

多溴联苯醚生物富集系数的定量结构-活性关系

李吉安¹, 薛秀玲¹, 卢桂宁²

(1. 华侨大学 化工学院, 福建 厦门 361021;

2. 华南理工大学 环境与能源学院, 广东 广州 510006)

摘要: 在 B3LYP/6-31G* 水平上,运用 Gaussian 03 对多溴联苯醚(polybrominated diphenyl ethers,PBDEs)分子进行结构优化.采用偏最小二乘法(partial least squares,PLS)对 PBDEs 的生物富集系数(bioconcentration factor,BCF)与分子结构参数进行关联.结果表明:BCF 结合苯环上溴原子的取代个数(N_{Br})可以较好地表达 PBDEs 的理化性质与其分子结构间的定量关系; $\log(\text{BCF}-N_{Br})$ 优化后模型的相关系数 $R^2_{Y(\text{cum})}$ 为 0.980,累积交叉有效性判别系数 Q^2_{cum} 为 0.921,该模型具有较高的拟合精度和良好预测能力;分子总能量 E_T 和电子运动空间广度 R_e 对 PBDEs 的 BCF 有着重要制约作用;具有较低 E_T 、较大 R_e 的 PBDEs 呈现出较低的 BCF.

关键词: 多溴联苯醚; 定量结构-活性关系; 生物富集系数; 偏最小二乘法

中图分类号: O 621.2

文献标志码: A

多溴联苯醚(PBDEs)是溴代阻燃剂类化合物,共有 209 种同系物,常作为阻燃添加剂加入到树脂、聚苯乙烯和聚氨酯泡沫等高分子合成材料中,广泛应用于塑料制品、纺织品、电路板和建筑材料等领域^[1-3].PBDEs 相对分子质量大、熔点高、蒸汽压低、水溶性低及辛醇-水分配系数高,因而具有亲脂性和生物易累积等特点^[4],能在生物体内的脂肪和蛋白质中蓄积,并通过食物链放大,对高营养级的生物造成影响.PBDEs 作为一种新型的全球性、持久性有机污染物,对环境生物以及人体健康的造成危害,并引起了广泛的关注.生物富集因子(BCF)是评价有机污染物生物累积性的重要指标^[5].在 PBDEs 的环境评估中,生物富集因子是一个重要的参数.虽然对于 PBDEs 的生物富集因子的测定有标准的方法,但要测定 209 种 PBDEs 的 BCF 值,实验显然耗时、耗力,测定费用高.因此,目前由实验测定 PBDEs 的 BCF 极其有限.定量结构-活性关系(QSAR)指化合物的分子结构与其活性之间的定量预测模型^[6-8].QSAR 可以根据化合物的结构参数来预测其活性,弥补数据的缺失,降低昂贵的测试费用.本文拟通过 Gaussian 03^[9]将 PBDEs 进行 B3LYP/6-31G* 水平上的分子结构优化,从中提取相关量子化学参数.

1 材料与方法

1.1 BCF 文献值的选取

多溴联苯醚(PBDEs)通用结构式,如图 1 所示,图 1 中: $m+n$ 的范围是 1~10.本研究所用 19 种 PBDEs 的 BCF 实测值取自 Mansouri 等^[10]的工作,其数据从寡毛纲淡水生物正颤蚓获得,生物富集系数,如表 1 所示.表 1 中:随机选取的 13 种 PBDEs 组成模型训练组,其他 6 种 PBDEs 则作为模型检验组(见 * 标注).

1.2 量子化学参数选取

应用量子化学软件包 Gaussian 03,以密度泛函理论的方法计算,选择考察包括 22 个量子化学参数在内的 24 种变量,所涉及的量子化学参数包括:最高占有轨道能量(E_{HOMO})、最低空轨道能量(E_{LUMO})、

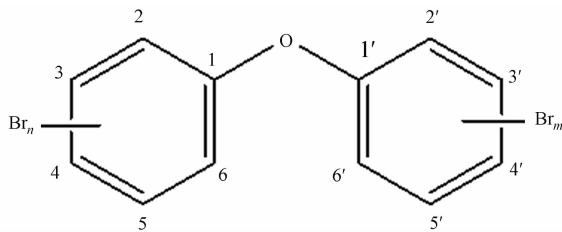


图 1 PBDEs 的化学结构图

Fig. 1 Chemical structure of PBDEs

收稿日期: 2013-10-18

通信作者: 薛秀玲(1975-),女,副教授,主要从事环境化学的研究. E-mail:xiulingxue@163.com.

基金项目: 国家自然科学基金资助项目(21207043, 21007017)

次高占有轨道能量(E_{NHOMO})、次低空轨道能量(E_{NLUMO})、分子总能量(E_{T})、电子空间广度(R_{e})、分子偶极矩(μ)、1,1'号碳原子与氧原子间的键长($R_{\text{C1-O}}$ 、 $R_{\text{O-C1'}}$)、分子骨架中 13 个原子的带电量(Q_{C1} , Q_{C2} , Q_{C3} , Q_{C4} , Q_{C5} , Q_{C6} , Q_{O} , $Q_{\text{C1'}}$, $Q_{\text{C2'}}$, $Q_{\text{C3'}}$, $Q_{\text{C4'}}$, $Q_{\text{C5'}}$, $Q_{\text{C6'}}$). 此外还考察了 $E_{\text{LUMO}}-E_{\text{NHOMO}}$, $E_{\text{LUMO}}+E_{\text{NHOMO}}$ 两种前线轨道能量的组合.

表 1 PBDEs 的生物富集系数
Tab.1 Log BCF of the PBDEs

编号	化合物	取代基位置	log BCF
1	BDE-28	2,4,4'-TriBDE	5.83
2*	BDE-47	2,2',4,4'-TeBDE	5.84
3	BDE-51	2,2',4,6'-TeBDE	5.88
4*	BDE-66	2,3',4,4'-TeBDE	6.05
5	BDE-77	3,3',4,4'- TeBDE	6.29
6	BDE-99	2,2',4,4',5-PeBDE	6.01
7*	BDE-100	2,2'4,4',6-PeBDE	6.36
8	BDE-119	2,3',4,4',6-PeBDE	6.29
9	BDE-126	3,3',4,4',5-PeBDE	6.43
10*	BDE-153	2,2',4,4',5,5'-HxBDE	6.13
11	BDE-154	2,2'4,4',5,6'-HxBDE	6.15
12	BDE-197	2,2',3,3',4,4',6,6'-OcBDE	5.30
13*	BDE-198	2,2',3,3',4,5,5',6-OcBDE	5.68
14	BDE-203	2,2',3,4,4',5,5',6-OcBDE	5.93
15	BDE-204	2,2',3,4,4',5,6,6'-OcBDE	5.75
16*	BDE-206	2,2',3,3',4,4',5,5',6-NoBDE	5.44
17	BDE-207	2,2',3,3',4,4',5,6,6'-NoBDE	5.75
18	BDE-208	2,2',3,3',4,4',5,6,6'-NoBDE	5.57
19	BDE-209	2,2',3,3',4,4',5,5',6,6'-DeBDE	5.19

1.3 模型的优化建立

应用 Simca-P(11.5 版)软件进行偏最小二乘法(PLS)分析. 模型的检验采用外部样本预测检验法,即将样本分为模型训练组和模型检验组. 表 1 用 Simca-P 软件建模,当某个 PLS 主成分交叉有效性判别系数的 Q^2 值大于程序缺省设置的显著性水平限(0.097 5)时,即认为该主成分有意义. 当累积交叉有效性判别系数 Q^2_{cum} 大于 0.5 时,认为所建立的模型有较好的预测可靠性.

考虑到 PBDE 溴取代的个数(N_{Br})对 PBDEs 的 BCF 可能有影响,为了获得统计学上合理的模型,将 log BCF 与 N_{Br} 进行组合,log BCF,log($\text{BCF}+N_{\text{Br}}$),log($\text{BCF}-N_{\text{Br}}$),log($\text{BCF}\times N_{\text{Br}}$)和 log(BCF/N_{Br})作为因变量,采用 PLS 方法拟合建立模型.

2 结果与分析

2.1 模型的建立

模型训练组不同因变量 PLS 模型最终拟合参数,如表 2 所示. 由表 2 可知:以 log BCF 为因变量建模的 Q^2_{cum} 小于 0.5,说明该模型预测性较差,无法建立可信用度较高的定量预测模型;以 log($\text{BCF}-N_{\text{Br}}$)

表 2 不同因变量的 PLS 模型拟合参数

Tab.2 Fitting results of different dependent variables in PLS model

因变量 Y	h	R^2_{Y}	$R^2_{\text{Y(cum)}}$	R^2_{Y}	$R^2_{\text{Y(cum)}}$	Eig	Q^2	Q^2_{cum}
log BCF	1	0.318	0.318	0.645	0.645	4.13	0.398	0.398
	2	0.386	0.505	0.085	0.965	1.54	0.288	0.865
log($\text{BCF}+N_{\text{Br}}^{\text{h}}$)	1	0.364	0.364	0.911	0.911	4.73	0.839	0.839
	2	0.158	0.522	0.087	0.980	2.05	0.051	0.921
log($\text{BCF}\times N_{\text{Br}}^{\text{h}}$)	1	0.365	0.365	0.893	0.893	4.75	0.817	0.817
log($\text{BCF}/N_{\text{Br}}^{\text{d}}$)	1	0.366	0.366	0.906	0.906	4.75	0.788	0.788

为因变量建立的 PLS 模型的模型拟合精度相对较高, Q_{cum}^2 为 0.921,包含了自变量 52.2%的变异信息,对因变量的解释能力高达 98.0%; $\log(\text{BCF} + N_{Br})$, $\log(\text{BCF} \times N_{Br})$ 和 $\log(\text{BCF}/N_{Br})$ 优化后的 Q_{cum}^2 为 0.865,0.817 和 0.788,所建立的模型分别包含了更多的自变量变异信息 50.5%,36.5%和 36.6%,对因变量的解释能力为 96.5%,89.3%和 90.6%. 综合比较分析 Q_{cum}^2 和 $R_{Y(cum)}^2$ 可知, $\log(\text{BCF} - N_{Br})$ 的拟合效果最好.

2.2 模型的验证

为了确定最优模型,本研究对模型进行了验证. 将所有样本(模型训练组和模型检验组)的自变量分别带入到各 PLS 模型中,并将得到的预测结果与实验数值进行比较^[10],结果如表 3 所示. 表 3 中: * 为模型检验组的数据,残差指预测值与文献中实验值之差. 由表 3 可知: $\log(\text{BCF} + N_{Br})$, $\log(\text{BCF} - N_{Br})$, $\log(\text{BCF} \times N_{Br})$ 和 $\log(\text{BCF}/N_{Br})$ 4 个模型均具有一定的预测能力,其残差范围分别为-0.60~0.69, -0.49~0.96, -1.53~1.13,0.21~6.02. 从残差范围来看, $\log(\text{BCF} + N_{Br})$ 与 $\log(\text{BCF} - N_{Br})$ 的预测效果明显优于其他两个模型,预测效果较好. PBDEs4 个模型检验组的残差,如图 2 所示.

表 3 各因变量预测值与实验值对比

Tab. 3 Comparison of predictive value and experimental value of different dependent variables

编号	文献值	log(BCF+N _{Br})		log(BCF−N _{Br})		log(BCF×N _{Br})		log(BCF/N _{Br})		
		预测值	残差	预测值	残差	预测值	残差	预测值	残差	
模型训练组	BDE-28	5.83	6.52	0.69	5.59	−0.24	6.96	1.13	6.04	0.21
	BDE-51	5.88	5.70	−0.18	5.75	−0.13	6.85	0.97	6.35	0.47
	BDE-77	6.29	6.28	−0.01	6.59	0.30	6.38	0.09	7.52	1.23
	BDE-99	6.01	5.92	−0.09	5.81	−0.20	6.98	0.97	7.71	1.70
	BDE-119	6.29	6.10	−0.19	6.54	0.25	4.76	−1.53	9.66	3.37
	BDE-126	6.43	6.38	−0.05	6.40	−0.03	5.86	−0.57	9.16	2.73
	BDE-154	6.15	5.90	−0.25	6.29	0.14	6.07	−0.08	8.90	2.75
	BDE-203	5.93	5.33	−0.60	5.98	0.05	5.46	−0.47	10.43	4.50
	BDE-204	5.75	5.54	−0.21	5.47	−0.28	5.68	−0.07	9.48	3.73
	BDE-207	5.75	5.59	−0.16	5.70	−0.05	5.31	−0.44	10.65	4.90
	BDE-208	5.57	6.07	0.50	5.08	−0.49	5.78	0.21	9.07	3.50
	BDE-209	5.19	5.07	−0.12	6.15	0.96	5.23	0.04	11.21	6.02
	BDE-197	5.30	5.98	0.68	5.00	−0.30	5.21	−0.09	11.24	5.94
模型检验组	BDE-47 *	5.84	5.76	−0.08	6.00	0.16	6.97	1.13	6.93	1.09
	BDE-66 *	6.05	6.52	0.47	5.92	−0.13	7.23	1.18	7.08	1.03
	BDE-100 *	6.36	5.43	−0.93	6.20	−0.16	6.20	−0.16	8.10	1.74
	BDE-153 *	6.13	5.96	−0.17	5.73	−0.40	6.35	0.22	9.10	2.97
	BDE-198 *	5.68	6.25	0.57	5.82	0.14	5.70	0.02	9.97	4.29
	BDE-206 *	5.44	5.81	0.37	6.57	1.13	5.20	−0.24	11.27	5.83

从图 2 可以看出: $\log(\text{BCF} - N_{Br})$ 模型检验组的残差值相对较低($\log(\text{BCF} + N_{Br})$, $\log(\text{BCF} \times N_{Br})$ 和 $\log(\text{BCF}/N_{Br})$ 检验组的残差绝对值分别为 2.59,2.95 和 16.95),其残差绝对值之和为 2.12,是 4 种模型中最低的. 这也说明了以 $\log(\text{BCF} - N_{Br})$ 为因变量建立的 PLS 模型为本研究的最佳预测模型.

2.3 BCF 影响因素的分析

以 $\log(\text{BCF} - N_{Br})$ 为因变量建立的模型,如表 4 所示. 表 4 中:x 为自变量;a 为系数. 表 4 列出该最优模型的自变量在方程中对应的系数及相应的 VIP 值,所得回归模型为

$$Y(\text{BCF}) = a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_3 + \cdots + C.$$

上式中:系数 a 为各自变量对应的系数值;常数为 C.

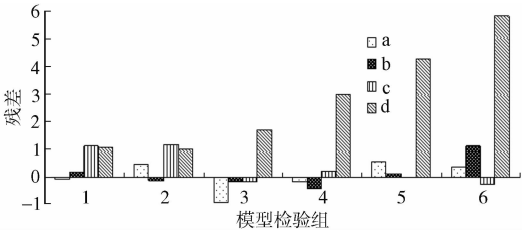


图 2 实验值与模型预测值残差的比较

Fig. 2 Residual comparison of the observed value and predicted value

表 4 自变量在模型方程中的系数及其重要性指标
Tab. 4 Variable importance in the projection and variable coefficient in the model equation

<i>x</i>	VIP 值	<i>a</i>	<i>x</i>	VIP 值	<i>a</i>
E_T	1.71	0.000 1	E_{NHOMO}	0.86	-16.900
R_e	1.48	-0.000 1	μ	0.79	0.215
$Q_{C6'}$	1.25	-2.588 0	Q_{C2}	0.79	1.359
M	1.20	-3.785 0	Q_{C4}	0.73	-1.059
$E_{LUMO}-E_{NHOMO}$	1.14	24.680 0	$Q_{C5'}$	0.73	-1.983
Q_O	1.11	-28.990 0	$RO-C1'$	0.66	-13.840
E_{NHOMO}	1.08	-11.710 0	$E_{LUMO}+E_{NHOMO}$	0.63	2.148
$Q_{C2'}$	1.08	-2.524 0	E_{NLUMO}	0.51	10.490
$Q_{C1'}$	0.94	6.826 0	$Q_{C3'}$	0.48	-1.434
Q_{C3}	0.87	2.233 0	常数	—	0.844

在 PLS 建模中,VIP 是一个反应自变量重要性相对大小的参数,自变量的 VIP 值越大,说明该自变量对因变量的影响越大.由表 4 可知:分子总能量 E_T 与电子空间广度 R_e 的 VIP 值分别为 1.71,1.48,远高于其他自变量的 VIP 值,说明这两个自变量对 PBDEs 的 BCF 有着重要的控制作用; R_e 在同系物中可用于表征不同分子间分子体积的大小,虽然某些情况下 R_e 与分子体积间的相关性不好,但对 PBDEs 这类结构上非常相似的同系物而言却具有较好的相关性^[11]; E_T 数值上的较大变化意味着分子组成中取代溴原子个数的变化,而相同分子组成的同分异构体之间也存在着微小的差异; R_e 的系数为正, E_T 的系数为负,说明在 PBDEs 的溴取代基数目相同的情况下,随着 R_e 数值的减小或 E_T 数值的增大, $\log(BCF-N_{Br})$ 的值逐渐减小,即相应的 PBDEs 更不易富集在土壤、沉淀物、生物体中.

Harrie 等^[12]和 Qin 等^[13]对二噁英及多氯联苯富集情况的研究,也说明了卤素(氯和溴)取代基数目对有机化合物的 BCF 数值有较大的影响.

3 结论

通过 Gaussian 03 将 PBDEs 进行 B3LYP/6-31G* 水平上的分子结构优化,应用 PLS 分析,建立了 PBDEs 生物富集系数的 QSAR 模型.结果表明:溴取代基数对 PBDEs 的 BCF 有较大影响,直接用 $\log BCF$ 建模无法得到较好的预测模型.

以 $\log(BCF+N_{Br})$, $\log(BCF-N_{Br})$, $\log(BCF\times N_{Br})$ 和 $\log(BCF/N_{Br})$ 作为 4 种因变量 Y 建模,发现用 $\log(BCF-N_{Br})$ 建立的模型具有较好的预测能力,并通过 $\log(BCF-NB)$ 模型建模,得到其拟合相关系数平方和累积交叉有效性判别系数分别为 0.980 和 0.921,可有效预测 PBDEs 的生物富集系数.此外研究发现 E_T , R_e 以及 N_{Br} 对 PBDEs 的 BCF 有着较大的影响.较低 E_T , 较大 R_e 的 PBDEs 呈现出较低的 BCF,随着溴原子数的递增,BCF 呈现先增大后减小的趋势.

参考文献:

[1] HITES R A. Polybrominated diphenyl ethers in the environment and in people: Ameta-analysis of concentrations [J]. Environ Sci Technol, 2004, 38(4): 945-956.

[2] 刘艺凯,唐建辉,潘晓辉,等.环境中多溴联苯醚分析方法的研究进展[J].环境化学,2012,31(12):1908-1915.

[3] RAHMAN F, LANGFORD K H, SCRIMSHAW M D, et al. Polybrominated diphenyl ether (PBDE) flame retardants[J]. Sci Total Environ, 2001, 275(1/2/3): 1-17.

[4] CHEN J W, HARNER T, YANG P, et al. Quantitative predictive models for octanol-air partition coefficients of polybrominated diphenyl ethers at different temperatures[J]. Chemosphere, 2002, 51(7): 577-584.

[5] 王连生. 有机污染化学[M]. 北京: 高等教育出版社, 2005: 253-254.

[6] 陈景文. 有机污染物定量结构-性质关系与定量结构-活性关系[M]. 大连: 大连理工大学出版社, 1999: 1-8.

[7] 谷成刚, 蒋新, 颜冬云, 等. 二噁英结构与正辛醇-水分配系数相关性的密度泛函理论[J]. 环境科学学报, 2008, 28(1): 185-191.

[8] 张幸川, 杨郭英, 王遵尧, 等. 密度泛函方法预测二噁英类化合物(PCDD/Fs)的正辛醇/水分配系数[J]. 环境科学学

报,2007,27(1):257-266.

- [9] RAMALINGAM M, RAMASAMI K, VENUVANALINGAM P, et al. Ab initio and DFT investigations of the mechanistic pathway of singlet bromocarbenes insertion into C—H bonds of methane and ethane[J]. Lecture Notes in Computer Science, 2007, 4488(2): 288-295.
- [10] MEYLAN W M, HOWARD P H, BOETHLING R S, et al. Improved method for estimating bioconcentration/bioaccumulation factor from octanol/water partition coefficient[J]. Environ Toxicol Chem, 1999, 18(4): 664-672.
- [11] LU Gui-ning, DANG Zhi, TAO Xue-qin, et al. Modeling and prediction of photolysis half-lives of polycyclic aromatic hydrocarbons in aerosols by quantumchemical descriptors[J]. Science of the Total Environment, 2007, 373(1): 289-296.
- [12] HARRIE A J, COVERS H, KROP B. Partition constants of chlorinated dibenzofurans and dibenzo-p-dioxins[J]. Chemosphere, 1998, 37(9): 2139-2152.
- [13] QIN Li-tang, LIU Shu-sheng, LIU Hai-ling. A new predictive model for the bioconcentration factors of polychlorinated biphenyls (PCBs) based on the molecular electronegativity distance vector (MEDV)[J]. Chemosphere, 2008, 70(9): 1577-1587.

Quantitative Structure-Activity Relationships on Bioconcentration Factor of PBDEs

LI Ji-an¹, XUE Xiu-ling¹, LU Gui-ning²

(1. College of Chemical Engineering, Huaqiao University, Xiamen 361021, China;

2. School of Environment and Energy, South China University of Technology, Guangzhou 510006, China)

Abstract: In this paper, the molecular structure of polybrominated diphenyl ethers (PBDEs) was optimized by using Gaussian 03 software at B3LYP/6-31G* level. Then the relationships between the bioconcentration factor (BCF) of PBDEs and the molecular structure parameters were correlated by partial least squares (PLS). The results showed that BCF combined with the number of bromine atoms on the benzene ring could better express the quantitative relationship between physicochemical properties and molecular structure of PBDEs. The correlation coefficients of fitting ($R_{Y(cum)}^2$) and cross validation (Q_{cum}^2) of the optimized model $\log(BCF - N_{Br})$ were 0.980 and 0.921 respectively, which showed that the model had a high fitting precision and a good predicting ability. Furthermore, molecular total energy (E_T) and electronic spatial extent (R_e) both played important roles in the BCF of PBDEs, lower E_T and larger R_e had less BCF.

Keywords: polybrominated diphenyl ethers; quantitative structure-activity relationship; bioconcentration factor; partial least squares

(责任编辑: 陈志贤 英文审校: 刘源岗)