

# 酚类化合物的三维-定量结构与生物降解性关系 (3D-QSBR)\*

罗 坤 高士祥\* 张 爱 茜 王 连 生

(南京大学环境学院, 污染控制与资源化研究国家重点实验室, 南京, 210093)

**摘 要** 利用比较分子场分析法 (CoMFA) 研究了 32 种酚类化合物的生物降解性与其结构间的三维定量关系, 并利用分子场聚焦 (Region Focus) 和调整网格大小对模型进行改善, 得到具有较强预测能力的 3D-QSBR 模型. 结果表明: 进行分子场聚焦和减小网格步长 (Grid Spacing) 均可改善模型质量, 得到的最佳模型主成分数为 4, 交叉验证相关系数  $Q^2$  为 0.587, 复相关系数  $R^2$  为 0.917,  $F$  值为 57.654.

**关键词** 酚类化合物, 生物降解性, 比较分子场分析, QSBR.

生物降解是有机污染物从环境中去除的主要途径, 通过建立 QSBR 模型, 根据化合物的结构可推测其生物降解性. 目前, QSBR 的研究方法主要有线性自由能相关方法<sup>[1]</sup>、基团贡献法<sup>[2]</sup>、分子连接性指数法<sup>[3]</sup>、人工神经网络法<sup>[4]</sup>、专家系统<sup>[5]</sup>和机制方法<sup>[6]</sup>等, 它们都是二维的方法. 比较分子场分析 (CoMFA) 是一种常用的三维 QSAR 方法<sup>[7]</sup>. Dearden 与 Stott<sup>[8]</sup> 曾用 CoMFA 对生物降解性进行了预测, 结果发现, 对于醇、羧酸和直链烷基苯磺酸的 BOD 值具有良好的预测能力; 对于酯和苯磺酸, 交叉验证的相关系数较差; 而对于酚类则未得出明显的相关性. 该法是一种有潜力的研究方法.

本文利用比较分子场分析的方法研究了 32 种酚类化合物的生物降解性, 建立了相应的 QSBR 模型, 并讨论了影响酚类化合物生物降解的因素.

## 1 实验部分

### 1.1 酚类化合物生物降解性评价

32 种酚类化合物的生物降解数据引自文献 [9], 以降解 12d 产生的二氧化碳量 (PCD) 来表征化合物的生物降解性, 单位为  $\text{mmol} \cdot \text{l}^{-1}$ . 根据实验结果对化合物进行分类<sup>[10]</sup>:  $\text{PCD} > 8 \text{ mmol} \cdot \text{l}^{-1}$ , 易生物降解;  $0 < \text{PCD} \leq 8 \text{ mmol} \cdot \text{l}^{-1}$ , 可生物降解;  $\text{PCD} \leq 0 \text{ mmol} \cdot \text{l}^{-1}$ , 难生物降解. 按结构选取 6 种酚类化合物作为测试组, 其余 26 种作为训练组, 见表 1.

### 1.2 分子构建及结构优化

采用 Tripos 公司 Sybyl 7.3 软件包中的 Sketch Molecule 构建了 32 种酚类化合物的三维结构. 用 Tripos 分子力场进行优化, 分子中的原子净电荷采用 Gasteiger-Huckel 法计算, 能量收敛标准为  $0.05 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-1}$ , 得到化合物的低能构象.

本文选择降解性最大的苯酚为模板分子, 以苯环上的碳和 1 号位上取代的羟基为重叠原子.

### 1.3 比较分子场分析

以带一个单位正电荷的  $\text{sp}^3$  碳原子为探针原子, 在计算机产生的矩形盒子 (Region) 内, 每隔  $0.2 \text{ nm}$  设定一个网格点, 在每个网格点上分别利用 Lennard-Jones 6-12 公式和 Coulomb 公式计算探针原子与各个分子之间的立体场能和静电场能. 立体场能和静电场能的阈值为  $30 \text{ kcal} \cdot \text{mol}^{-1}$ .

对酚类化合物的生物降解性数据 (PCD) 和 CoMFA 值进行偏最小二乘法 (PLS) 分析. 以  $\delta(2.0 \text{ kcal} \cdot \text{mol}^{-1})$  作为列过滤, 去除场能变化小于  $2.0 \text{ kcal} \cdot \text{mol}^{-1}$  的网格点, 从而加速运算和减少噪音. 通过逐一剔除 (LOO) 交叉验证得出最佳主成分数 (NOC) 和预测相关系数 ( $Q^2$ ), 再进行非交叉验证, 得到最佳的拟合模型.

表 1 32种酚类化合物生物降解产生的 CO<sub>2</sub>量 (PCD)及定性评价Table 1 Production of CO<sub>2</sub> and qualitative evaluation for biodegradation of 32 phenols

序号	化合物	PCD	定性评价	序号	化合物	PCD	定性评价
1	苯酚	18.32	易 <sup>1)</sup>	17	2,3-二氯酚	2.00	可
2	邻苯二酚	16.30	易	18	2,4-二氯酚	1.69	可
3	间苯二酚*	17.94	易	19	2,5-二氯酚	2.06	可
4	对苯二酚	11.47	易	20	3,4-二氯酚	1.04	可
5	间苯三酚	15.32	易	21	3,5-二氯酚*	8.92	易
6	邻氨基酚	9.50	易	22	3,5-二羟基甲苯	13.58	易
7	间氨基酚	10.44	易	23	2,4-二甲基酚*	3.01	可
8	对氨基酚	9.42	易	24	2,6-二甲基酚*	4.67	可
9	邻甲酚	8.32	易	25	3,5-二甲基酚	14.22	易
10	间甲酚	8.76	易	26	2,4-二硝基酚	-2.12	难 <sup>3)</sup>
11	对甲酚	9.00	易	27	2,6-二硝基酚	-2.40	难
12	邻氯酚	5.36	可 <sup>2)</sup>	28	3-甲基-5-乙基酚	-0.23	难
13	对氯酚	7.27	可	29	5-亚硝基邻甲酚	2.94	可
14	邻硝基酚*	12.72	易	30	4-亚硝基间苯二酚	8.94	易
15	对硝基酚	11.02	易	31	2,6-二叔丁基对甲酚	-5.07	难
16	间乙酰氨基酚*	0.73	可	32	2,4,6-三氯苯酚	-2.78	难

注: \* 测试组化合物; 1) 表示易生物降解, 2) 表示可生物降解, 3) 表示难生物降解.

## 2 结果与讨论

对 32种酚类化合物的生物降解性进行比较分子场分析的结果(模型 1)见表 2. 为研究分子场聚焦和网格步长对模型的影响, 对模型 1进行分子场聚焦后分析得模型 2. 将网格步长从 0.2nm 变为 0.1nm 后分析得模型 3. 再对模型 3进行分子场聚焦后分析得模型 4, 模型结果见表 3.

在 QSBR 分析过程中,  $R^2$ 和  $Q^2$ 是检验模型质量的重要标准. 一般认为  $R^2 \geq 0.9$ 表明所建立的模型具有良好的相关性. 本研究所得四个模型除了第一个模型的  $R^2$ 略低于 0.9 外, 其它三个模型的  $R^2$ 均大于 0.9. 表明所建模型具有良好的相关性. 交叉验证相关系数  $Q^2$ 可检验模型的预测能力.  $Cramer^{[10]}$ 等认为,  $Q^2$ 大于 0.5 表明模型具有较强的预测能力. 本研究所得四个模型的  $Q^2$ 均大于 0.5. 表明所得模型具有良好的预测能力.

表 2 QSBR 模型对酚类化合物生物降解性的预测值与实验值比较

Table 2 Comparison between observed biodegradability values and calculated values of phenols by QSBR models

数据集	化合物	实验值	模型预测值及残差							
			模型 1	残差	模型 2	残差	模型 3	残差	模型 4	残差
训练组	苯酚	18.32	14.04	-4.28	14.29	-4.03	14.04	-4.28	14.80	-3.52
	邻苯二酚	16.30	16.91	0.61	16.11	-0.19	17.22	0.92	16.61	0.31
	对苯二酚	11.47	10.52	-0.95	9.69	-1.78	10.73	-0.74	11.24	-0.23
	间苯三酚	15.32	14.36	-0.96	15.51	0.19	17.19	1.87	16.07	0.75
	邻氨基酚	9.50	11.19	1.69	10.89	1.39	9.45	-0.05	9.54	0.04
	间氨基酚	10.44	13.08	2.64	12.30	1.86	12.36	1.92	12.21	1.77
	对氨基酚	9.42	9.96	0.54	11.16	1.74	10.76	1.34	10.88	1.46
	邻甲酚	8.32	7.87 <sup>*</sup>	-0.45	6.75 <sup>*</sup>	-1.57	7.41 <sup>*</sup>	-0.91	7.51 <sup>*</sup>	-0.81
	间甲酚	8.76	11.32	2.56	10.50	1.74	10.03	1.27	10.08	1.32
	对甲酚	9.00	9.63	0.63	10.66	1.66	9.64	0.64	10.33	1.33
	邻氯酚	5.36	5.82	0.46	5.27	-0.09	6.30	0.94	5.82	0.46

续表 2

数据集	化合物	实验值	模型预测值及残差							
			模型 1	残差	模型 2	残差	模型 3	残差	模型 4	残差
训练组	对氯酚	7.27	8.55 <sup>e</sup>	1.28	8.87 <sup>f</sup>	1.60	7.88	0.61	8.30 <sup>e</sup>	1.03
	对硝基酚	11.02	6.35 <sup>e</sup>	-4.67	7.41 <sup>f</sup>	-3.61	8.32	-2.70	9.61	-1.41
	2,3-二氯酚	2.00	2.97	0.97	2.15	0.15	2.09	0.09	1.27	-0.73
	2,4-二氯酚	1.69	0.25	-1.44	-0.25 <sup>e</sup>	-1.94	0.18	-1.51	-0.66 <sup>e</sup>	-2.35
	2,5-二氯酚	2.06	4.43	2.37	3.30	1.24	4.54	2.48	3.58	1.52
	3,4-二氯酚	1.04	5.54	4.50	5.63	4.59	3.48	2.44	3.44	2.40
	3,5-二羟基甲苯	13.58	13.92	0.34	14.63	1.05	13.87	0.29	13.52	-0.06
	3,5-二甲苯酚	14.22	9.18	-5.04	9.88	-4.34	9.41	-4.81	9.94	-4.28
	2,4-二硝基酚	-2.12	-3.63	-1.51	-3.42	-1.30	-1.94	0.18	-2.41	-0.29
	2,6-二硝基酚	-2.40	-1.18	1.22	-1.09	1.31	-1.30	1.10	-1.87	0.53
	3-甲基-5-乙基酚	-0.23	-0.38	-0.15	0.54 <sup>f</sup>	0.77	1.01 <sup>e</sup>	1.24	1.14 <sup>e</sup>	1.37
	5-亚硝基邻甲酚	2.94	3.06	0.12	1.65	-1.29	0.38	-2.56	0.35	-2.59
	4-亚硝基间苯二酚	8.94	8.63	-0.31	9.24	0.30	9.32	0.38	10.19	1.25
	2,6-二叔丁基对甲酚	-5.07	-5.39	-0.32	-4.76	0.31	-5.48	-0.41	-4.07	1.00
2,4,6-三氯苯酚	-2.78	-2.61	0.17	-2.53	0.25	-2.50	0.28	-3.05	-0.27	
测试组	间苯二酚	17.94	12.35	-5.59	11.75	-6.19	10.90	-7.04	11.20	-6.74
	邻硝基酚	12.72	4.02 <sup>e</sup>	-8.70	3.40 <sup>f</sup>	-9.32	3.49 <sup>e</sup>	-9.23	2.44 <sup>e</sup>	-10.28
	间乙酰氨基酚	0.73	11.11 <sup>e</sup>	10.38	10.53 <sup>f</sup>	9.80	9.53 <sup>e</sup>	8.80	9.08 <sup>e</sup>	8.35
	3,5-二氯酚	8.92	9.73	0.81	9.10	0.18	8.05	-0.87	7.84 <sup>e</sup>	-1.08
	2,4-二甲苯酚	3.01	3.70	0.69	3.54	0.53	3.15	0.14	3.27	0.26
	2,6-二甲苯酚	4.67	4.60	-0.07	4.37	-0.30	6.18	1.51	6.20	1.53

注: \* 表示模型预测定性评价与实际不符合。

从表 3 可以看出, 分子场聚焦可以提高模型的相关性和预测能力。 $Q^2$ 和  $R^2$ 从 0.525 和 0.886 增加到 0.632 和 0.901;  $F$  值从 401631 增加到 471618,  $SE_{L00}$ 和  $s$  则从 41865 和 21389 降低到 41283 和 212251。然而训练组和测试组中定性评价预测发生错误的个数却增加了, 这可能是因为分子场聚焦在降低噪音的同时也丢失了部分信息。此外, 网格步长对模型结果也会产生影响。将网格步长从 0.12nm 减小到 0.11nm 后, 模型的相关性和预测能力都有一定的提高。 $Q^2$ 和  $R^2$ 从 0.525 和 0.886 增加到 0.587 和 0.917;  $F$  值从 401631 增加到 571654,  $SE_{L00}$ 和  $s$  则从 41865 和 21389 降低到 41539 和 210401。从模型的统计参数看, 模型 4 最优, 但对化合物定性评价结果较差。而评价模型的一个重要标准是对测试组化合物的预测能力。模型 3 对测试组的预测偏差比模型 4 小。因此, 综合模型的拟合能力与预测能力, 认为模型 3 为最优模型, 其对训练组和测试组的预测正确率分别为 92.13% 和 66.17%, 且均无将不可降解预测为易降解或将易降解预测为不可降解这样的显著预测误差。

表 3 模型结果与文献结果的比较

Table 3 Summary of the model results and comparison with literatures

模型	$Q^2$	NOC	$SE_{L00}$	$R^2$	$s$	$F$	$S$	$E$	ER <sub>1</sub> /%	ER <sub>2</sub> /%
模型 1	0.525	4	41865	0.886	21389	401631	0.1614	0.1386	111.5	331.3
模型 2	0.632	4	41283	0.901	21225	471618	0.1609	0.1391	191.2	331.3
模型 3	0.587	4	41539	0.917	21040	571654	0.1485	0.1515	71.7	331.3
模型 4	0.655	4	41149	0.933	11823	731560	0.1523	0.1477	151.4	501.0
文献 [11]	)	3	)	0.931 <sup>e</sup>	2152	50132	)	)	191.2	161.7
文献 [12]	)	3	)	0.718	)	231794	)	)	71.7	501.0

注: \* 表示的为相关系数  $R$ ;  $Q^2$ 为交叉验证相关系数; NOC为最佳主成分数;  $SE_{L00}$ 为逐一剔除交叉验证预测标准误差;  $R^2$ 为复相关系数;  $s$ 为预测标准差;  $F$ 为  $F$  统计参数;  $S$ 为立体场贡献;  $E$ 为静电场贡献; ER<sub>1</sub>为训练组中定性评价预测误差率; ER<sub>2</sub>为测试组中定性评价预测误差率。

Dearden<sup>[8]</sup>等用 CoMFA 方法对酚类化合物的生物降解性进行研究, 未发现明显的交叉验证相关性, 而本研究发现酚类化合物的生物降解性与其结构之间的相关性是存在的。陈勇生等<sup>[11]</sup>曾用逐步回归方法对二氧化碳产生量和分子连接性指数建立线性回归方程:

$$\text{PCD} = 53118 - 41537^3 v_p - 20173^3 v_p + 13178^4 v_{PC}$$

运用此方程对训练组和测试组进行预测得到的定性评价正确率分别为 8018% 和 8313%; 何菲等<sup>[12]</sup>通过筛选描述符建立了二氧化碳产生量与生成热  $v H_f$ 、分子摩尔质量  $M_w$  以及偶极矩  $L$  之间的相关方程:

$$\text{PCD} = 241239 - 01115v H_f - 01179M_w + 11399L$$

对 32 种化合物预测与实际不符的共计有 5 个, 总体预测正确率为 8414%。因此, 本研究所建模型的拟合能力与文献模型相当, 但预测能力偏差。这可能是因为测试组中的间乙酰氨基酚含有训练组中未曾出现的乙酰氨基基团, 因此, 导致预测发生较大的偏差。

CoMFA 法优于二维 QSAR 的特征之一是能够得到立体场和静电场的等高图 (图 1, 图中分子为间乙酰氨基酚), 能直观反映立体场和静电场对分子活性的影响。立体场中的 Favored 和 Disfavored 区域分别用绿色和黄色 (图中未显示) 表示, 绿色区域内取代基体积愈大活性愈大, 而黄色区域内增大取代基降低活性; 静电场中的 Favored 和 Disfavored 区域分别用蓝色和红色表示, 蓝色区域内取代基正电性愈强活性愈大, 而红色区域内负电性愈强活性愈大。由图 1 可以看出, 苯环上方为红色区域, 而苯酚的羟基上的孤对电子与苯环上的大  $\pi$  键形成共轭作用使苯环上的电子云密度增大, 负电性增强, 因此, 生物降解性增大。苯酚的间位附近为绿色区域, 3, 5-二甲基酚比间甲酚多了一个间位甲基, 体积增大, 生物降解性提高。测试组中间乙酰氨基酚的乙酰氨基位于绿色区域, 乙酰氨基中的氢原子靠近静电场中的蓝色区域, 这都会导致生物降解性预测值增大。但训练组中苯酚间位未出现和乙酰氨基类似大小的基团, 乙酰氨基与立体场发生冲撞, 不适合用本模型进行预测, 从而预测偏差较大。因此, 在利用 CoMFA 进行 QSBR 研究时必须考虑基团的体积因素, 使训练组尽可能具有代表性。

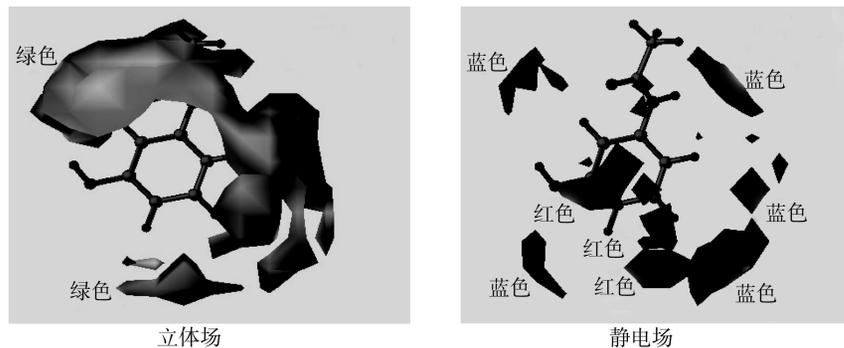


图 1 CoMFA 模型 3 的三维等高图

Fig 1 3D<sub>2</sub>contour maps of CoMFA model 3 for steric field and electrostatic field

### 3 结论

运用 CoMFA 方法对 32 种酚类化合物的生物降解性与其结构进行相关分析得到了较好的模型; 进行分子场聚焦和减小网格步长可以改善模型的质量; 得到的最优模型对训练组和测试组的正确评价率分别达 9213% 和 6617%。比较分子场分析是有机物生物降解性预测的一种潜在的三维方法, 有待于进一步研究。

#### 参 考 文 献

- [1] Vaishnav D D, Boethling R S, Babeu L et al, Quantitative Structure-Biodegradability Relationships for Alcohols, Ketones and Aliphatic Compounds [J] *J Chemosphere*, 1987, **16** (44) B 695-703
- [2] Desai S M, Govind R, Tabak H H et al, Development of Quantitative Structure-Activity Relationships for Predicting Biodegradation Kinetics [J] *Environmental Toxicology and Chemistry*, 1990, **9B** 473-477
- [3] Boethling R S Application of Molecular Topology to Quantitative Structure-Biodegradability Relationships [J] *Environmental Toxicology*

- and Chemistry*, 1986, **5B** 797) 806
- [ 4 ] Ziko V, Prediction of Biodegradability of Organic Chemicals by an Artificial Neural Network [ J ] *Chemosphere*, 1991, **23** ( 3 ) B 305) 312
- [ 5 ] Klopman G, Tu M, Structure-Biodegradability Study and Computer-Automated Prediction of Aerobic Biodegradation of Chemicals [ J ]. *Environmental Toxicology and Chemistry*, 1997, **16** ( 99 ) B 1829) 1835
- [ 6 ] Peijnenburg W J Damborsky J, Biodegradability Prediction. Huwer Academic Publishers, 1996, 75) 92
- [ 7 ] Cramer R D, Patterson D E, Bunce J D, Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins [ J ] *Journal of the American Chemical Society*, 1988, **110** ( 18 ) B 5959) 5967
- [ 8 ] Dearden J C, Stott I P, CoMFA Analysis of Biodegradability [ J ] *SAR and QSAR in Environmental Research*, 1995, **4** ( 4 ) B 189) 196
- [ 9 ] 陈勇生, 陈丽侠, 杨杰等, 32种芳香化合物的好氧生物降解性表征 [ J ] *环境化学*, 1997, **16** ( 1 ) B 43) 48
- [ 10 ] Cramer R D, Bunce J D, Patterson D E, Crossvalidation, Bootstrapping and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies [ J ] *Quantitative Structure-Activity Relationships*, 1988, **7B** 18) 25
- [ 11 ] 陈勇生, 陈丽侠, 杨杰等, 拓扑指数在定量结构生物降解性关系中的应用 [ J ] *环境化学*, 1997, **16** ( 3 ) B 208) 214
- [ 12 ] 何菲, 袁星, 程香菊等, 酚类化合物好氧生物降解的 QSBR 研究 [ J ] *中国环境科学*, 2001, **21** ( 2 ) B 152) 155

## THREE DIMENSIONAL QUANTITATIVE STRUCTURE-BIODEGRADABILITY RELATIONSHIP (3D QSBR) ON PHENOLS

LUO Kun    GAO Shi-xiang    ZHANG Ai-qian    WANG Lian-zheng

( State Key Laboratory of Pollution Control and Resource Reuse, School of the Environment, Nanjing University, Nanjing 210093, China )

### ABSTRACT

The quantitative relationships between the biodegradability of 32 phenols and their structure were studied by using the Comparative Molecular Field Analysis (CoMFA) method. Refinements of the model were made by using Region Focus and adjustment of the Grid Spacing. A 3D QSBR model with fine predictability was obtained comparing with literature. As indicated by the results, both Region Focus and decrease of the Grid Spacing can improve the model quality. The number of components  $Q^2$ ,  $R^2$  and  $F$  of the best model are 4.01587, 0.01917 and 571.654 respectively. Factors affecting chemicals biodegradability were discussed.

**Keywords** phenols, biodegradability, CoMFA, QSBR