信息理论指数及其在有机磷农药毒性预测研究中的应用*

彭 勤 宋瑞琨 林琼芳 鲁生业

(同济医科大学环境医学研究所，武汉 430030)

摘要 为探讨化合物结构与毒性关系研究方法，将定量描述分子结构特征的信息理论指数、价分子连接性指数应用于有机磷农药毒性预测研究，提出了基团信息参数。运用判别分析、多元非线性回归分析，依据114种有机磷农药急性毒性数据资料，按动物不同给药途径，分别建立定性、定量预测模型，定性判别回代符合率均在83.33%以上，定量回归方程相关指数 R^2 均在0.81以上，回归残差呈正态分布。预测分析，预测值均在 LD_{50} 测定值的95%可信区间内。预测效果满意。

关键词 信息理论指数，价分子连接性指数，结构活性关系，有机磷农药，毒性预测。

19世纪后期，Crum-Brow和Fraser指出[1]：有可能通过化学结构的改变，并将这种改变与生物反应联系起来进行构效关系(SAR)研究。但由于缺乏从数字上描述化合物分子的理化参数，这一理论未能得到充分的发展和应用。直到本世纪50、60年代，SAR还处于经验的和半定性阶段。近十几年来，描述化合物分子结构特征的各种参数的建立和健全，生物活性定量客观指标的确定，使定量结构活性关系(QSAR)得以问世[2,3]，并应用于化学物质的毒性研究，形成了定量结构毒性关系(QSTR)理论。

自30年代末Schrader发现有机磷中性酯的杀虫性以来，全世界有机磷农药品种已超过150种，但其中很多化合物毒性评价尚不健全，利用QSTR方法对该类化合物进行全面的毒性预测评价具有重要意义。

1 材料与方法

1.1 生物活性资料来源

董华模编《化合物毒性及其环保手册》；[美]韦兰·丁·小海斯编著，陈炎磐、夏世钧主译《农药毒理学各论》；WHO 1991，1992《International Programme on Chemical Safety，IPCS》等。

1.2 结构参数选择

拓扑学信息理论指数[4]，价分子连接性指数[5]。

1.3 统计分析方法

定性预测分析，采用多元逐步判别分析方法；定量预测采用多元非线性回归分析方法。全部数据经SAS(6.04版)软件处理分析。

2 结果与分析

2.1 结构参数的数学描述

信息理论参数核心计算公式

$$\text{信息量 IC} = -\sum_{i=1}^{K}(P_i \cdot \log_2 P_i) \qquad (1)$$

式中，$P_i = N_i/N$，为分子中第 i 个亚结构中某元素随机选择概率；i 为某种结合；N_i 为第 i 个亚结构中的元素个数，N 为分子总原子数。

$$\text{结构信息量 SIC} = \text{IC}/\log_2 N \qquad (2)$$

$$\text{剩余信息量 CIC} = \sum_{i=1}^{K}(P_i \cdot \log_2 N_i) \qquad (3)$$

$$\text{比率信息量 PIC} = \sum_{i=1}^{K}\left(P_i \cdot \frac{\log_2 N_i}{\log_2 N}\right) \qquad (4)$$

2.2 参数的评价

(1) 从上述计算公式可以看出，当分子中第 i 个亚结构元素个数为1时，$\log_2 N_i = 0$，可造成分子亚结构信息的部分损失，为弥补这一缺

* 国家自然科学基金资助项目

收稿日期：1994-11-07

陷，本研究将上述式中 P_i 转换成 R_i，$R_i=W_i/W$，W 为分子量，W_i 为第 i 个亚结构各元素原子量之和，得到基团信息参数：

基团信息量 $$RIC=-\sum_{i=1}^{K}(R_i\cdot\log_2 R_i) \quad (5)$$

基团结构信息量 $$RSIC=RIC/\log_2 W \quad (6)$$

基团剩余信息量 $$RCIC=\sum_{i=1}^{K}(R_i\cdot\log_2 W_i) \quad (7)$$

基团比率信息量 $$RPIC=\sum_{i=1}^{K}\left(R_i\cdot\frac{\log_2 W_i}{\log_2 W}\right) \quad (8)$$

（2）信息理论指数对化合物分子的“加合性”及“构成性”等结构信息进行了定量描述，但在有机磷类化合物构效研究中，尚未充分反应杂原子(如 Cl、Br、F 等)对分子结构的影响，因此，本研究中结构数据矩阵引入了价分子连接性指数，对 R 取代基进行了定量描述。价分子连接性指数数学描述为：

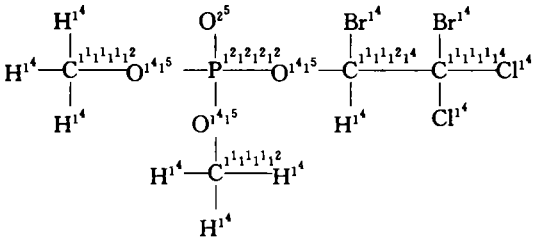
$$^{m}X_i^V=\sum_{j=1}^{n_m}\sum_{i=1}^{m+1}(\delta_i^V),\quad \delta^V=Z^V-h \quad (9)$$

式中，n_m 为某原子的序号；m 为子图边数；δ^V 为价 δ 值，是子图中 i 的点价；Z 为价电子数；h 为氢原子数。

2.3 参数的计算方法

（1）R 取代基价分子连接性指数计算方法详见文献[5]。

（2）分子信息理论指数计算，以二溴磷为例，描述其计算过程。二溴磷分子平面结构及元素价连接形式为：



等价原子组	N_i	W_i	P_i	R_i
1^4	11	7×1+2×35.45+2×79.9	11/20	237.3/380.8
$1^1 1^1 1^1 1^2$	2	2×12.01	2/20	24.02/380.8
$1^1 1^1 1^1 1^4$	1	1×12.01	1/20	12.01/380.8
$1^1 1^1 1^2 1^4$	1	1×12.01	1/20	12.01/380.8
$1^4 1^5$	3	3×16	3/20	48/380.8
2^5	1	1×16	1/20	16/380.8
$1^2 1^2 1^2 2^2$	1	1×30.97	1/20	30.97/380.8

则有：

$$IC=-\left(\frac{11}{20}\log_2\frac{237.3}{380.8}+\frac{2}{20}\log_2\frac{24.04}{380.8}+\cdots+\frac{1}{20}\log_2\frac{30.97}{380.8}\right)=1.953776$$

$$CIC=\frac{11}{20}\log_2 11+\frac{2}{20}\log_2 2+\cdots+\frac{1}{20}\log_2 1=1.867027$$

同理有：

PIC=0.4072063，RIC=1.853532

RCIC=6.719231，RPIC=0.7837883

SIC=0.4261269，RSIC=0.2162117

2.4 预测方法与数学模式

对 114 种有机磷农药依据 R_1、R_2 取代基不同划归为 3 大类，即二甲氧取代基类、二乙氧取代基类、非甲乙氧取代基类；根据已知的毒性实验结果，大鼠经口、经皮 LD_{50} 按 WHO 农药危害性建议标准，小鼠经口 LD_{50} 按我国 1978 年全国“工业企业设计卫生标准”科研协作会议建议的化学毒物毒性分级标准，将该类化合物划分为极毒、高毒、中等毒、低毒 4 级。

毒性等级的定性预测采用多元逐步判别分析方法。不同给药途径按化合物结构分类，以毒性等级为分组变量，以结构参数为自变量建立最优判别方程。9 类模型共 28 个判别式，判别回代符合率最低为 83.33%，平均回代符合率最低为 87.88%。

毒性 LD_{50} 定量预测采用多元非线性回归分析方法，不同给药途径分别建立毒性分级多元

定量预测模型和结构分类多元定量预测模型。共14个回归方程，相关指数 R^2 均在0.81以上，回归残差呈正态分布。

11种有机磷农药验证性预测，定性判别符合率为90.91%，1种化合物判定错误原因在于实测值 LD_{50} 贴近毒性等级划分标准。定量预测值平均相对误差为6.98%，预测值均在实验值95%可信区间内。预测效果满意。

3 讨论

在QSAR研究中，关键在于构造化合物理化结构参数及生物学效应数据矩阵，在此基础上，建立定量数学式，并实现其算法。QSAR研究所用参数众多，大体可分为3大类，即超热力学参数、量子化学参数和拓扑学参数。超热力学参数中诸如正辛醇/水分配系数 $\log P$、电子效应常数 σ_X、Taft立体参数 E_S 等都是一些经验参数，多依赖于实验或通过经验方程计算获得，常出现观测值与计算值的偏差，高分子化合物结构复杂，其理化性质影响因素较多，经验参数各自仅能反应结构信息的一些侧面，因此超热力学参数应用受到了限制。量子化学参数诸如分子总能 E_t、轨道能量 $E_{O(X)}$、原子电荷密度 Q^e 等，由于其计算上的复杂性，现在未能得到广泛应用。拓扑学参数是以拓扑图论为基础产生的，主要包括分子信息理论指数、分子连接性指数及Wiener途径数、Alternburg多项式指数和Hosoya的 Z 指数等。拓扑学参数就是以定量描述分子中原子的组成及其排列，弥补了其他参数不能对分子结构进行定量描述的缺陷。但：

(1) 结构信息理论指数和价分子连接性指数只能反映分子平面结构所包含的结构信息，而不能体现分子的立体空间构象，对于几何异构体不能区别。

(2) QSAR有许多不成熟的地方，特别是它忽略了化合物在生物体内的转运、代谢转化等过程，不能反映化合物与受体或酶的特异性作用及微观动态变化，从而影响了系统的预测精度。

(3) QSAR只能应用于作用机制相同化合物和结构相近同源化合物，而该类化合物 R 取代基变化复杂，目前尚没有一种合理分类标准来满足该类化合物分类毒性预测的需要。

QSTR研究逐渐成为解决毒性评价的重要工具。由于这门学科尚处于起步阶段，许多问题有待解决。诸如，为取得可靠的生理活性资料，需进一步改进生物鉴定法，需要积累生理活性物质的作用点与作用机制方面的药理、生理学知识，与此同时探讨新的参数，以便说明活性物质的作用，要进一步开展体内分布的隔室分布分析研究，以及寻求新的模型。所以QSTR研究只能作为毒理学的重要辅助手段，而不能在卫生学评价中完全代替毒理学试验，但它仍不失为一种极有前途的方法。

参考文献

1 Hansch C. On the Structure of Medical Chemistry. J. Med. Chem. 1976, **19**(1): 1

2 Tichy M et al.. QSAR analysis and data extrapolation among mammals in series of aliphatic alcohols. Environ. Health Perspect. 1985, **61**: 321

3 Speece R E. Environ. Science Technology. 1988, **22**(6): 606

4 Ray S K et al.. A QSARs study of N- Alkylnorketobemidones and Triazinones using SIC. Arzneim Forsch. /Drug Research. 1982, **32**(I), Nr. **4**: 322

5 李栋等译. 价分子连接性用于构效关系研究. 国外医学卫生学分册. 1979, **3**: 137

Abstracts

Chinese Journal of Environmental Science

Response Characteristics of Phenol on a Tyrosinase Biosensor. Hu Xiaoya et al. (Dept. of Chem. Teacher's School, Yangzhou Univ., Yangzhou 225002): *Chin. J. Environ. Sci.*, **16** (3), 1995, pp. 1—3

The response time decreases in the presence of catechol. The principal factors influencing the sensitivity were described. The ampeprometric biosensor of tyrosinase was prepared firstly by immobilizing tyrosinase onto preactivated polyamide support which contacted tightly with wax-impregnated graphite electrode. Amperometric measurements of phenol in water were carried out by applying a potential of -0.200 $V_{vs}SCE$ in a 0.1 mol/L phosphate buffer of pH6.50 and 5×10^{-7} mol/L catechol at 25℃. The preparation, storage and replacement of immobilized enzyme are convenient. No activity of immobilized enzyme was lost during five months of storage, with linear range of $2\times10^{-7}-1.25\times10^{-5}$ mol/L of phenol.

Key words: electrode, biosensor, enzyme, tyrosinase, phenol.

Effect of Magnetic Field on Dehydrogenase Activity of Purple Nonsulfur Photosynthetic Bacteria. Ma Haizhen et al. (Taiyuan Univ. of Technology, Taiyuan 030024): *Chin. J. Environ. Sci.*, **16**(3), 1995, pp. 4—7

Some properties of native and immobilized cells of purple nonsulfur photosynthetic bacteria (PSB) were studied and compared under the condition of magnetic field. The results show that under the condition of optimum magnetic field both types of cells were thermally and pH stable, with an optimum dehydrogenase temperature of 30—40℃, and an optimum pH of 8. As compared to the control, the dehydrogenase activity was 10%—20% hig-her for native cells and was 20%—30% higher for immobilized cells. The magnetic field augmented the sensitivity of PSB to metal inos (Fe^{2+}, Mn^{2+}, Mg^{2+}). But these two types of cells were different in the value of optimum magnetic field and the accumulative effect of magnetic field. The effect of magnetic field on the dehydrogenase activity of immobilized cells remained constant after 45—210 days of storage at 4℃, with a dehydrogenase activity improved by 20%. Immobilized cells had a dehydrogenase activity of 150%—200% higher than native cells under the same conditions but without an applied mgnetic field.

Key words: magnetic field, magnetic biological effect, purple nonsulfur photosynthetic bacteria, immobilized cells, dehydrogenase.

Study on Two-phase Anaerobic Digestion Process for Treating High Strength Organic Wastewater Containing a High Level of Sulphate. Yang Jingliang et al. (Dept. of Environ. Eng., Tsinghua Univ., Beijing 100084): *Chin. J. Environ. Sci.*, **16**(3), 1995, pp. 8—11

A two-phase anaerobic digestion process has been developed for treating high strength organic wastewater containing a high level of sulphate. The system comprises: ① anaerobic filter packed with Rasching rings which was used as an acidogenic reactor with a volumetric loading of 5 kg $SO_4^{2-}/(m^3\cdot d)$ and a SO_4^{2-} removal efficiency of 80%; ② a desulphiding reactor with a sulphide removal rate of more than 90%. of which more than 95% was converted to S^0; ③ a methanogenic reactor with a COD volumetric loading rate of 15.8 kg COD/($m^3\cdot d$) and a COD removal efficiency of 83.3%. The whole system has an overall COD removal efficiency of 87.6% and a SO_4^{2-} removal efficiency of 99.4%—100%.

Key words: organic wastewater containing sulphate, sulphate reducing bacteria, two-phase anaerobic digestion process.

Information Theoretic Indices and Application to Predicting the Toxicities of Organophosphorus Pesticides. Peng Ji et al. (Institute of Environmental Medicine, Tongji Medical Univ., Wuhan 430030): *Chin. J. Environ. Sci.*, **16**(3), 1995, pp. 12—14

The information theoretic indices and the valence molecular connectivity indices as the characteristics of molecular structure were applied to predicting the toxicities of organophosphorus pesticides. Radical information parameter was suggested. According to the data on acute toxicities of 114 organophosphorus pesticides and different channels that the pesticides were given, the qualitative and quantitative prediction models were estabi-

lished along with a discriminatory analysis and a multiple non-linear regression analysis. Rates of qualitative discriminatory analysis were not less than 83.33%. The correlation indices of quantitative regression models (R^2) were over 0.81. Residuals showed the normal distribution. Predicted values LD_{50} by the regression equations were within a 95% confidence interval of observed values.

Key words: information theoretic indices, valence molecular connectivity indices, structure-toxicity relationship, organophosphorus pesticides, prediction of toxicity.

Design and Implementation of the Jiangsu Province's Environmental Information System. Cheng Shengtong et al. (Dept. of Environ. Eng., Tsinghua Univ., Beijing 100084): *Chin. J. Environ. Sci.*, **16**(3), 1995, pp. 15—18

The objects of Jiangsu Province's Environmental Information System (JSEIS) project and the implementation of the first phase of the project were discussed, the development procedure, hardware, software and network environment were also presented. The process of JSEIS development includes the data normalization, the design of unified data collection report system, the establishment of a basic database for data sharing and a applied database for each of 13 environmental management subsystems developed, the data exchange system among different database products and the data communication at both provincial and municipal levels. The JSEIS allows the data to have integrity, consistency and good practicality.

Key words: environmental management, environmental information system, database management systems.

Remote Sensing Investigation on the Ground Thermal Field and Its Influence on Air Pollution in an Industrial and Mining City. Sheng Yihua et al. (Dept. of Mining, China Univ. of Mining and Technology, Xuzhou 221008): *Chin. J. Environ. Sci.*, **16**(3), 1995, pp. 19—22

In Jincheng City, Shanxi Province, the two temporal thermal infrared images were taken by means of airborne thermal infrared scanner in early morning (from 5 : 30 to 6 : 30) and midday (from 13 : 30 to 14 : 30) in winter. With reference to simultaneous ground radiation temperature data, some of digital image processing methods, such as image quantization, geometric correction, and thermal field classification, were used for discovering the ground thermal field intensity in the city. Based on these, it was concluded that there is a very obvious ground thermal island effect in early morning and a vague one in midday, and the ground radiation temperature differences for both of the two tempral phases are about 10℃. In addition, the linear relationship between the low altitude air temperature and the ground radiation temperature were computed, and a bad influence of the ground thermal filed on air pollution was found.

Key words: thermal infrared remote sensing, ground thermal field, air pollution.

Identification of Oils Spilt at Sea Surface Based on a Fuzzy Similarity Preference Ratio. Xu Hengzhen et al. (Institute of Marine Environ. Protection, State Oceans Administration, Dalian 116023): *Chin. J. Environ. Sci.*, **16**(3), 1995, pp. 23—27

A model based on a fuzzy similarity preference ratio has been established to identify different kinds of oils spilt at sea surface, by introducing a concept of the fuzzy area of confidence level $\bar{\lambda}$ of a fuzzy similarity preference ratio. The identification of oils spilt at sea surface was found being largely influenced by kinds of oils, experimental errors, and weathering of oils. A Renqiu crude oil after a 30 days weathering was found to have an identifiable confidence level of fuzzy similarity preference ratio in the range of 0.005819 to 0.04641 making it different from other 13 kinds of oils. A Daqing crude oil after a 28 days weathering at sea surface was found to have such an identifiable confidence level in the range of 0.1101 to 0.2510 inclusive making it different from other 5 kinds of oils. The results from the actual identification of oils spilt at sea surface show that the present method was more precise than a fingerprint spectrograms-based identification method.

Key words: fuzzy, similarity preference ratio, identification, oil spill.

Production of Sintered Ceramsite from Coal Gangue, Fly Ash and Waste Gypsum. Sheng Zhaoqi et al. (Dept. of Environ. Eng., East China Univ. of Technology, Shanghai 200237): *Chin. J. Environ. Sci.*, **16**(3), 1995, pp. 28—31

A lightweight construction aggregate was suc-