

羟肟酸浮选分离选择性的定量构效关系

谭鑫¹,尚衍波²,何发钰³,刘龙利²

(1. 东北大学 资源与土木工程学院,辽宁 沈阳 110819;2. 北京矿冶研究总院 矿物加工科学与技术国家重点实验室,北京 102628;
3. 中国五矿集团总公司,北京 100010)

摘 要:采用定量构效关系方法和遗传算法对13种芳烷基羟肟酸类捕收剂的锡石-萤石浮选分离选择性指数进行了定量构效关系研究,旨在预测羟肟酸捕收剂对锡石-萤石浮选分离的选择性能.采用快速分子描述符和量子化学描述符建立了两种选择性指数的二元因子模型,选择性指数1的模型5和选择性指数2的模型6的相关系数 R^2 分别为0.868和0.796, R_{cv}^2 分别为0.779和0.590.使用留一法对两模型进行内部验证的相关系数 R^2 分别为0.785和0.615,使用测试集对两模型进行外部验证,所得两模型的平均预测绝对误差分别为0.095和1.939.指数1模型具有真实有效性和较好的预测能力,能够有效预测羟肟酸的选择性.

关 键 词:定量构效关系;羟肟酸捕收剂;遗传算法;选择性;描述符

中图分类号:TD 923 文献标志码:A 文章编号:1005-3026(2016)08-1198-05

QSAR Research on Flotation Selectivity of Hydroximic Acid Collectors

TAN Xin¹, SHANG Yan-bo², HE Fa-yu³, LIU Long-li²

(1. School of Resources & Civil Engineering, Northeastern University, Shenyang 110819, China; 2. State Key Laboratory of Mineral Processing, Beijing General Research Institute of Mining & Metallurgy, Beijing 102628, China; 3. China Minmetals Corporation, Beijing 100010, China. Corresponding author: TAN Xin, E-mail: tanxin0222@hotmail.com)

Abstract: Quantitative structure-activity relationships (QSAR) of 13 kinds of hydroximic acid collectors were studied to predict the flotation selectivity of cassiterite flotation separation based on genetic algorithm. Using fast descriptor and quantum chemical descriptor, two kinds of two-parameter models were developed, i. e., Model 5 for selectivity index 1 and Model 6 for selectivity index 2. The correlative coefficients of Model 5 and Model 6 were 0.868 and 0.796, with R_{cv}^2 of 0.779 and 0.590, respectively. For the two models, internal reliability was validated using the leave-one-out approach, with $R^2 = 0.785$ for Model 5 and $R^2 = 0.615$ for Model 6, and an external validation was also performed using test sets and the mean absolute deviations were 0.137 and 1.939, respectively. The results show that Model 5 for selectivity index 1 has better predictability, and can effectively predict flotation selectivity of hydroximic acid collectors.

Key words: QSAR; hydroximic acid collector; genetic algorithm; selectivity; descriptor

定量构效关系(quantitative structure-activity relationship, QSAR)是利用物质的结构-性质之间的关系来关联特定物化性质与分子参数的经验方法,通过建立关联模型能够有效地预测各类分子、化合物的物化性质、反应活性、生物毒性等,广泛应用于化学、生物、医药等领域.在矿物加工领域,也有浮选捕收剂 QSAR 研究的报道. Chen

等^[1]通过计算7种硫化矿捕收剂的12种物化和量化参数,建立捕收剂生物降解性的 QSAR 模型,其分析认为分子的电参数是影响捕收剂生物降解性的最重要因素;Natarajan 等^[2-3]运用拓扑化学指数、结构指数等参数,采用回归分析法建立二阶多项式回归 QSAR 模型预测17种芳香基羟肟酸对硫化铜锌矿的分离系数. Hu, Yang 等^[4-5]

采用基于线性回归算法的 QSAR 先后研究了季铵盐捕收剂结构与其对铝土矿的浮选效率和硫酸类捕收剂结构与其对硫化铜锌矿选择性之间的关系 获得了满意的结果 然而在浮选领域有关复杂氧化矿捕收剂的浮选分离选择性研究还未曾报道.

遗传算法是非数值算法中最优秀的算法之一 在许多领域的二维定量构效关系研究中得到了广泛应用 其结果精度往往优于多元线性回归分析方法^[6]. 本文采用基于遗传算法的 QSAR 方法研究 13 种芳烷基羟肟酸系列捕收剂的结构、物化、量化性质与其对锡石 – 萤石的分离选择性之间的关系 通过采用不同描述符建立羟肟酸捕收剂分离选择性的 QSAR 模型 从多角度对羟肟酸的选择性构效关系进行阐释 选择最佳预测模型将其用于新型羟肟酸选择性的预测 为羟肟酸捕收剂的筛选和设计提供参考和指导.

1 研究对象及方法

1.1 实验数据采集

训练集和测试集的选择性指数由实验数据计算得到. 具体实验条件如下 取锡石或萤石纯矿物 2 g 在 40 mL XFG5 – 35 挂槽式浮选机进行实验 , 搅拌速度 1 680 r/min. 加入 30 mL 去离子水 再加入 33 g/t 羧甲基纤维素(CMC)作抑制剂搅拌 2 min 用 HCl 或 NaOH 溶液调节至适宜 pH 值后 加入 100 g/t 捕收剂与 24 g/t 2[#]油 搅拌 3 min 后开始浮选 浮选时间 4 min. 所得泡沫产品和槽内产品分别烘干后称重 计算回收率.

1.2 结构优化及模型建立

在 Material studio7.0 软件的 DMol⁰³ 模块中采用 GGA/PW91 函数对分子进行几何构型优化和量化计算 采用 QSAR 模块计算优化后分子的分子连接性指数、几何拓扑描述符等参数. 采用遗传算法对计算描述符进行筛选并建模 参数设置如下 : 模型方程的最大因子数为 3 , 种群数 200 , 子代数 5 000 , 变异概率为 0. 1.

1.3 模型验证

采用训练集数据拟合的相关系数 R^2 评估模型拟合能力 采用 R_{cv}^2 来评估模型的预测能力 当 $R_{cv}^2 > 0. 6$ 时 模型即具有可靠的预测能力 R_{cv}^2 越接近 1 模型预测能力越强. 用遗传算法特有的 Friedman LOF(lack-of-fit)函数作为模型评价函数^[7]. 个体的 LOF 得分值越小 被入选的概率越大 同时采用内部验证留一交互验证法(leave-

one-out cross-validation , LOO-CV)和外部测试集验证考察模型的有效性和预测能力.

2 结果与讨论

2.1 羟肟酸选择性指数

浮选分离选择性体现在对两种矿物浮选回收率的差异上 在获得较高的有用矿物回收率的前提下 对脉石浮选回收率的高低可以体现捕收剂对两种矿物分离的选择性高低. 因此 本文采用锡石有效回收率大于 75% (戊羟肟酸捕收力弱 故选 65%)时的浮选条件来浮选萤石 以“ 1 – 萤石回收率 ”作为捕收剂的分离选择性指数 1 来建立羟肟酸的选择性 QSAR 模型. 因此 指数 1 越大 捕收剂的选择性越高. 采用“ 锡石回收率/萤石回收率 × 100% ”作为选择性指数 2 建立模型进行对比. 经过多组条件实验探索 在最佳的浮选分离条件下获得的测试集 13 种羟肟酸的锡石、萤石回收率及选择性指数 1 和选择性指数 2 见表 1.

表 1 羟肟酸捕收剂对锡石和萤石的浮选回收率及两种选择性指数

Table 1 Cassiterite/fluorite flotation recovery of hydroximitic acid collectors and two selectivity indices

序号	羟肟酸分子	锡石回收率/%	萤石回收率/%	选择性指数 1	选择性指数 2
1	戊羟肟酸	65. 70	3. 21	0. 968	20. 47
2	己羟肟酸	86. 33	16. 69	0. 833	5. 17
3	庚羟肟酸	80. 16	69. 62	0. 304	1. 15
4	辛羟肟酸	94. 60	58. 80	0. 412	1. 61
5	壬羟肟酸	99. 00	95. 12	0. 049	1. 04
6	苯甲羟肟酸	84. 96	25. 37	0. 746	3. 35
7	水杨羟肟酸	85. 55	19. 98	0. 800	4. 28
8	苯乙羟肟酸	77. 36	7. 62	0. 924	10. 15
9	对苯二羟肟酸	85. 28	62. 74	0. 373	1. 36
10	邻苯二羟肟酸	80. 92	46. 50	0. 535	1. 74
11	H ₂₀₅	83. 74	39. 33	0. 607	2. 13
12	环己羟肟酸	81. 07	9. 80	0. 902	8. 27
13	N – 苯甲酰基苯基羟肟(BPHA)	76. 47	26. 74	0. 733	2. 86

选择性指数 1 = 1 – 萤石回收率 ;

选择性指数 2 = $\frac{\text{锡石回收率}}{\text{萤石回收率}} \times 100\% .$

2.2 选择性指数 1——快速分子描述符模型
首先采用 QSAR 模块中的快速分子描述符 (fast descriptor)来建立模型 快速分子描述符是 MS 软件 QSAR 模块中提供的一系列能够应用于

构建分子 2D QSAR 模型的分子性能有效参数,其计算简便快捷,并且能够提供较为准确的结构性能之间的关系模型。

采用遗传算法,实验最终获得一元、二元、三元因子的 QSAR 方程分别如下:

$$Y = -0.119\ 192\ 446X_1 + 1.262\ 329\ 137 \quad (1)$$

X_1 : 可旋转键数;

$$Y = -0.181\ 021\ 511X_1 + 0.144\ 926\ 369X_{18} + 1.473\ 362\ 621, \quad (2)$$

X_1 : 可旋转键数,

X_{18} : 电拓扑状态指数 S_{sCH_3} ;

$$Y = -0.504\ 180\ 546X_4 - 0.401\ 712\ 770X_8 - 0.531\ 322\ 686X_{24} + 3.176\ 012\ 899, \quad (3)$$

X_4 : 油水分分配系数,

X_8 : Kier-Hall 分子连接指数,

X_{24} : 电拓扑状态指数 S_{ssNH} 。

式中,分子中的可旋转键数是一种分子结构描述符,其与分子中直链烷基 C 原子数成正比,而苯环、双键及环状结构的可旋转键数则较小,电拓扑状态指数通过计算分子中每个原子自身的拓扑环境以及和其他原子相互之间的交互电子作用两方面的信息获得,油水分分配系数是分子热力学描述符,是描述分子亲水-疏水性差异的重要参数,一直被认为是影响捕收剂性能的重要参数。

计算结果表明三个 QSAR 方程的 LOF 值分别为 0.039 8, 0.044 9, 0.069 6, 而模型相关系数 R^2 则依次为 0.649, 0.780, 0.852, R_{cv}^2 分别为 0.525, 0.623, 0.694。其中,一元因子方程(1)LOF 值最小,具有最大入选概率,但其相关系数 R^2 , R_{cv}^2 略低;三元因子方程(3) R_{cv}^2 系数最高,预测能力最强,但本模型训练集样品容量仅为 13,模型基函数数目应小于 3。因此二元因子方程(2)为快速描述符最适宜预测模型。

对二元方程进行主成分分析,可以发现可旋转键数与计算结果相关系数高达 0.893,其对预测结果影响更大。由于系数为负值,故可旋转键数值越小,捕收剂选择性越大。这表明捕收剂具有苯环或环状结构,对锡石-萤石的分离效果更好。捕收剂所含的烷基直碳链越长,选择性越低。另一因子,电拓扑状态指数 S_{sCH_3} 的相关系数为 -0.047 9。该电性拓扑状态指数表示分子中与三个氢原子和一个单键相连的 C 原子的加和指数。这两种描述符因子在一定程度上体现了分子的结构特征和基团的连接方式,表明捕收剂选择性受到碳链骨架影响较大。

2.3 选择性指数 1——量子化学描述符模型

早期 QSAR 研究多采用拓扑指数来建立构效关系。随着 QSAR 研究的深入开展,人们开始更加注重模型的理论性,期待通过建立合理模型,从本质上揭示物质的构效关系,找到影响物质性能和活性的最关键内在因素,为研究开发新药剂提供指导。人们开始探讨用量子化学参数建立构效关系模型。采用量子化学描述符获得最佳二元因子方程(4),其中 LOF 值为 0.043 6, R^2 为 0.786, R_{cv}^2 为 0.656。

$$Y = 3.725\ 665\ 263 X_{11} - 0.008\ 475\ 008X_{19} + 3.244\ 739\ 149, \quad (4)$$

X_{11} : N 原子上的 Mulliken 电荷,

X_{19} : 分子表面积。

从上述方程可以看出,羧基捕收剂的选择性大小与分子表面积及分子官能团经基中的 N 原子上的 Mulliken 电荷有重要联系^[8]。对方程进行主成分分析,分子表面积与模型预测结果相关系数高达 -0.940,分子表面积大小对选择性影响极大。该描述符系数为负数,表明分子表面积越大,捕收剂的选择性越小。另一方面,羧基基团中 N 原子的 Mulliken 电荷与选择性相关系数为正系数 0.246,表明 N 原子上所带负电荷越多,羧基对锡石-萤石的分离选择性越低。

2.4 选择性指数 1——综合模型

为了进一步提高模型的可靠性以及预测能力,本节综合采用快速分子和量子化学描述符建立 QSAR 模型。实验最终获得了 LOF 值为 0.0270, R^2 为 0.868, R_{cv}^2 为 0.779 的二元基因方程(5)如下:

$$Y = -0.209\ 573\ 462 X_7 + 0.422\ 930\ 582X_{85} + 1.504\ 663\ 586, \quad (5)$$

X_7 : 分子弹性指数,

X_{85} : LUMO 特征值。

模型(5)的可靠性和预测能力远高于前述模型,因此选择其作为选择性指数 1 的最佳模型。分子弹性指数是一种 Kier-Hall 分子连接指数描述符,是一种分子空间结构无限弹性程度的体现。方程中其系数为负数,表明捕收剂分子中直链 C 原子数少,或者存在苯环、分支,均有利于该指数降低,而捕收剂的选择性增加。这与实验中芳香基和环状羧基选择性整体优于长直链烷基羧基的实际情况相符。

而 LUMO 特征值是一种基于半经验分子轨道理论的量子化学描述符。捕收剂的 LUMO 特征值越低,其越容易接受来自金属原子的反馈电子;

而不同矿物表面的不同金属原子反馈能力有差异,因而有利于扩大捕收剂与金属原子相互作用的差异,捕收剂表现的选择性越高.对该两因子描述符进行主成分分析,可以发现两因子与预测值的相关系数分别为-0.872和-0.269.分子弹性指数与羟肟酸锡石-萤石的分离选择性的相关系数更大.

采用 LOO-CV 评价和检验模型,验证结果见表 2,序号后为不参与建模的分子.

表 2 羟肟酸选择性指数 1LOO-CV 的实验值及计算值
Table 2 Experimental and calculated selectivity 1 of LOO-CV approach to hydroximic acid collectors

序号	不参与分子	实验值	计算值	残差
1	戊羟肟酸	0.968	0.817	0.151
2	己羟肟酸	0.833	0.619	0.214
3	庚羟肟酸	0.304	0.479	-0.175
4	辛羟肟酸	0.412	0.333	0.079
5	壬羟肟酸	0.049	0.232	-0.183
6	苯甲羟肟酸	0.746	0.852	-0.106
7	水杨羟肟酸	0.800	0.858	-0.058
8	苯乙羟肟酸	0.924	1.081	-0.157
9	对苯二羟肟酸	0.373	0.345	0.028
10	邻苯二羟肟酸	0.535	0.511	0.024
11	H ₂₀₅	0.607	0.618	-0.011
12	环己羟肟酸	0.902	0.908	-0.006
13	BPHA	0.733	0.584	0.149

通过计算,LOO-CV 相关系数 R^2 为 0.785,平均预测绝对误差为 0.103,可见模型稳定可靠.使用测试集对模型进行外部验证,结果如表 3 所示.

表 3 测试集选择性指数 1 的实验值和计算值对比
Table 3 Experimental and calculated values of selectivity 1 for compounds in testing set

序号	测试集分子	实验值	计算值	绝对误差
1	十碳羟肟酸	0.017	-0.142	-0.159
2	2-乙基己基羟肟酸	0.335	0.517	0.182
3	对甲苯羟肟酸	0.723	0.798	0.07

模型对测试集预测的平均绝对误差为 0.137,说明该模型同时具有一定的外部预测能力.但烷基羟肟酸的误差远大于芳香基羟肟酸的误差,模型整体上不利于烷基羟肟酸的预测,尤其是模型为了体现捕收剂分子的选择性差异,十碳羟肟酸预测值甚至出现了负数,这与实际不符.除去该异常点,剩下两个测试样,模型对预测集的平均绝对误差为 0.095,所获得的预测模型具有较

好的外部预测能力.

2.5 选择性指数 2——综合模型

浮选分离选择性指数具有多种表达方式,研究采用选择性指数 2 建立 QSAR 模型.由于计算过程中发现 5C 羟肟酸与其余数据差异较大,为模型异常点,故建模过程中除去该点,参加训练的样本容量为 12.

最终获得 LOF 值为 4.688, R^2 为 0.796, R_{cv}^2 为 0.590 的最佳二元因子方程(6)如下:

$$Y = -2.143\,087\,378X_7 + 5.619\,781\,325X_{44} + 12.905\,563\,541, \tag{6}$$

X_7 : 分子弹性指数,
 X_{44} : LUMO 特征值.

该模型(6)具有与模型(5)相同的描述符因子,进一步证明这两因子对捕收剂的选择性影响较大. LOO-CV 检验模型(6)的验证结果见表 4,通过计算,相关系数 R^2 为 0.615,绝对平均误差 1.049. 外部验证模型(6)对三个测试集分子的预测结果见表 5. 与模型(5)的外部检验相似,十碳羟肟酸的误差较大,除去该异常点,剩下两模型的绝对平均误差为 1.939.

表 4 羟肟酸选择性指数 2LOO-CV 的实验值及计算值
Table 4 Experimental and calculated selectivity 2 of LOO-CV approach to hydroximic acid collectors

序号	不参与分子	实验值	预测值	残差
1	己基羟肟酸	5.173	4.848	0.324
2	庚基羟肟酸	1.151	3.227	-2.075
3	辛基羟肟酸	1.609	2.458	-0.850
4	壬基羟肟酸	1.041	-0.632	1.673
5	苯甲羟肟酸	3.349	5.935	-2.586
6	水杨羟肟酸	4.282	5.871	-1.589
7	苯乙羟肟酸	10.152	7.184	2.968
8	对苯二羟肟酸	1.359	-1.603	2.962
9	邻苯二羟肟酸	1.740	1.799	-0.059
10	H205	2.129	2.733	-0.604
11	环己羟肟酸	8.272	6.262	2.010
12	BPHA	2.860	2.847	0.012

表 5 测试集选择性指数 2 的实验值和计算值对比
Table 5 Experimental and calculated values of selectivity 2 for compounds in testing set

序号	测试集分子	实验值	计算值	绝对误差
1	十碳羟肟酸	0.976	-3.137	-4.113
2	2-乙基己基羟肟酸	1.435	3.524	2.089
3	对甲苯羟肟酸	3.159	4.948	1.789

选择性指数 2 的模型(6)的可靠性、预测能力及精确度均远小于选择性指数 1 的模型(5)。这主要是由训练集中因变量的数据分布引起的,在模型建立时,训练集的数据分布会对模型预测能力造成较大影响。对两种选择性指数因变量进行单变量分析,得到选择性指数分布如图 1 所示。

图 1 中横坐标为羟肟酸捕收剂选择性指数均分的十个区间,纵坐标为该区间选择性指数数目所占总数据的百分比。由图 1 可见,选择性指数 1 比选择性指数 2 更加接近标准正态分布,因而该指数参数更加具有统计学意义,建立的模型(5)更加合理,更能反映羟肟酸结构与选择性关系的内部规律,因此得到更好的拟合效果和预测能力。

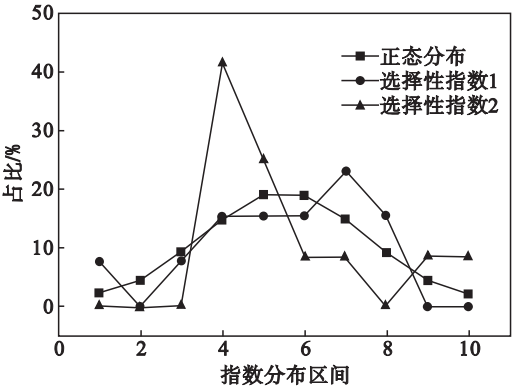


图 1 羟肟酸选择性指数 1 和 2 实验值数据分布图
Fig. 1 Distribution of experimental selectivity 1 and 2 of hydroxamic acid collectors in training set

3 结 论

1) 以选择性指数 1 为因变量的最佳模型(5) 相关系数 R^2 为 0.868 , R^2_{cv} 为 0.779。留一法交互验证相关系数 R^2 为 0.785 ,该模型真实有效;测试集外部验证所得模型的平均预测绝对误差为 0.095 ,该模型也具有较好的预测能力。

2) 采用选择性指数 2 的预测模型(6) ,模型相关系数 R^2 和预测系数 R^2_{cv} 分别为 0.796 和 0.590 ,留一法交互验证相关系数 R^2 为 0.615 ,外部验证平均绝对偏差为 1.939。该预测模型的可靠性及预测能力均差于选择性指数 1 的模型(5)。这主要是由于选择性指数 1 比指数 2 数据更接近正态分布,更具有统计意义,因此模型(5)

更合理。

3) 羟肟酸选择性指数 1 预测模型(5)真实可靠,理论阐释充分,且模型的外推性较好,可应用于羟肟酸类系列整合捕收剂的锡石-萤石分离选择性预测,能为羟肟酸系列捕收剂的筛选提供指导。

参考文献 :

[1] Chen S ,Gong W ,Mei G ,et al. Quantitative structure-biodegradability relationship of sulfide mineral flotation collector[J]. *International Journal of Mineral Processing* , 2011 ,101(1/2/3/4) :112 -115.

[2] Natarajan R , Nirdosh I. Quantitative structure-activity relationship (QSAR) approach for the selection of chelating mineral collectors[J]. *Minerals Engineering* ,2008 ,21(12/13/14) :1038 -1043.

[3] Natarajan R , Nirdosh I. Application of topochemical , topostructural physicochemical and geometrical parameters to model the flotation efficiencies of *N*-arylhydroxamic acids [J]. *International Journal of Mineral Processing* ,2003 ,71 (1/2/3/4) :113 -129.

[4] Hu Y ,Chen P ,Sun W. Study on quantitative structure-activity relationship of quaternary ammonium salt collectors for bauxite reverse flotation[J]. *Minerals Engineering* 2012 , 26 24 -33.

[5] Yang F ,Sun W ,Hu Y. QSAR analysis of selectivity in flotation of chalcopyrite from pyrite for xanthate derivatives : xanthogen formates and thionocarbamates [J]. *Minerals Engineering* 2012 ,39 :140 -148.

[6] 闫超群 ,万辉 ,管国锋. 1,3-二取代咪唑类离子液体熔点预测 [J]. 物理化学学报 2008 ,24(12) :2198 -2202.

(Yan Chao-qun ,Wan Hui ,Guan Guo-feng. Prediction of melting points for 1,3-disubstituent imidazolium ionic liquids [J]. *Acta Physico-Chimica Sinica* , 2008 , 24 (12) : 2198 -2202.)

[7] 潘咏梅 ,计明娟. 基于遗传算法的 PTP1B 抑制剂的二维定量构效关系研究 [J]. 物理化学学报 2003 ,19(8) :695 -700.

(Pan Yong-mei ,Ji Ming-juan. Application of genetic algorithms on 2D-QSAR analysis of benzofuran and benzothiophene biphenyls as PTP1B inhibitors [J]. *Acta Physico-Chimica Sinica* 2003 ,19(8) :695 -700.)

[8] Sreenivas T ,Padmanabhan N P H. Surface chemistry and flotation of cassiterite with alkyl hydroxamates [J]. *Colloids and Surfaces A :Physicochemical and Engineering Aspects* , 2002 205 (1/2) :47 -59.