

DOI: 10.7524/j.issn.0254-6108.2015.10.2015032605

范德玲, 刘济宁, 王 蕾, 等. 羟基自由基反应常数定量预测模型[J]. 环境化学, 2015, 34(10): 1924-1931

FAN Deling, LIU Jining, WANG Lei, et al. QSAR model for predicting hydroxyl radical reaction constant of organic chemicals[J]. Environmental Chemistry, 2015, 34(10): 1924-1931

羟基自由基反应常数定量预测模型*

范德玲 刘济宁** 王 蕾 周林军 石利利

(环保部南京环境科学研究所, 南京, 210042)

摘 要 羟基自由基($\cdot\text{OH}$)反应常数对于表征有机污染物在大气环境中持久性具有重要意义. 依据经济合作与发展组织(OECD)关于 QSAR 模型构建与验证的导则, 采用量子化学方法对覆盖了不同种类的 722 个化合物进行结构优化, 遗传算法筛选最优结构描述符, 运用多元线性回归构建化学品羟基自由基反应常数预测模型. 拟合结果显示, 多元线性回归模型决定系数 R^2 和标准误差分别为 0.819 和 0.508, 基于 leverage 法评价模型的应用域, 结果表明模型具有较强的稳健性、预测性和拟合能力. 美国环保局 EPI Suite 中 AOPWIN 模块羟基自由基反应常数预测模型没有给出明确的应用域, 利用所建模型与美国 EPI Suite 对化学物质进行比较, 其中, 有 85 个化学物质预测优于 EPI Suite 软件. 通过定量结构-活性关系(QSAR)预测技术可弥补羟基自由基反应常数测试数据的缺失, 减少测试费用和评估数据的不确定性.

关键词 化学品, 羟基自由基反应常数, 定量结构-性质/活性关系.

QSAR model for predicting hydroxyl radical reaction constant of organic chemicals

FAN Deling LIU Jining** WANG Lei ZHOU Linjun SHI Lili

(Nanjing Institute of Environmental Sciences, Ministry of Environmental Protection, Nanjing, 210042, China)

Abstract: Reaction rate constants of hydroxyl radicals (k_{OH}) are of vital importance for assessing the persistence of organic pollutants in the atmosphere. 722 chemicals were optimized by quantum chemistry method and screened by genetic algorithm for optimal descriptors. A quantitative structure-activity relationship (QSAR) model for hydroxyl radical reaction constant was developed by employing multiple linear regression (MLR) according to the QSAR model development and validation guideline issued by Organization for Economic Cooperation and Development (OECD). The QSAR modeling results indicated that the R^2 and SE of multiple linear regressions model were 0.819 and 0.508, The application domain was evaluated based on the leverage method. The results showed that the model had high goodness-of-fit, robustness and good predictive ability. The hydroxyl radical reaction constant prediction model in AOPWIN module of the EPI Suite software did not have a definite application domain and the prediction of 85 chemicals was better than EPI software Suite. Thus, the QSAR model can be employed to make up the missing data, reduce the cost of testing and evaluated of data uncertainty.

Keywords: organic chemicals, hydroxyl radical reaction constant, quantitative structure-activity relationship (QSAR).

2015 年 3 月 26 日收稿.

* 国家高技术研究发展计划(863 计划)(2013AA060A308); 环保公益性行业科研专项经费(2013467028)资助.

** 通讯联系人, E-mail: ljn@nies.org

目前世界上大约有 700 多万种化学物质,其中常用化学物质超过 7 万多种,并且每年 1000—2000 种新的化学物质问世.我国已生产和上市销售的现有化学物质大约有 4.6 万多种,每年申报新化学物质约上千种^[1].化学品在生产、存储、销售、运输、使用及废物处置中,会对人类健康和生态环境产生不同程度的危害影响.有机污染物通过与大气化学过程的主要氧化剂,包括·OH 自由基、臭氧、NO₃ 自由基等反应是降解的重要途径,其中,与·OH 自由基的反应在化学过程中起着关键作用^[2].化合物与·OH 自由基反应常数是表征有机污染物在大气持久性的重要参数,是进行环境持久性和生态风险评估的重要指标.仅通过实验方法获取化学品的 OH 自由基反应常数成本高、费时费力,难以满足化学物质生态风险评估的需要.所以,迫切需要发展科学快捷有效的预测氧化反应常数的理论计算方法.经济合作与发展组织(OECD)于 2007 年发布了 QSAR 模型构建与验证的导则^[3],提出了 QSAR 模型应满足的标准:(1)具有明确定义的环境指标;(2)具有清晰和明确的数学算法;(3)定义了模型的应用域;(4)模型具有适当的拟合优度、稳健性和预测能力;(5)尽可能进行模型机理解释.

针对·OH 自由基反应速率常数,国内外开展了一系列应用 QSAR 方法的预测研究,主要包括:Atkinson 等发展的基团贡献法^[4],分子轨道法^[5]、电离能估算法和分子描述符定量结构-活性关系的模型.1995 年,Atkinson^[6]等人利用基团贡献法对 485 个有机化合物进行·OH 自由基反应速率常数的预测,在建模过程中加入了新的取代基因子和官能团,例如氟化醚,提高了模型预测效果.2004 年,Gramatica^[7]等人采用 D-最优实验设计和 Kohnoen 人工神经网络(K-ANN)对数据集进行训练集和验证集的划分,采用遗传算法-多元线性回归分析对训练集进行描述符选择和建模,对 456 个不同类型的挥发性有机化合物进行了 QSAR 的研究,模型具有较好的稳健性、预测能力.2005 年,Öberg^[8]采用 DRAGON 软件计算 743 个挥发性和半挥发性的有机污染物分子结构描述符,利用偏最小二乘法构建·OH 反应常数的预测模型.2009 年,Fatemi^[9]使用 CODESSA 来计算 98 个烯烃类分子结构描述符,用逐步多元线性回归和遗传算法筛选最优描述符,采用多元线性回归(MLR)、人工神经网络(ANN)和支持向量机(SVM)等多种建模方法进行建模.

美国 EPI Suite 中 AOPWIN 程序采用基团贡献法进行·OH 自由基反应常数预测,根据有机化合物与大气中的自由基反应的途径,针对化合物分子所具有的特定基团,利用取代因子和官能团速率因子来预测反应速率,但是未给出模型的应用域^[10].由于有机污染物在大气对流层的浓度极低,阳光产生的羟基自由基浓度稳定,所以化学物质与羟基自由基的反应可以看做是准一级动力学反应.大气中化学物质的半衰期为:

$$t_{1/2} = 0.693 / (K_{OH} [\cdot OH])$$

式中, K_{OH} 为化学物质与羟基自由基反应的速率常数($\text{cm}^3 \cdot \text{mol}^{-1} \cdot \text{s}^{-1}$), $[\cdot OH]$ 为单位空间内羟基自由基的分子个数($\text{mol} \cdot \text{cm}^{-3}$).

EPIWIN 软件用于化学物质理化参数预测,用户在缺少化学品测试信息的情况下,仅需输入化学物质名称、CAS 号或 SMILES 码进行理化参数属性查询.表 1 给出了 AOPWIN 使用的部分碎片及其羟基自由基反应速率常数.

表 1 碎片羟基自由基反应速率常数

Table 1 Hydroxyl radical reaction constant of some organic groups

碎片	$K_{OH}(\times 10^{12}) / (\text{cm}^3 \cdot \text{mol}^{-1} \cdot \text{s}^{-1})$	碎片	$K_{OH}(\times 10^{12}) / (\text{cm}^3 \cdot \text{mol}^{-1} \cdot \text{s}^{-1})$
(—CH ₃)	0.136	(—OH)	0.14
(>N— NO ₂)	1.3	(—NH ₂)	21
(P(=S))	53	(—NH—)	63
(>C<)	0	(>N—)	66

目前,已有研究发展了关于 K_{OH} 的 QSAR 模型.然而,绝大多数模型并不满足 OECD 关于 QSAR 模型的标准,影响了其在化学品管理上的应用.美国 EPI Suite 软件中 AOPWIN 模块羟基自由基反应常数预测模型没有给出明确的应用域,不利于使用者评估预测的化合物是否处于模型应用域内.

基于上述分析,为了能够快捷、有效、准确地预测化学物质的羟基自由基反应常数,同时,弥补测试

数据的缺失、减少测试费用.本研究采用遗传算法-多元线性回归(GA-MLR)方法构建羟基自由基反应常数预测模型,基于 leverage 法评价模型的应用域,并且与美国环保局的 EPI 组件中 AOPWIN 模块对有机化学物质羟基自由基反应常数的预测进行比较.

1 数据与方法

1.1 样本采集及筛选

不同类型有机化合物化学品的羟基自由基反应常数是从文献[11]中整理得到的.为了建立有效的 QSAR 模型,采用分组方法是 Kennard & Stone 方法^[12],把数据集分成训练集和验证集,在一定程度上能够避免训练集样本分布不均匀,能够很好地将数据集划分为训练集和验证集,训练集和验证集样本数分别为 506 和 216 个化合物.

1.2 分子描述符计算

采用 ChemDraw 软件中构建出 722 个有机化合物的分子结构,导入 HyperChem 程序对分子进行优化,采用 MM+分子力场方法和半经验量子力学 AM1 方法进行构型优化.优化后的结构导入到 DRAGON 5.4 软件中计算 1664 个不同类型的理论分子描述符^[13].建模前对这些描述符进行了预处理,即将常数项、接近常数的项和具有高度相关(相关系数大于 0.96 的两个分子描述符中与目标值相关系数较小的)的分子描述符删除.最后筛选 717 个描述符用于后面的变量选择过程^[14].

2 模型构建与应用

2.1 模型构建

采用 MobyDigs 软件进行遗传算法来选择与羟基自由基反应常数具有高度相关的描述符^[15-16].经过遗传算法变量选择后,用多元线性回归(MLR)方法建立线性 QSAR 模型,即 GA-MLR 模型.模型评价函数选择留一法交互检验(Leave-one-out cross validation),增加一个描述符后模型的性能没有明显变化时(增加一个描述符, Q^2 增加小于 0.02),达到最佳描述符个数,本研究中模型最佳描述符个数为 7.建模过程中相关参数:种群大小(Population size)为 100,初始模型允许的做大变量数(Maximum allowed variables)为 7,变异均衡值(Mutation trade-off, T)为 0.5,交叉(Crossover)和变异(Mutation)概率均基于 T 参数.

2.2 模型验证

经过遗传算法变量选择后,用多元线性回归方法建立线性 QSAR 模型,即 MLR 模型.线性 MLR 方程如下:

$$Y = -14.628 - 0.449\text{MAXDN} + 0.155\text{IC1} + 0.295\text{SPH} - 0.181\text{H2m} + 0.468\text{R1u} + 0.148\text{nR} \\ = \text{Cp} + 0.170\text{nRCHO}$$

$$n_{\text{tr}} = 506 \quad R^2 = 0.819 \quad \text{RMSE} = 0.508 \quad Q_{\text{Loo}}^2 = 0.803 \quad \text{RMSE}_{\text{Loo}} = 0.517$$

$$n_{\text{test}} = 216 \quad R_{\text{ext}}^2 = 0.655 \quad \text{RMSE}_{\text{ext}} = 0.433 \quad Q_{\text{ext}}^2 = 0.807.$$

其中,MAXDN 表示最大负静电拓扑变化,IC1 表示邻近对称性,SPH 表征分子三维分子结构,H2m 表示分子的分子质量信息,nR=Cp 表示分子中末端伯碳(sp²)的数目,nRCHO 表示脂肪族分子中醛官能团的数目,R1u 描述化合物结构片段的相关信息.表 2 列出了 7 个描述符的名称、物理化学意义以及标准回归系数.训练集的均方根误差 RMSE 为 0.508,相关系数 R^2 、抽一法交互检验相关系数 q^2 分别为 0.819 和 0.803;验证集预测的均方根误差 RMSE 为 0.433,相关系数 R^2 为 0.655,多元线性回归模型实验值与预测值之间的相关性见图 1.

2.3 模型的应用域

利用杠杆(Leverage)方法进行模型应用范围的分析评价^[17],不仅可以将模型应用范围量化,而且以图形的方式,即 Williams 图(图 2)直观表达^[18].通过模型应用域研究,研究数据集的 645 个化合物大部分都位于应用域范围内(图 2),即化合物的 \hat{h} 值 $\leq h^*$ ($h^* = 0.048$),并且大部分化合物的预测误差都

在 3 倍标准偏差以内,说明模型具有较好的稳健性^[19].X 例外点包括 1 个烯烃类,8 个酮类,1 个烷烃类,Y 例外点包含 1 个乙酰氯,1 个 1,3,5-三嗪和 1 个 2-硝基萘.

表 2 羟基自由基反应常数的 GA-MLR 模型中所用描述符的物理化学意义和相应系数

Table 2 Physical chemistry definition and the corresponding coefficients of GA-MLR model

描述符	意义	回归系数	回归系数偏差	标准回归系数
常数项		-14.628	0.147	
MAXDN	最大负静电拓扑变化	-0.31	0.017	-0.449
IC1	邻近对称性	0.405	0.058	0.155
SPH	表征分子三维分子结构	1.447	0.111	0.295
H2m	描述分子的分子质量信息	-0.493	0.06	-0.181
R1u	描述化合物结构片段的相关信息	1.448	0.067	0.468
nR=Cp	分子中末端伯碳(sp ²)的数目	0.522	0.07	0.148
nRCHO	脂肪族分子中醛官能团的数目	0.708	0.082	0.17

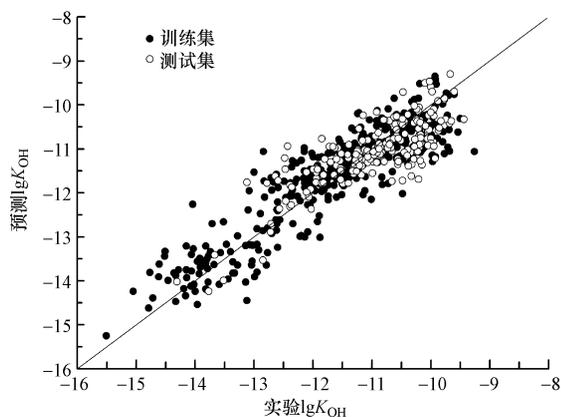


图 1 羟基自由基反应常数的 MLR 模型的预测值与实验值的相关性

Fig.1 Plot of experimental versus predictive $\lg K_{OH}$ values by MLR method

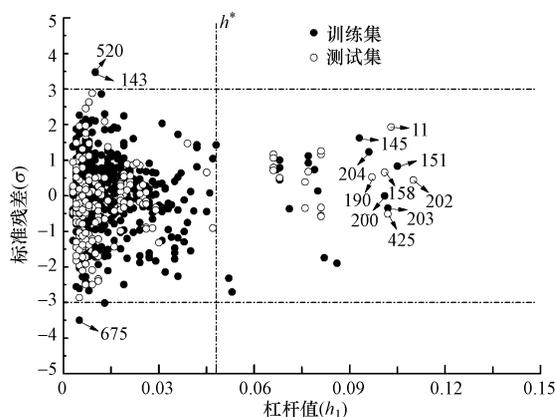


图 2 羟基自由基反应常数的 MLR 模型的应用域 Williams 图

Fig.2 Williams plot of the developed model by MLR method

2.4 模型与现有模型比较

将构建的羟基自由基反应常数 QSAR 模型与现有预测模型进行比较^[7,9,20],通过比较可以得出:(1)本文模型数据集是相对比较大,覆盖了不同种类的有机化合物, QSAR 模型具有较广的应用域.(2)本文构建的 QSAR 模型相对简单和快捷,模型中描述符依靠分子结构信息得到,而不需要复杂的计算.另外一方面模型通过多元线性回归分析构建,可以明确地看出哪些分子结构特征对化合物分子的氧化降解具有较大影响.综上所述,本文所构建的羟基自由基反应常数 QSAR 模型是一个非常有效的预测模型,可以用来预测不同类型的有机化合物.

2.5 模型与 EPI Suite 预测结果比较

将 216 个化学物质作为验证集,用于检验该模型的实际应用能力.同时与美国环保局的 EPI 组件中 AOPWIN 模块对该组物质进行羟基自由基反应常数的预测进行比较,具体预测及实验数据如表 3 所示.

本文所建模型与美国 EPI Suite 组件中的 AOPWIN 模块对该 216 个化学物质预测结果进行比较(表 3).对比实测结果发现,有 85 个化学物质的预测优于 EPI Suite 的预测结果.可能是由于化学物质结构比较复杂,部分 EPI Suite 软件中选取的结构碎片不能完全表达分子结果,所建模型预测结果优于 EPI Suite 软件中的 AOPWIN 模块的预测,23 个化学物质的预测结果与 EPI Suite 的预测结果相近,57 个化学物质的预测误差均较大,51 个化学物质 EPI Suite 的预测结果较好.然而,EPI Suite 软件中未给出模型的应用域,本研究所建模型,弥补了 EPI Suite 软件的不足.

表 3 216 个化学物质的羟基自由基反应常数预测结果比较

Table 3 Comparison of predicted K_{OH} for 216 chemicals

CAS	K_{OH} (实测值)	预测值	EPI 预测值	CAS	K_{OH} (实测值)	预测值	EPI 预测值
000074-84-0	-12.57	-12.36	-12.57	000513-35-9	-10.06	-10.06	-10.06
000074-98-6	-11.94	-11.93	-11.90	000590-18-1	-10.25	-10.96	-10.25
000106-97-8	-11.6	-11.61	-11.58	000624-64-6	-10.19	-10.19	-10.25
000109-66-0	-11.4	-11.4	-11.39	000563-79-1	-9.96	-10.84	-9.96
000078-78-4	-11.41	-11.41	-11.39	000763-29-1	-10.2	-10.21	-10.27
000463-82-1	-12.07	-12.09	-12.17	000627-20-3	-10.19	-10.82	-10.24
000096-14-0	-11.24	-11.02	-11.24	000646-04-8	-10.17	-10.82	-10.24
000079-29-8	-11.21	-11.07	-11.26	000674-76-0	-10.22	-10.75	-10.23
000287-92-3	-11.29	-11.36	-11.34	000690-08-4	-10.26	-10.9	-10.24
004551-51-3	-10.76	-10.88	-10.89	000563-46-2	-10.21	-10.23	-10.28
000493-02-7	-10.69	-10.84	-10.73	000625-27-4	-10.05	-10.04	-10.05
000593-53-3	-13.77	-14.24	-13.89	014686-13-6	-10.17	-10.18	-10.22
000075-09-2	-12.85	-13.53	-12.87	000142-29-0	-10.17	-10.18	-10.23
000075-43-4	-13.52	-13.53	-13.58	000110-83-8	-10.17	-10.18	-10.21
000075-00-3	-12.39	-12.39	-12.39	000591-49-1	-10.03	-10.56	-10.04
000074-96-4	-12.46	-12.47	-12.49	000106-99-0	-10.18	-10.35	-10.18
000079-00-5	-12.71	-12.7	-12.64	000590-19-2	-10.58	-10.56	-10.51
000811-97-2	-14.31	-14.02	-14.21	000598-25-4	-10.24	-10.1	-10.24
000540-54-5	-11.95	-11.93	-12.03	001574-41-0	-10	-10.01	-9.98
000075-29-6	-12.04	-12.05	-12.08	000078-79-5	-10	-9.99	-9.98
000142-28-9	-12.11	-12.01	-12.00	020237-34-7	-9.95	-10.1	-9.97
000507-20-0	-12.39	-11.82	-12.39	000591-93-5	-10.28	-10.28	-10.27
000109-69-3	-11.82	-12.80	-11.66	000926-56-7	-9.88	-9.87	-9.85
000078-86-4	-11.64	-11.59	-11.84	000763-30-4	-10.1	-9.5	-10.10
000543-59-9	-11.51	-11.43	-11.45	005194-50-3	-9.87	-10.88	-9.85
000058-89-9	-12.72	-12.89	-12.24	005194-51-4	-9.87	-10.88	-9.85
000075-07-0	-10.8	-10.79	-10.77	000764-13-6	-9.68	-10.87	-9.58
000123-38-6	-10.71	-10.63	-10.66	007319-00-8	-10.04	-10.04	-10.08
000078-84-2	-10.58	-10.60	-10.63	004049-81-4	-10.02	-9.47	-10.10
000590-86-3	-10.56	-10.65	-10.53	004054-38-0	-9.86	-10.52	-9.84
000107-02-8	-10.7	-10.76	-10.59	000628-41-1	-10	-10.01	-9.94
000078-85-3	-10.47	-9.47	-10.46	002612-46-6	-9.96	-10.25	-9.95
004170-30-3	-10.44	-10.39	-10.44	000821-07-8	-9.95	-10.25	-9.95
022056-82-2	-10.78	-10.79	-10.16	000123-35-3	-9.67	-9.67	-9.71
000078-93-3	-11.94	-11.90	-11.88	000080-56-8	-10.27	-10.28	-10.04
000096-22-0	-11.7	-11.69	-11.61	005989-27-5	-9.77	-9.78	-9.84
000591-78-6	-11.04	-11.19	-11.17	000099-85-4	-9.75	-10.33	-9.75
000563-80-4	-11.54	-11.57	-11.58	000099-86-5	-9.44	-9.45	-9.58
006004-44-0	-10.15	-10.16	-10.06	000555-10-2	-9.77	-9.78	-9.82
020334-52-5	-9.93	-11.22	-10.06	000586-62-9	-9.65	-10.37	-9.70
000120-92-3	-11.53	-11.44	-11.17	004497-92-1	-10.1	-10.11	-10.05
000108-94-1	-11.19	-11.32	-10.92	000107-00-6	-11.1	-11.1	-11.09
005878-19-3	-11.17	-11.19	-11.28	000627-19-0	-10.95	-10.98	-11.02
005077-67-8	-11.11	-11.12	-11.42	000693-02-7	-10.9	-10.97	-10.96
000513-86-0	-10.99	-11.00	-11.23	000075-02-5	-11.25	-11.25	-11.26
003393-64-4	-10.79	-11	-10.81	000075-01-4	-11.16	-10.35	-11.26
000820-69-9	-10.34	-11.08	-10.34	000075-35-4	-10.96	-11.14	-11.64

续表3

CAS	K_{OH} (实测值)	预测值	EPI 预测值	CAS	K_{OH} (实测值)	预测值	EPI 预测值
000107-22-2	-10.94	-10.71	-10.60	000156-59-2	-11.58	-11.69	-11.60
002363-83-9	-10.62	-10.60	-10.39	000156-60-5	-11.63	-11.62	-11.60
000067-56-1	-12.03	-12.02	-12.21	000127-18-4	-12.78	-12.66	-12.67
000064-17-5	-11.49	-11.48	-11.45	010061-02-6	-10.85	-10.97	-11.03
000071-23-8	-11.26	-11.44	-11.26	000075-08-1	-10.33	-10.32	-10.40
000067-63-0	-11.29	-11.25	-11.14	000107-03-9	-10.32	-11.35	-10.37
000071-36-3	-11.07	-11.08	-11.16	000075-33-2	-10.38	-11.30	-10.32
000071-41-0	-10.95	-10.98	-11.08	000109-79-5	-10.29	-11.28	-10.35
006032-29-7	-10.93	-11.14	-10.94	000513-53-1	-10.4	-10.25	-10.28
000584-02-1	-10.91	-11.17	-10.88	000075-66-1	-10.48	-11.47	-10.48
000096-41-3	-10.97	-11.3	-10.99	000624-89-5	-11.07	-10.98	-10.99
000598-75-4	-10.91	-11.16	-10.94	000111-47-7	-10.7	-10.69	-10.65
000115-18-4	-10.19	-10.67	-10.56	000067-68-5	-10.21	-11.69	-10.21
000107-07-3	-11.85	-11.78	-11.76	000074-89-5	-10.66	-10.66	-10.65
000111-35-3	-10.66	-11.05	-10.67	000075-04-7	-10.56	-11.55	-10.53
002517-43-3	-10.63	-10.77	-10.68	000124-40-3	-10.18	-11.27	-10.18
000928-96-1	-9.96	-10.85	-10.20	000075-50-3	-10.22	-10.21	-10.16
000057-55-6	-10.92	-11.31	-10.89	000151-56-4	-11.21	-11.20	-11.19
000115-10-6	-11.53	-11.7	-11.78	000060-34-4	-10.19	-10.9	-10.19
000060-29-7	-10.88	-11.15	-10.91	003710-84-7	-10	-10.87	-10.08
001634-04-4	-11.53	-11.49	-11.65	000108-01-0	-10.05	-10.92	-10.08
000110-88-3	-11.21	-10.79	-10.99	000124-68-5	-10.55	-11.29	-10.59
000628-28-4	-10.83	-11.08	-10.88	000062-75-9	-11.6	-11.23	-11.60
000628-81-9	-10.67	-10.97	-10.73	000108-03-2	-12.36	-11.87	-12.37
000503-30-0	-10.99	-11.55	-11.35	000627-05-4	-11.82	-11.77	-11.79
000109-99-9	-10.79	-11.29	-10.85	003638-64-0	-11.92	-11.47	-11.27
000142-68-7	-10.86	-11.12	-10.63	000625-46-7	-10.91	-11.08	-10.90
000505-22-6	-11.04	-11.62	-10.71	000107-29-9	-11.66	-10.95	-11.68
000497-26-7	-11.03	-11.23	-10.86	000079-16-3	-11.28	-11.26	-11.21
007778-85-0	-10.51	-10.81	-10.63	001187-58-2	-11.12	-11.02	-11.14
000109-87-5	-11.31	-11.57	-11.29	000107-13-1	-11.39	-10.91	-11.38
000110-71-4	-10.56	-11.23	-10.80	000110-86-1	-12.43	-10.94	-12.43
000107-98-2	-10.73	-10.98	-10.78	000288-32-4	-10.44	-10.66	-10.44
000109-59-1	-10.69	-10.87	-10.63	000288-47-1	-11.85	-10.77	-11.85
000111-90-0	-10.24	-11.07	-10.50	000100-69-6	-10.25	-10.32	-10.57
000110-00-9	-10.39	-11.02	-10.39	000624-91-9	-12.59	-11.72	-12.81
000534-22-5	-10.21	-10.54	-9.98	003013-02-3	-10.88	-10.88	-10.87
000098-01-1	-10.45	-11.08	-10.43	000105-40-8	-11.08	-11.04	-11.08
000627-42-9	-11.33	-11.5	-11.46	006135-31-5	-10.98	-11.06	-10.98
006482-24-2	-11.18	-11.59	-11.44	001066-40-6	-11.4	-11.78	-11.41
000421-14-7	-13.67	-13.41	-13.57	000108-88-3	-11.22	-10.54	-11.28
000075-21-8	-13.12	-11.76	-12.55	000100-41-4	-11.15	-10.24	-11.23
000106-88-7	-11.72	-10.95	-11.74	000622-96-8	-10.92	-10.16	-11.13
000079-09-4	-11.94	-11.53	-11.86	000873-66-5	-10.23	-10.37	-10.23
000079-20-9	-12.47	-11.46	-12.59	000105-67-9	-10.15	-10.44	-10.30
000141-78-6	-11.8	-11.23	-11.77	000095-87-4	-10.1	-10.46	-9.97
000623-42-7	-11.52	-11.11	-11.72	000576-26-1	-10.18	-10.4	-10.30
000107-31-3	-12.64	-11.62	-12.66	000095-65-8	-10.09	-10.48	-9.97

续表3

CAS	K_{OH} (实测值)	预测值	EPI 预测值	CAS	K_{OH} (实测值)	预测值	EPI 预测值
000109-94-4	-11.99	-11.46	-11.78	000108-90-7	-12.11	-11.53	-11.86
000108-05-4	-10.6	-10.44	-10.58	000062-53-3	-9.95	-10.94	-12.56
000140-88-5	-10.8	-10.34	-10.96	000121-69-7	-9.83	-10.37	-9.69
000005-09-7	-11.26	-11.51	-11.38	000100-47-0	-12.48	-11.21	-12.46
000074-85-1	-11.07	-10.8	-11.07	000096-18-4	-12.37	-12.28	-12.45
000115-07-1	-10.58	-10.56	-10.58	000541-31-1	-10.28	-10.95	-10.34
000106-98-9	-10.5	-10.44	-10.56	000542-56-3	-11.27	-11.13	-11.51
000109-67-1	-10.5	-10.31	-10.54	027400-72-2	-9.98	-9.7	-9.64
000592-76-7	-10.39	-10.22	-10.50	003779-61-1	-9.6	-9.7	-9.64
000563-45-1	-10.5	-10.42	-10.54	000096-17-3	-10.48	-10.45	-10.57
000558-37-2	-10.55	-10.76	-10.57	000123-15-9	-10.48	-10.32	-10.54
000115-11-7	-10.29	-10.41	-10.29	000104-87-0	-10.89	-10.63	-10.73

注: 黑体为预测结果优于 EPI.

2.6 机理解释

通过对化合物羟基自由基反应常数的 QSAR 模型中所选的描述符物理化学意义的解释, 可以获得影响化合物羟基自由基反应常数高低的结构信息. 其中, R1u 是 GETAWAY (Geometry, Topology and Atom-Weights Assembly) 类描述符, 描述化合物结构片段的相关信息, 因此能够描述出同系物的分子结构差异, 同时可以反映分子的大小和一些特殊原子的性质. 另一个与分子的羟基氧化常数有密切负相关的是 MAXDN, 该描述符是拓扑描述符, 表示最大负静电拓扑变化. IC1 是一个信息指数描述符, 是基于包含氢原子分子图中的邻近自由度和边缘多重性来计算得到的拓扑指数. SPH (Spherosity) 是一个几何描述符, 是根据协方差矩阵的特征值函数计算得到的, SPH 描述符的值允许在 0 到 1 之间变化, 当分子为平面时 (苯) 则 SPH 为 0, 当分子是一个完全的球形分子的时候, 它的 SPH 则为 1, 表征分子的三维分子结构, 能有效的识别出具有微小差别的分子结构, 例如异构体. H2m 也是一个 GETAWAY 描述符, 描述分子的分子质量信息, nR=Cp 和 nRCHO 是官能团数目的描述符, nR=Cp 代表的是分子中末端伯碳 (sp²) 的数目, nRCHO 表征的是脂肪族分子中醛官能团的数目.

3 结论

采用多元线性回归方法构建的羟基自由基反应常数预测模型可以较好地预测不同类别的化合物的羟基自由基反应常数, 是一个简单、明确和有效的预测模型. 数据集涵盖了大量不同种类的有机化合物, 使得 QSAR 模型有较广的应用范围. 与美国 EPI Suite 组件中的 AOPWIN 模块比较, 有 85 个化学物质的预测结果优于 EPI 预测结果, 进一步解释了影响羟基自由基反应常数的描述符及结构信息. 通过定量结构-活性关系 (QSAR) 预测技术可弥补羟基自由基反应常数测试数据的缺失, 减少测试费用和评估数据的不确定性.

参 考 文 献

- [1] Brown V J. Reaching for chemicals safety [J]. Environmental Health Perspectives, 2003, 111: A766-A769
- [2] Sabljic A, Peijnenburg W. Modeling lifetime and degradability of organic compounds in air, soil, and water systems (IUPAC Technical Report)[J]. Pure and Applied Chemistry, 2001, 73: 1331-1348
- [3] Organisation for Economic Co-operation and Development (OECD), Guidance document on the validation of (quantitative) structure activity relationships [(Q)SAR] models, Technical Report for OECD Environment, Health and Safety Publications Series on Testing and Assessment No. 69; Paris, 2007[R]
- [4] Atkinson R. Estimation of gas-phase hydroxyl radical rate constants for organic chemicals [J]. Environmental toxicology and chemistry, 1988, 7 (6): 435-442
- [5] Klamt A. Estimation of gas-phase hydroxyl radical rate constants of organic compounds from molecular orbital calculations [J].

- Chemosphere, 1993, 26 (7): 1273-1289
- [6] Kwok E S, Atkinson R. Estimation of hydroxyl radical reaction rate constants for gas-phase organic compounds using a structure-reactivity relationship: An update[J]. Atmospheric Environment, 1995, 29 (14): 1685-1695
- [7] Gramatica P, Pilutti P, Papa E. Validated QSAR prediction of OH tropospheric degradation of VOCs: Splitting into training-test sets and consensus modeling[J]. Journal of chemical information and computer sciences, 2004, 44 (5): 1794-1802
- [8] Öberg T. A QSAR for the hydroxyl radical reaction rate constant: Validation, domain of application, and prediction[J]. Atmospheric Environment, 2005, 39 (12): 2189-2200
- [9] Fatemi M, Baher E. Quantitative structure-property relationship modelling of the degradability rate constant of alkenes by OH radicals in atmosphere[J]. SAR and QSAR in Environmental Research, 2009, 20 (1-2): 77-90
- [10] Meylan W M, Howard P H. A review of quantitative structure-activity relationship methods for the prediction of atmospheric oxidation of organic chemicals[J]. Environmental Toxicology and Chemistry, 2003, 22: 1724-1732
- [11] Wang Y N, Chen J W, Li X H, et al. Predicting rate constants of hydroxyl radical reactions with organic pollutants: Algorithm, validation, applicability domain, and mechanistic interpretation[J]. Atmospheric Environment, 2009, 43 (5): 1131-1135
- [12] Kennard R W, Stone L A. Computer Aided Design of Experiments[J]. Technometrics, 1969, 11 (1): 137-148
- [13] Mauri A, Consonni V, Pavan M, et al. DRAGON software: An easy approach to molecular descriptor calculations[J]. Match, 2006, 56, 237-248
- [14] Todeschini R, Consonni V. Handbook of Molecular Descriptors[M]. Weinheim: Wiley-VCH, 2008
- [15] Liu H, Gramatica P. QSAR study of selective ligands for the thyroid hormone receptor beta[J]. Bioorganic and Medicinal Chemistry, 2007, 15: 5251-5261
- [16] Lei B, Ma Y, Li J, et al. Prediction of the adsorption capability onto activated carbon of a large data set of chemicals by local lazy regression method[J]. Atmospheric Environment, 2010, 44: 2954-2960
- [17] Cho S J, Hermsmeider M A. Genetic algorithm guided selection: Variable selection and subset selection[J]. Journal of Chemical Information and Computer Sciences, 2002, 42 (4): 927-936
- [18] Puzyn T, Falandysz J. Computational estimation of logarithm of *N*-octanol/air partition coefficient and subcooled vapor pressures of 75 chloronaphthalene congeners[J]. Atmospheric Environment, 2005, 39 (8): 1439-1446
- [19] Tropsha A, Gramatica P, Gombar V K. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models[J]. QSAR and Combinatorial Science, 2003, 22: 69-77
- [20] Pompe M, Veber M, Randić M, et al. Using variable and fixed topological indices for the prediction of reaction rate constants of volatile unsaturated hydrocarbons with OH radicals[J]. Molecules, 2004, 9(12): 1160-1176