数及各隐含层结点数选择

固定学习参数 α 和动量参数 η ,选择隐含层数和各隐含层结点数,计算结果:隐含层取 1 层,结点由 10 个逐次递减到 1 个, D 值依次为 0.0140,0.0140,0.0140,0.0170,0.0160,0.0137,0.0150,0.0150,0.0139,0.0330。5 个结点为好。再取隐含层为 2 层,结点再由 10 个逐次递减为 1 个, D 值依次为 0.0141,0.0140,0.0140,0.0140,0.0140,0.0138,0.0137,0.0140,0.0138,0.0142。仅取 1 层 (5 个结点) 与取 2 层 (第二层,4 个结点) 结果相差甚小,但取 1 层,计算速度较快。再取 3 层,效果欠佳。

3.2 学习参数 α 和动量参数 η 数值优化

η 是动量因子,合理选择它的值可在不影响预报精确度的前提下,克服迭代振荡。基于上述优化结果,隐含层取 1 层,其结点数为 5 个,学习参数 α 取 0.90,动量因子 η 由 0.70 每次降低 0.05,直至 0.30, D 值依次为 0.0138,0.0138,0.0139,0.0138,0.0138,0.0137,0.0136,0.0143 和 0.0162。以 η 值取 0.40 为佳。 α 是学习参数,合理选择它的值可以加快学习过程。动量因子 η 取 0.40,学习参数 α 由 0.90,每次降低 0.05,直至 0.45, D 值依次为 0.0136,0.0137,0.0137,0.0137,0.0138,0.0138,0.0137,0.0138,0.0138 和 0.0139。选择 α 为 0.90 为好。

3.3 BP 网络优选结构

BP 网络优选结构如图 2。

4 环氧化合物构效关系 BP 网络模式

BP 网络程序用 C 语言自编,运算在微机实现。在优化的网络结构基础上,86 个样本每次随机抽出

10% 作为预报集,剩下样本作为学习集,直至每个样本皆被预测一遍,最终预报结果亦示于表 1。预测值与实验值比较,对于低和高 2 类相对的致变活性,错分了 7 个,正确分类率达到 92%,满足统计完好标准。

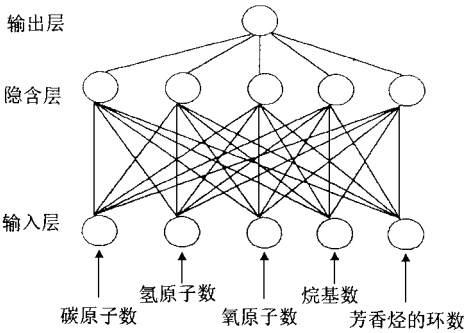


图 2 优化的神经网络结构图

Fig. 2 Optimized structure of BP network

5 小结

与芳香烃环氧化物相比较,大部分脂肪烃环氧化物致变活性为低;当与环氧化物相连接的是 3~4 个苯核稠环,如花,苯并[a]蒽等,则它们的致变活性非常高。致变活性传达了化学物质与 DNA 相互作用中的某些分子水平信息^[1]。最终致变 (致癌) 物环氧化物中环氧基打开,形成正碳离子,进攻 DNA 上的胺基,造成 DNA 损伤 (图 1)。当正碳离子与连接的较大芳香环共轭时,趋于稳定,促进了上述反应。

参考文献:

- 1 Lewis David F V, Oarke Dennis V. The renotoxicity of benzanthracenes: a quantitative structure-activity study. *Mutation Research*, 1995, (328):207~214.
- 2 Passino reader D R, Hickey J P. Toxicity to *Daphnia Pulex* and QSAR predictions for polycyclic hydrocarbon representative of Great Lakes contaminants. *Bull. Environ. Contam. Toxicol.*, 1997, **127**(59):834~840.
- 3 Langer Thierry, Hoffmann Remy D. New principal components derived parameter describing molecular diversity of heteroaromatic residues. *Quant. Struct. Act. Relat.*, 1998, **129**(17):211~223.
- 4 Langer Thierry. Molecular similarity determination of heteroaromatic ring fragments using the multivariate data analysis. *Quant. Struct. Act. Relat.*, 1996, **126**(15):469~474.
- 5 唐桂刚 白乃彬. 醛类化合物结构与对鼠急性毒性关系研究. *环境科学*, 2000, **21**(5):89~93.