Minerals Engineering 39 (2012) 140–148

Contents lists available at SciVerse ScienceDirect

Minerals Engineering

journal homepage: www.elsevier.com/locate/mineng

QSAR analysis of selectivity in flotation of chalcopyrite from pyrite for xanthate derivatives: Xanthogen formates and thionocarbamates

Fan Yang, Wei Sun *, Yuehua Hu

School of Resources and Bioengineering, Central South University, Changsha 410083, Hunan, China

ARTICLE INFO

Article history:
Received 6 March 2012
Accepted 1 June 2012
Available online 4 October 2012

Keywords:
Sulfide ores
Flotation collectors
Modeling

ABSTRACT

A quantitative structure–activity relationship (QSAR) study of selectivity in flotation of chalcopyrite from pyrite for xanthates and xanthate derivatives was performed. The genetic function approximation (GFA) algorithm was applied to select the descriptors and to generate the correlation models. Three combinations of topological, structural, physicochemical, spatial and electronic descriptors were applied. Six QSAR models with acceptable R^2 correlation coefficients and Q^2 cross-validation correlation coefficients ($R^2 > 0.75$, $Q^2 > 0.6$) were developed. The models show that shadow area fraction *YZ*, LUMO density of carbonyl oxygen atom and Hirshfeld Fukui indices (+) of thiocarbonyl sulfur atom are the three most significant variables for the selectivity. The descriptor shadow area fraction *YZ* revealed that the planar structure is important for the selectivity. The descriptors LUMO density of carbonyl oxygen atom and Hirshfeld Fukui indices (+) of thiocarbonyl sulfur atom demonstrated LUMO distributions over carbonyl oxygen atom should be large for favorable selectivity. This study concluded that selectivity of xanthates and xanthate derivatives is determined by the strength of the dative bonding between copper atom on the chalcopyrite surface and carbonyl oxygen atom of these collectors and the stability of the six-member ring formed as a result.

1. Introduction

Flotation separation chalcopyrite from pyrite has always been a problem as it lacks effectively selective collectors. Traditionally, xanthates are used as collector though a large amount of lime is required to obtain a highly alkaline condition to achieve the separation (Liu, 2004). Addition of lime, however, may depress chalcopyrite and other coexisting precious metal minerals and cause the erosion of the equipments (Li and Sun, 2000). To overcome this problem, several series of xanthate derivatives such as xanthogen formats and thionocarbamates were invented. They work well in the natural circumstances of copper sulfide ores, i.e. at pH of about five sometimes, with high selectivity against pyrite (Crozier, 1978; Aplan and Chander, 1987; Harris and Fishback, 1954; Seryakova et al., 1975; Woods and Hope, 1999). Especially, in 1980s, based on N-alkyl thionocarbamate N-alkoxycarbonyl thionocarbamates were invented (Fu and Wang, 1987), the selectivity against pyrite in neutral circumstance of which is much better than the former ones.

All along, most of researches concerning xanthates and its derivatives have been focused on their reaction mechanism with sulfide minerals (Fairthorne et al., 1997, 1998; Nagaraj and Brinen, 2001; Hope et al., 2005, 2007; Liu et al., 2006), while few tried to reveal the relation between their structures and selectivity. Ackerman et al. (2000) evaluated the influence of different substituent structures on the flotation of copper sulfide minerals and pyrite and demonstrated a relation between them does exist. Liu et al. (2008) found the relation of efficiencies of N-substituent thionocarbamates with the properties such as energies and compositions of frontier molecular orbital and atomic charges and explained why IBECTC and IBACTC have excellent collecting power for copper sulfide minerals and selectivity against iron sulfide minerals on the basis of the order of electron-donating ability and feedback-electron-accepting ability. In these work, the relations obtained are mostly qualitative and without statistical validations. This study was to systematically investigate the relation between molecular structures and selectivity using quantitative structure–activity relationship (QSAR) modeling technique.

A quantitative structure activity relationship (QSAR) is a method that correlates an activity of a set of compounds quantitatively to chemicals descriptors (structure or property) of those compounds, which are generally obtained from experiments and quantum chemical calculations. The exact form of the relationship between structure and activity can be determined using a variety of statistical methods and computed molecular descriptors. QSAR has the objective of prediction but maintaining a relationship to mechanistic interpretation. Recent years, quantitative structure–activity

* Corresponding author. Tel.: +86 371 88830482.
E-mail address: sunmenghu@mail.csu.edu.cn (W. Sun).

0892-6875/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.mineng.2012.06.001

relationship (QSAR) modeling has been applied extensively in pharmaceutical chemistry and predictive toxicology (Yangali-Quintanilla et al., 2010; Wang et al., 2011; Mandal and Roy, 2009; Iyer et al., 2007; Moorthy et al., 2011). There are also several applications in flotation. Natarajan and Nirdosh (2008) selected 10N-arylhydroxamic acids from a virtual database of 3800 compounds with the method of molecular similarity clustering and synthesized and tested them as collectors for sphalerite. Hu et al., (2012) carried out a QSAR research on the quaternary ammonium salt collectors for bauxite reverse flotation and established a robust model.

The purpose of this work was to establish QSAR models of the selectivity in flotation of chalcopyrite from pyrite for xanthates and xanthate derivatives, including xanthogen formates and thionocarbamates, and to give the models mechanistic interpretations to reveal the favorable structural characteristics that enhance their selectivity.

2. Methods

2.1. Data set

In this study the selectivity of 16 sample molecules including xanthates, xanthogen formates and thionocarbamates was collected from Fu et al. (1989). The values are reported in the form of I_{cu}, which was defined and calculated in accordance with Eq. (1):

$$I_{Cu} = \frac{(100 - \%\ \text{Pyrite recovered})}{(100 - \%\ \text{Copper recovered})^2} \quad (1)$$

The names and structures of the 16 collectors were listed in Table 1 with their I_{cu} values.

Generally, the data set is divided into a training set to develop models and a test set to externally validate the constructed QSAR models. As there are no such large data sets available, the set of the 16 molecules were wholly used as the training set and as the test set flotation results of a new set of 6 collectors were obtained. The names and structures of the 6 collectors were listed in Table 2.

The individual-mineral flotation tests were carried out in the SFG micro flotation cell (40 ml). Chalcopyrite and pyrite used in this test were obtained from Donggua Hill copper mine, Anhui, China. They were crushed with hammer wrapped up with cloth to avoid contaminating the minerals before hand picking the pure minerals, which, afterwards, were milled by porcelain ball mill respectively. And the part with size of $-150\sim+200$ mesh of each was used in flotation tests. Measured by X-ray Diffraction and chemical elemental analysis, the purity of chalcopyrite is up to 85.32% containing trace amount of silica and fluorite while the purity of pyrite is as high as 93.56% and no other impurities are tested. First, 2 g minerals were placed in the cell and added 30 ml distilled water. Then, the mixture was agitated for 1 min. Collectors and pine oil were added successively at the interval of 3 min and every time the mixture was blended for 3 min. Last, the froth was collected for 3 min. All tests were run at collector concentration of 1E−4 mol/l and frother concentration of 35 mg/l at natural pH. The flow sheet of flotation is listed in Fig. 1.

Obviously, the two data sets were obtained in two different experimental conditions. And their absolute values are not comparable. So the external validation cannot be performed quantitatively in the general way. Actually, selectivity is different from many other physical and chemical properties. Flotation results that selectivity is obtained from depend largely on the artificial operations and they vary from person to person. Hence, even the values of the predicted selectivity with externally validated models are meaningless for the compounds investigated. The models of selectivity only make sense qualitatively. Therefore, it is acceptable to externally validate the models qualitatively with a data set from a different source for the training set. And in this case, it was performed through comparing the order of selectivity of the test set predicted with QSAR models constructed with the one obtained from flotation tests above.

2.2. Quantum chemical calculations

All molecules were built using the Material Studio 4.0 software. The geometry optimizations and single point calculations were both carried out with the Dmol3 module in Material Studio with Generalized Gradient Approximation (GGA) functional PBE (Perdew et al., 1996) and the basis set DND. Conformers were evaluated to make sure the optimized structures were the global-minimum structures. Xanthates have no stable conformers and the local-minimum structure is global-minimum. For xanthogen formates and thionocarbamates, three most stable conformers (Tobón et al., 2009) were all optimized and evaluated with the single point energies to identify the global-minimum conformations. Finally, vibrational analysis was performed to confirm the structures obtained.

In this study, QSAR models were developed using topological, structural, physicochemical, spatial and electronic descriptors. These descriptors were divided into three categories using different combinations of descriptors. The first category was a combination of topological and structural descriptors while the second category included structural, physicochemical, spatial and electronic descriptors. All these parameters were clubbed together in the third category. All descriptors were calculated using Material Studio 4.0 software and are listed categorically in Table 3. The meanings of all descriptors are available in Material Studio 4.0 Tutorials (Accelrys, 2005).

2.3. QSAR analysis

The correlation coefficients for all pair of descriptor variables used in the models were evaluated to identify highly correlated descriptors in order to detect redundancy in the data set. Hence, some highly correlated ($r \geqslant 0.9$) and constant descriptors were removed from the further consideration. Furthermore, the models with the descriptors that had cross-correlation coefficient of more than 0.6 were also removed from consideration.

Typically, a ratio of five or more measured values for every descriptor should be sought in order to prevent overfitting (Walker et al., 2003). Hence, only two and three-parameter models were considered in this study. For the development of equations, GFA regression analysis was utilized with Material studio 4.0 software.

The genetic function approximation (GFA) algorithm is a useful technique for searching in a large parameter space when the data set is small. This method provides multiple models that are created by evolving random initial models using different descriptors. Models are improved by performing a crossover operation to recombine terms providing better scoring models. The GFA algorithm approach has a number of important advantages over other techniques such as, it builds multiple models rather than a single model, it automatically selects which features are to be used in the models, and it can build models using either a linear relation or higher order polynomial, splines, and Gaussians. (Ponnurengam et al., 2006).

The application of GFA to QSAR/QSPR studies has been successfully used in a wide number of QSPR/QSAR research (Rogers and Hopfinger, 1994; Shi et al., 1999; Hou et al., 2009; Couling et al., 2006; Fan et al., 2001). This method combines Friedman's multivariable adaptive regression splines MARS (Friedman, 1991) and Holland's genetic algorithm (Holland, 1975). This method also used the Friedman's lack-of-fit (LOF) along with the common R-squared, cross validated R-squared, and *F*-value to evaluate the significance

Table 1
Selectivity of the 16 sample molecules obtained from Fu et al. (1987).

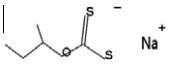
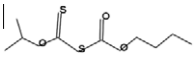
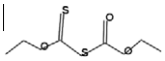
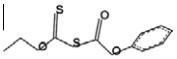
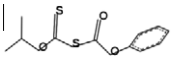
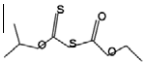
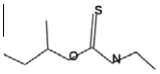
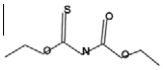
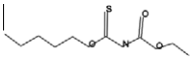
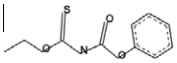
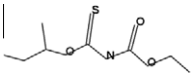
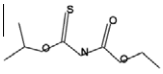
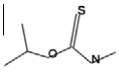
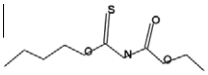
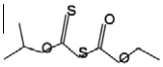
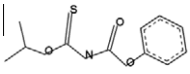
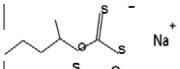
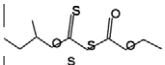
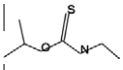
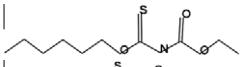
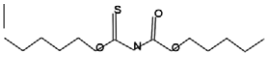
Name	Abbreviation	Structure	I_{cu} (Obsd)
Sodium isobutyl xanthate	IBX		0.02
Isopropyl xanthogen butyl formate	IPB		0.28
Ethyl xanthogen ethyl formate	EXE		0.25
Ethyl xanthogen phenyl formate	EXPhenyl		0.20
Isopropyl xanthogen phenyl formate	IPPhenyl		0.27
Isopropyl xanthogen ethyl formate	IPE		0.18
N-Ethyl-O-isobutyl thionocarbamate	IBETC		0.10
N-Ethoxy carbonyl-O-ethyl thionocarbamate	EECTC		0.22
N-Ethoxycarbonyl-O-amyl thionocarbamate	AECTC		0.53
N-Phenoxycarbonyl-O-ethyl thionocarbamate	EPCTC		0.29
N-Ethoxycarbonyl-O-isobutyl thionocarbamate	IBECTC		0.21
N-Ethoxycarbonyl-O-isopropyl thionocarbamate	IPECTC		0.39
N-methyl-O-isopropyl thionocarbamate	IPMTC		0.06
N-Ethoxycarbonyl-O-n-butyl thionocarbamate	NBECTC		0.36
Isopropyl xanthogen ethyl formate	IPE		0.30
N-Phenoxycarbonyl-O-isopropyl thionocarbamate	IPPCTC		0.39

Table 2
Names and structures of the 6 molecules of the test set.

Name	Abbreviation	Structure
Sodium isoamyl xanthate	IAX	
Isobutyl xanthogen ethyl formate	IBXF	
N-Ethyl-O-isopropyl thionocarbamate	IPETC	
N-Ethoxycarbonyl-O-hexyl thionocarbamate	HECTC	
N-Amoxycarbonyl-O-amyl thionocarbamate	AACTC	

of the regression in the QSPR/QSAR model. In Material Studio, the LOF measures the fitness of a GFA model during the evolution process and is obtained using a slight variation of the original Friedman formula as shown in Eq. (2).

$$LOF = \frac{\mathrm{SSE}}{M\left[1 - \lambda\left(\frac{c+dp}{M}\right)\right]^2} \quad (2)$$

where SSE is the sum of square errors, c is the number of terms in the model, other than the constant term, d is a scaled smoothing parameter, p is the total number of descriptors contained in all model terms (ignoring the constant term), M is the number of samples in the training set, λ is a safety factor, with the value of 0.99, to ensure that the denominator of the expression can never become zero.

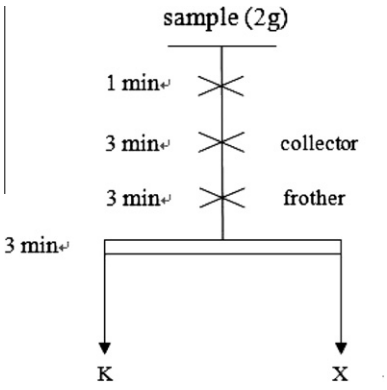


Fig. 1. Flow sheet of flotation tests.

The scaled smoothing parameter, d, is related to the user-defined smoothness parameter, R, by the expression shown in Eq. (3), where c_{max} is the maximum equation length.

$$d = \alpha\left(\frac{M - c_{max}}{c_{max}}\right) \tag{3}$$

The GFA algorithm was applied to the training set, using the R-squared value as a scoring function and a LOF smoothness parameter equal to 0.5. The program was run repeatedly varying the length and order of the equations. The set of equations returned were evaluated using the following parameters: (a) Friedman LOF measure, (b) R-squared, (c) cross-validated R-squared, and (d) F-value. These values were calculated with the statistical models in the Material Studio software.

3. Results and discussion

3.1. QSAR using topological and structural descriptors

3.1.1. Two-variable model

The following equation describes a two-variable model that has the highest r^2 coefficient among all the two-variable models generated and selected by GFA procedure using topological structural descriptors.

$$Y = -1.895761107\ X13 + 0.324164845\ X15 + 0.773528810 \tag{4}$$

$$(n = 16, R^2 = 0.767, R^2_{adj} = 0.732, Q^2 = 0.666, \text{Friedman LOF} = 0.019, F = 21.456)$$

where $X13$ = shadow area fraction YZ, $X15$ = shadow ratio.

This equation could explain and predict 73.2% and 66.6% of the variance of the selectivity, respectively. The Q^2 value resulting from the cross-validations was similar to the r^2 value, which suggests that the models are relatively stable (Sabljic, 2001; Hall et al., 2002b). The selectivity increases with the increase of shadow ratio and the decrease of shadow area fraction YZ. And shadow area fraction YZ has the major contribution to the selectivity. Such two descriptors are both molecular shadow indices, which help to characterize the shape of molecules (Rohrbaugh and Jurs, 1987). Shadow area fraction YZ is defined as the fraction of area of molecular shadows in the YZ plane over the area of the enclosing rectangle. Shadow ratio is the ratio of largest to smallest dimension in X, Y, Z directions. The negative coefficient of shadow area fraction YZ and the positive of shadow ratio imply that the planar shape of the molecules favors the selectivity. This is consistent with that O-isobutyl-N-ethoxycarbonyl-thionocarbamate (IBECTC) could react chemically with copper atom on the surface of copper sulfide minerals to form a stable six-member complex with —C(=O)—N—C(=S)— group, thereby making IBECTC more powerful than O-isobutyl-N-ethyl-thionocarbamate (IBETC) (Liu et al., 2008).

3.1.2. Three-variable model

The following equation describes a three-variable model that has the highest r^2 coefficient among all the 3-variable models generated and selected by GFA procedure using topological structural descriptors.

$$Y = -1.619495019\ X12 + 0.265202024\ X14 - 0.097914746\ X19 + 0.701296898 \tag{5}$$

$$(n = 16, R^2 = 0.806, R^2_{adj} = 0.758, Q^2 = 0.675, \text{Friedman LOF} = 0.018, F = 16.666)$$

where $X12$ = shadow area fraction YZ, $X14$ = shadow ratio, $X19$ = E-state S_dssC.

Better explained variance ($R^2_{adj} = 0.758$) and predictivity (Q^2 = 0.675) were shown by this three-variable model in comparison to Eq. (4). This equation is similar to the two-variable model except including E-state S_dssC. Like Eq. (4), shadow area fraction YZ show negative effects on selectivity and shadow ratio show

Table 3
Categorical list of descriptors used in the development of QSAR models.

Category of descriptors	Name of the descriptors
Structural	MW, Rotlbonds, Hbondacceptor, Hbonddonor, Chiralcenters, Atom count, Element count, Connolly surface area, Connolly surface occupied, solvent surface area, solvent surface occupied, Cavity volume, Molecular flexibility, information content, Bond information, Complementary information content, Structural information content, Atomic composition, Molecular area vdw, Molecular volume vdw, Molecular density, Principal moments
Physicochemical	AlogP, AlogP98, log P, MR, MolRef
Topological	Jx, Balaban index, Wiener index, Zagreb index, Kappa indices, Subgraph counts, Chi indices, Edge adjacency, Edge distance, Vertex indices, E-state indices
Electronic	Atomic charge, Dielectric energy, salvation energy, surface energy, atomic Fukui indices, total energy, HOMO energy, LUMO energy, LUMO–HOMO energy, dipole, Electron affinity, Ionization Potential, atomic HOMO density, atomic LUMO density, atomic Electrophilic frontier density, atomic Nucleophilic frontier density, atomic Radical frontier density, atomic Electrophilic Superdelocalizability, atomic Nucleophilic Superdelocalizability, atomic Radical Superdelocalizability, atomic IGLO chemical shift, atomic LORG chemical shift
Spatial	Radius of gyration, Jurs descriptors, Ellipsoidal volume, Shadow area XY plane, Shadow area YZ plane, Shadow area XZ plane, Shadow area fraction XY, Shadow area fraction YZ, Shadow area fraction XZ, Shadow length LX, Shadow length LY, Shadow length LZ, Shadow ratio, Dipole moment spatial

positive effects in Eq. (5). E-state S_dssC (E-state keys sums) descriptor was the sum of electrotopological state for carbon with two single bonds and one double bond and it has a negative index, which means the fragment >C=decrease the selectivity. However, compared to shadow area fraction *YZ*, the contribution of E-state S_dssC to the selectivity is much smaller and could be neglected.

3.2. QSAR using structural, physicochemical, spatial and electronic descriptor

3.2.1. Two-variable model

Eq. (6) lists the two-variable model with the highest r^2 coefficient among all the two-variable models generated and selected by GFA procedure using structural, physicochemical, spatial and electronic descriptors.

$$Y = 0.041094175\ X11 + 2.715760376\ X19 - 0.132618779 \quad (6)$$

$$(n = 16, R^2 = 0.757, R_a^2 dj = 0.720, Q^2 = 0.631, \text{Friedman LOF} = 0.006, F = 20.277)$$

where *X*11 = Molecular flexibility, *X*19 = LUMO (lowest unoccupied molecular orbital) density of carbonyl oxygen atom.

This equation could explain and predict 72% and 63.1% of the variance of the selectivity, respectively. It shows similar R^2 and Q^2 values for the training set to Eq. (4), while it gives a magnitude less Friedman LOF value than Eq. (4). In Eq. (6), molecular flexibility and LUMO density of carbonyl oxygen both have positive influence on the selectivity. And LUMO density of carbonyl oxygen has the major contribution. LUMO density describes the distributions of LUMO over atoms within the molecule (Singh et al., 2011). And the LUMO density of carbonyl oxygen is a measure of relative reactivity of the LUMO at the carbonyl oxygen atom. Liu et al. (2008) proposed that besides IBECTC offering its HOMO (highest occupied molecular orbital) electrons of thiocarbonyl sulfur atom to copper atom on the surface of chalcopyrite forming σ-bond copper atoms also donate its d-orbital electrons to the LUMO of IBECTC forming dative-bond and as a result form a stable six-member ring with IBECTC, which enhances the collectors' selectivity. And Eq. (6) suggests that the carbonyl oxygen is critical to forming the dative-bond and that the more it contributes to LUMO the more selective the collectors are Molecular flexibility also has positive effect on selectivity, although its effect is negligible compared to the one of LUMO density of carbonyl oxygen atom. And its positive effect could be explained as molecular flexibility is favorable to collectors forming the six-member ring with surface copper atoms.

3.2.2. Three-variable model

Eq. (7) lists the three-variable model with the highest r^2 coefficient among all the three-variable models generated and selected by GFA procedure using structural, physicochemical, spatial and electronic descriptor.

$$Y = 0.001010463\ X1 + 0.032020727\ X11 + 2.319610547\ X19 - 0.247321190 \quad (7)$$

$$(n = 16, R^2 = 0.780, R_a^2 dj = 0.725, Q^2 = 0.611, \text{Friedman LOF} = 0.021, F = 20.277)$$

where *X*1 = Connolly surface occupied volume, *X*11 = Molecular flexibility, *X*19 = LUMO density of carbonyl oxygen atom.

This model is a little worse than Eq. (6) with lower Q^2 (Q^2 = 0.611) and higher Friedman LOF (LOF = 0.021) values. Besides molecular flexibility and LUMO density of carbonyl oxygen atom, Connolly surface occupied volume was included in this equation and its coefficient is positive too. Similar to Eq. (6), LUMO density of carbonyl oxygen atom has the major contribution. While, the contribution of Connolly surface occupied volume to selectivity is even a magnitude less than molecular flexibility. The positive influence of Connolly surface occupied volume suggests that the size of the molecule increases the selectivity (Connolly, 1984).

3.3. QSAR using combined (topological, structural, physicochemical, spatial and electronic) set of descriptors

3.3.1. Two-variable model

Using combined set of topological, structural, physicochemical, spatial and electronic descriptors the following equation was the two-variable model with the highest r^2 coefficient among all the two-variable models selected and generated by GFA procedure.

$$Y = 2.450868414\ X17 + 0.042794062\ X20 - 0.155704610 \quad (8)$$

$$(n = 16, R^2 = 0.771, R_{adj}^2 = 0.736, Q^2 = 0.655, \text{Friedman LOF} = 0.019, F = 21.883)$$

where *X*17 = LUMO density of carbonyl oxygen atom, *X*20 = Kappa-3.

The ability of Eq. (8) to explain ($R_{adj}^2 = 0.736$) and predict (Q^2 = 0.655) was better than the Eq. (6). Similar to Eq. (6), LUMO density of the carbonyl oxygen atom is the major contribution to the selectivity. The other descriptor in this equation, Kappa-3, is the third-order kier shape indices. And its positive coefficient suggests the level of branching at the center of the molecule increase the selectivity (Adenot et al., 1999).

3.3.2. Three-variable model

$$Y = -4.263367670\ X3 + 0.065652445\ X20 - 0.322309832\ X32 + 1.195719997 \quad (9)$$

$$(n = 16, R^2 = 0.813, R_{adj}^2 = 0.767, Q^2 = 0.658, \text{Friedman LOF} = 0.018, F = 17.419)$$

where *X*3 = Hirshfeld Fukui indices (+) of thiocarbonyl sulfur atom, *X*20 = Kappa-3, *X*32 = Chi3 cluster valence modified.

In this equation, values of $R_a^2 dj = 0.767$ were better than Eq. (5) and (7). However, the Q^2 value (Q^2 = 0.658) was less than that of Eq. (5) and more than that of Eq. (7). Hirshfeld Fukui indices (+) of thiocarbonyl sulfur atom and Chi3 cluster valence modified are responsible for lowering the selectivity, while Kappa-3 has positive impact on the selectivity. Fukui functions introduced by Parr and Yang (1984) are partial derivatives of electron density with respect to the number of electrons. The atomic Fukui indices (+) indicate reactivity toward nucleophiles and in some cases they are approximated with LUMO densities (Cioslowski et al., 1993). Obviously, the effect of the LUMO density of carbonyl oxygen atom on selectivity, as mentioned in Eq. (6) and (7), is opposite to the effect of Fukui indices (+) of thiocarbonyl sulfur atom in this equation. This is not contradictory but consistent. Because quantum chemical calculations indicate that LUMO of xanthogen formates and thionocarbamates is mainly consisted of p-orbital of the thiocarbonyl sulfur atom and the carbonyl oxygen atom and that the LUMO density of these two atoms generally changes in opposite ways. That is when the Fukui indices (+) of thiocarbonyl sulfur atom of a molecule increases LUMO densities of the carbonyl oxygen atom decreases and consequently the selectivity decreases. And this equation proves dative-bonding of these collectors with copper atoms on the chalcopyrite surface critical to their selectivity again. Chi3 cluster valence modified indices measure the valence molec-

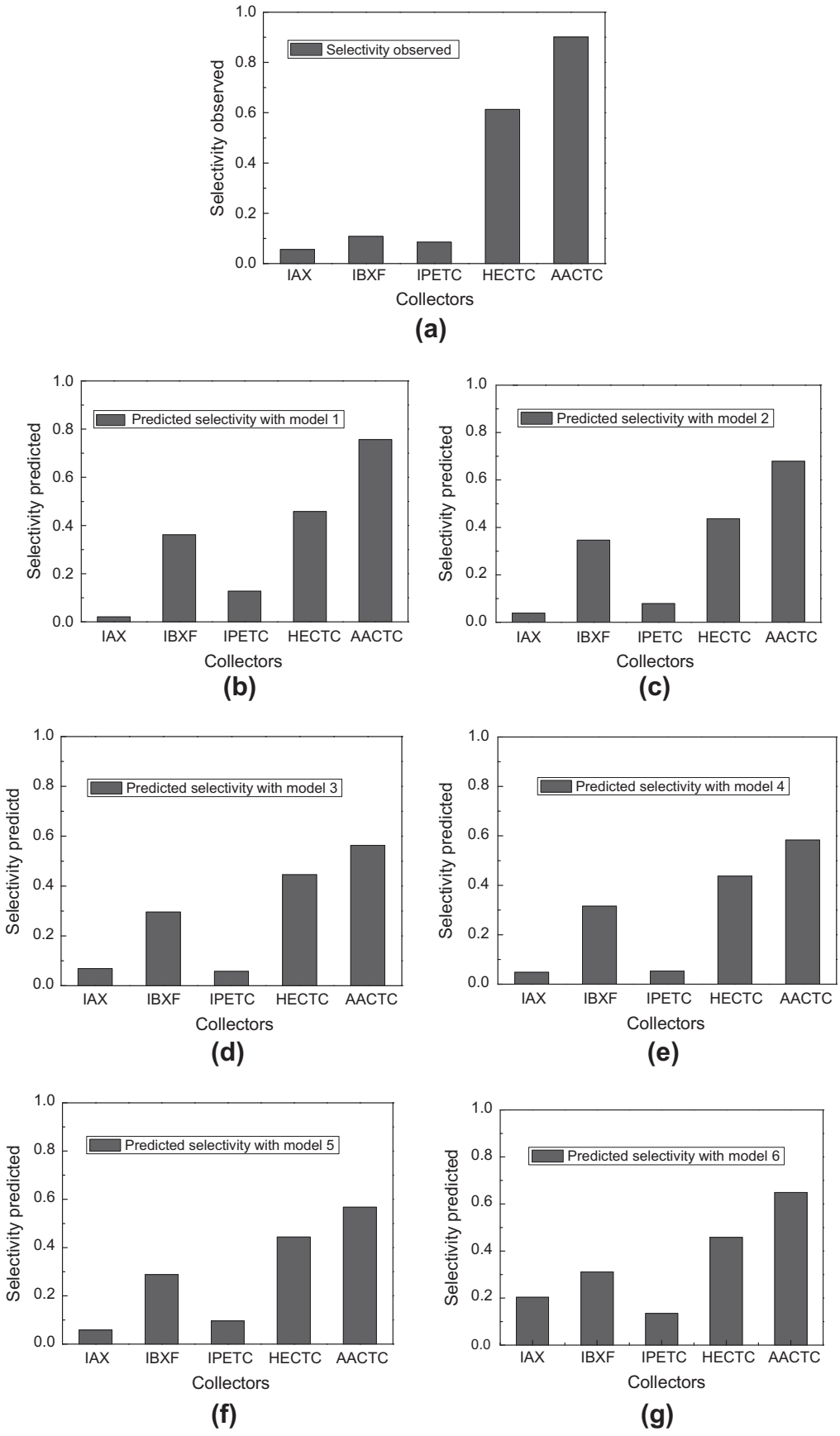


Fig. 2. Histograms of observed and predicted with the six models of the external set compounds. (a) Observed selectivity, (b) predicted selectivity with Eq. (4), (c) predicted selectivity with Eq. (5), (d) predicted selectivity with Eq. (6) and (e) predicted selectivity with Eq. (7).

ular connectivity index of 3rd order cluster and demonstrate that the presence of electronegative atoms (nitrogen, oxygen and sulfur) instead of carbon atoms in the molecules increases the molecular connectivity index value (Kier and Hall, 1976). And in this equation, its negative coefficient could suggest that the replacement of single bonded sulfur with nitrogen in the molecule increases the selectivity as it decreases Chi3 cluster valence modified index because nitrogen is less negative than sulfur.

3.4. Tests on external predictivity

A data set of selectivity collected from flotation tests specified above was considered to qualitatively validate the external predictivity of the six models obtained above. The set is comprised of five compounds including xanthates, xanthogen formates and thionocarbamates, which all appeared in the training set. So it could be implied that these compounds fall within the applicability domain of the models obtained above. Fig. 2 shows the observed and predicted selectivity of the external set compounds.

Fig. 2a shows that the order of selectivity observed is AACTC > AECTC > IBP > IPETC > IAX, and that the selectivity of IBP, IPETC and IAX is close and much lower than that of AACTC and AECTC. The predicted selectivity with the six equations generally follows this trend. However, it is worth to note that all of these six models overestimate the selectivity of IBP. And the possible reason is that IBP is characterized by an unusual structural feature for the training set, which is a double bond exists in its substituted alkyl groups. Therefore, IBP could be regarded as an outlier of the applicability domain of the six models.

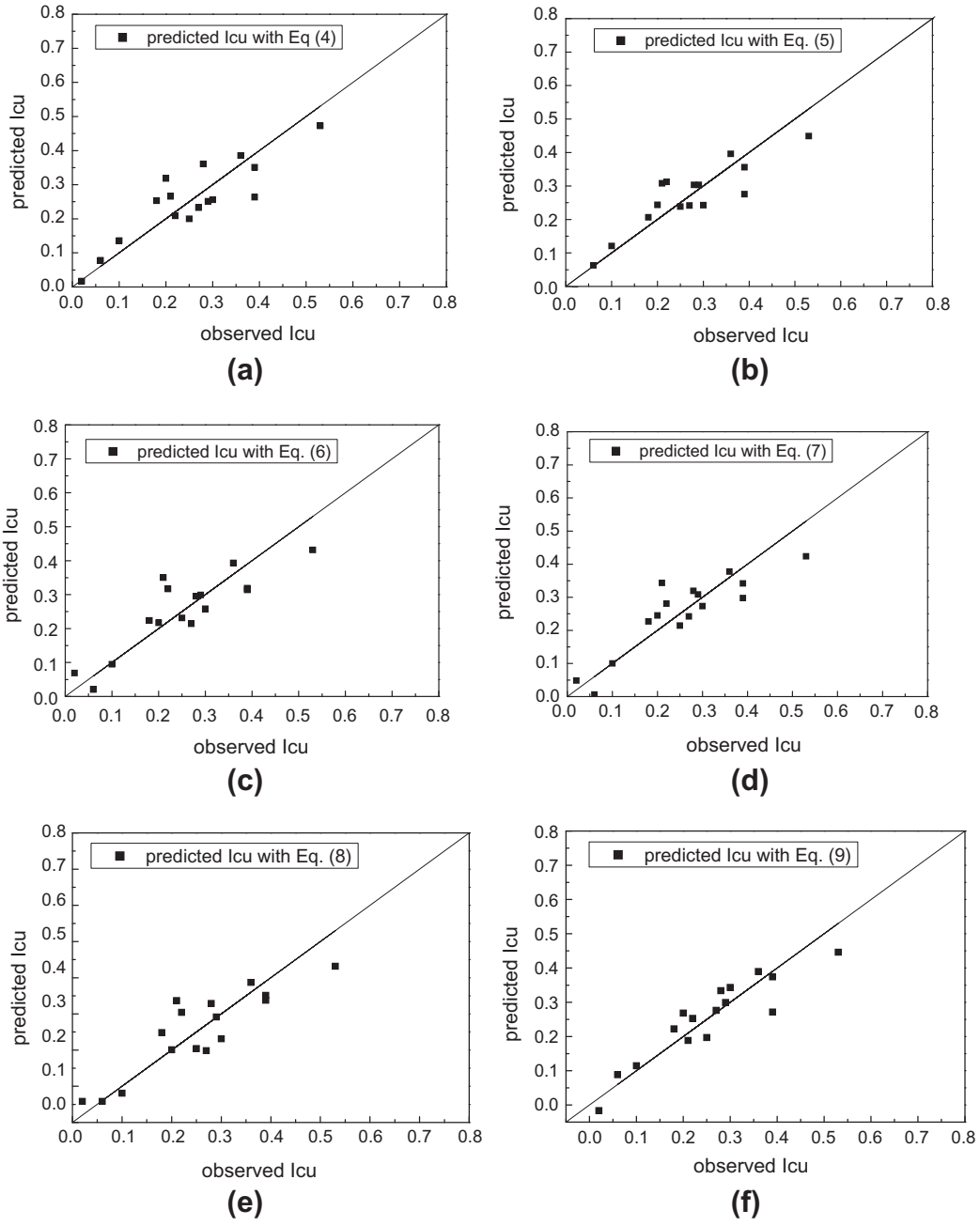


Fig. 3. A plot of observed and predicted I_{cu} values for the training set. (a) Predicted selectivity with Eq. (4), (b) predicted selectivity with Eq. (5), (c) predicted selectivity with Eq. (6), (d) predicted selectivity with Eq. (7), (e) predicted selectivity with Eq. (8) and (f) predicted selectivity with Eq. (9).

3.5. Overview

Genetic function approximation (GFA) has been applied to model xanthates, xanthogen formates and thionocarbamates as collectors of chalcopyrite using different combinations of topological, structural, physicochemical, electronic and spatial descriptors. All equations showed acceptable internally cross validated squared correlation coefficient ($Q^2 > 0.6$) and the best equation ($Q^2 = 0.675$) was developed using topological and structural descriptors with three variables. The predicting power of the equations were proved qualitatively acceptable through an externally validation with a data set from flotation tests. Scatter plots of observed versus predicted values of the training set obtained from the six models are given in Fig. 3.

Shadow area fraction *YZ*, LUMO density of carbonyl oxygen atom and Hirshfeld Fukui indices (+) of thiocarbonyl sulfur atom is the three most significant of descriptors in the six models. It implies that atomic LUMO properties of these collectors explained the largest variance in their selectivity among various structural characteristics. This seems a little surprising because in the case of explaining the difference of xanthates and xanthate derivatives in selectivity are usually used HOMO properties such as HOMO energy (Liu et al., 2006 and 2008). Actually, HOMO is just an indication of collectors' collecting power. High HOMO energy means strong collecting power but not necessarily equals with low selectivity. And this QSAR analysis revealed that the ability of accepting d-orbital electrons of surface copper atom, which is measured by atomic LUMO density and atomic Fukui indices (+) in this case, is more critical for these collectors' selectivity than the ability of donating their HOMO electrons. It can be explained in such two aspects: (1) electrons in occupied d-orbital of copper atom on chalcopyrite surface are more easily transferring to unoccupied orbital of collector than that of iron atom on pyrite as the chalcopyrite band gap was 0.6 eV and the values of pyrite band gap were reported 0.95 eV. (2) Together with the σ bonding through donating HOMO electrons to surface copper atom, the dative Π bonding through accepting d-orbital electrons of copper atom results in a stable six-ring bonding between collectors and chalcopyrite.

Other descriptors appear in the six models including shadow ratio, E-state_dssC, molecular flexibility, Connolly surface occupied volume, Kappa-3 and Chi3 cluster valence modified. These descriptors describe the structural characteristics in favor of forming stable six-member rings and as a result enhancing the selectivity.

4. Conclusion

The study on the QSAR analysis of the selectivity in flotation of chalcopyrite from pyrite for xanthates and xanthate derivatives were performed with theoretical descriptors. The QSAR analysis results obtained from the study provide significant statistical parameters and the validation studies confirm that the six models are statistically reliable and robust. The six QSAR models identified that Shadow area fraction *YZ*, LUMO density of carbonyl oxygen atom and Hirshfeld Fukui indices (+) of thiocarbonyl sulfur atom are the three descriptors that influence the selectivity most. It implies that the selectivity of xanthates and xanthate derivatives is determined by the strength of the dative bonding between copper atoms on the chalcopyrite surface and carbonyl oxygen atom of these collectors and the stability of the six-member ring formed as a result. And this explained why the selectivity xanthates and xanthate derivatives generally follows the trend: ethoxycarbonyl thionocarbamate ≫ xanthogen formates > N-alkyl-thionocarbamates > xanthates.

On the other hand, the six models could be used together to assess the selectivity in flotation of chalcopyrite from pyrite for new candidate molecules in the series of compounds we studied.

Acknowledgements

The authors would like to thank the Supporting Program of the "Eleventh Five-year Plan"for Sci & Tech Research of China (2006BAB02A06) for financial support.

References

Ackerman, P.K., Harris, G.H., Klimpel, R.R., Aplan, F.F., 2000. Use of xanthogen formates as collectors in the flotation of copper sulfides and pyrite. Int. J. Miner. Process 58, 1–13.

Adenot, M., Menthière, C.S., Chavanieu, A., Calas, B., Grassy, G., 1999. Peptides quantitative structure–function relationships: an automated mutation strategy to design peptides and pseudopeptides from substitution matrices. J Mol Graph Model 17 (5–6), 292–309.

Aplan, F.F., Chander, S., 1987. Collectors for sulfide mineral flotation. In: Somasundran (Ed.), Reagents in Mineral Technology. Marcel Dekker, New York, pp. 335–369.

Cioslowski, J., Martinow, M., Mixon, S.T., 1993. Atomic fukui indices from the topological theory of atoms in molecules applied hatree-fock and correlated electron densities. J. Phys. Chem. 97, 10948–10951.

Connolly, M.L., 1984. Computation of molecular volume. J. Am. Chem. Soc. 107, 1118–1124.

Couling, D.J., Bernot, R.J., Docherty, K.M., Dixon, J.K., Maginn, E.J., 2006. Assessing the factors responsible for ionic liquid toxicity to aquatic organisms via quantitative structure–property relationship modeling. Green Chem. 8 (1), 82–90.

Crozier, R.D., 1978. Processing of copper sulfide ores: froth flotation reagents – a review. Mining Mag. 138, 332–339.

Fairthorne, G., Fornasiero, D., Ralston, J., 1997. Interaction of thionocarbamate and thiourea collectors with sulfide minerals: a flotation and adsorption study. Int. J. Miner. Process. 50 (4), 227–242.

Fan, Y., Shi, L.M., Kohn, K.W., Pommier, Y., Weinstein, J.N., 2001. Quantitative structure–antitumor activity relationships of camptothecin analogues: cluster analysis and genetic algorithm-based studies. J. Med. Chem. 44 (20), 3254–3263.

Friedman, J.H., 1991. Multivariate adaptive regression splines. Ann. Stat. 19 (1), 1–67.

Fu, Y.L., Wang, S.S., 1987. Neutral hydrocarboxycarbonyl thionocarbamate sulfide collectors. US Patent 4 (657), 688.

Hall, L.H., Kier, L.B., Hall, L.M., 2002. The Guide for Development of QSAR with MDL QSAR. MDL Information System, San Leandro, CA.

Hari Narayana Moorthy, N.S., Ramos, M.J., Fernandes, P.A., 2011. Structural analysis of α-glucosidase inhibitors by validated QSAR models using topological and hydrophobicity based descriptors. Chemometr. Intell. Lab. Syst. 109, 101–112.

Harris, G.H., Fishback, B.C. 1954. Process for the manufacture of dialkyl thionocarbamates. Dow Chemical Company, US Patent 2691635.

Holland, J., 1975. Adaptation in Artificial and Natural Systems. University of Michigan, Ann Arbor, MI.

Hope, G.A., Buckley, F.M., Munce, C.G., Woods, R., 2007. Gold enhanced spectroelectrochemical investigation of 2-mercaptobenzothiazole, isopropyl xanthate and butylethoxycarbonylthiourea adsorption on minerals. Miner. Eng. 20, 964–969.

Hope, G.A., Woods, R., Parker, G.K., Watling, K.M., Buckley, F.M., 2005. Spectroelectrochemical investigations of flotation reagent-surface interaction. Miner. Eng. 19, 561–570.

Hou, T., Li, Y., Zhang, W., Wang, J., 2009. Recent developments of In silico predictions of intestinal absorption and oral bioavailability. Comb. Chem. 12, 497–506.

Hu, Y.H., Chen, P., Sun, W., 2012. Study on quatitative structure-activity relationship of quaternary ammonium salt collectors for bauxite reverse flotation. Miner. Eng. 26, 24–33.

Iyer, M., Zheng, T., Hopfinger, A.J., Tseng, Y.J., 2007. QSAR analyses of skin penetration enhancers. J. Chem. Inf. Model 47, 1130–1149.

Kier, L.B., Hall, L.H., 1976. Molecular connectivity indices in chemistry and drug research. Med. Chem., 14.

Li, C.D., Sun, C.Y., 2000. The research progress of Cu–S flotation separation. Metal. Ore Dressing Abroad 37 (8), 2–7 (in Chinese).

Liu, G.Y. 2004. Research on the comprehensive utilization for copper sulfide ores with new collectors. Ph.D. Thesis, Central South University, Changsha, Hunan.

Liu, G.Y., Zhong, H., Dai, T.G., 2006. The separation of Cu/Fe sulfide minerals at slightly alkaline conditions by using ethoxycarbonyl thionocarbamates as collectors: theory and practice. Miner. Eng. 19, 1380–1384.

Liu, G.Y., Zhong, H., Dai, T.G., Xia, L.Y., 2008. Investigation of the effect of N-substituents on performance of thionocarbamates as selective collectors for copper sulfides by ab initio calculations. Miner. Eng. 21, 1050–1054.

Mandal, A.S., Roy, K., 2009. Predictive QSAR modeling of HIV reverse transcriptase inhibitor TIBO derivatives. Eur. J. Med. Chem. 44, 1509–1524.

Material Studio 4.0 Tutorials. 2005. Accelrys, San Diego, CA, USA.

Nagaraj, D.R., Brinen, J.S., 2001. SIMS study of adsorption of collectors on pyrite. Int. J. Miner. Process 63, 45–57.
Natarajan, R., Nirdosh, I., 2008. Quantitative structure–activity relationship (QSAR) approach for the selection of chelating mineral collectors. Miner. Eng. 21, 1038–1043.
Parr, R.G., Yang, W., 1984. Density functional approach to the frontier-electron theory of chemical reactivity. J. Am. Chem. Soc. 106 (14), 4049–4050.
Perdew, J.P., Burke, K., Ernzerhof, M., 1996. Phys. Rev. Lett. 77, 3865.
Ponnurengam, M.S., Sethu, K.G., Doble, M., 2006. QSAR studies on chalcones and flavonoids as anti-tuberculosis agents using genetic function approximation (GFA) method. Chem. Pharm. Bull. 55 (1), 44–49.
Rogers, D., Hopfinger, A.J., 1994. Application of genetic function approximation to quantitative structure–activity relationships and quantitative structure–property relationships. J. Chem. Inf. Comput. Sci. 34 (4), 854–866.
Rohrbaugh, R.H., Jurs, P.C., 1987. Descriptions of molecular shape applied in studies of structure/activity and structure/property relationships. Anal. Chim. Acta 199, 99–109.
Sabljic, A., 2001. QSAR models for estimating properties of persistent organic pollutants required in evaluation of their environmental fate and risk. Chemosphere 43, 363–375.
Seryakova, I.V., Vorobiova, G.A., Glembotskii, A.V., 1975. Extraction of metals by neutral sulfur-containing extractants Part I. O-isopropyl-N-ethylthionocarbamate. Anal. Chim. Acta. 77, 183–190.
Shi, L.M., Fan, Y., Lee, J.K., Waltham, M., Andrews, D.T., Scherf, U., Paull, K.D., Weinstein, J.N., 1999. Mining and visualizing large anticancer drug discovery databases. J. Chem. Inf. Comput. Sci. 40 (2), 367–379.
Singh, R.K., Verma, S.K., Sharma, P.D., 2011. DFT based study of interaction between frontier orbitals of transition metal halides and thioamides 3, 1571–1579.
Tobón, Y.A., Romano, R.M., Hey-Hawkins, E., Boese, R., Della Védova, C.O., 2009. A comprehensive study of $(CH_3)_2CHOC(S)SC(O)OCH_3$ using matrix isolation technique, X-ray analysis, spectroscopic studies and theoretical calculations. J. Phys. Org. Chem. 22, 815–822.
Walker, J.D., Jaworska, J., Comber, M.H.I., Schultz, T.W., Dearden, J.C., 2003. Environ. Toxicol. Chem. 22, 1653–1665.
Woods, R., Hope, G.A., 1999. A SERS spectroelectrochemical investigation of the interaction of O-isopropyl-N-ethylthionocarbamate with copper surfaces. Colloids Surf. A: Physicochem. Eng. Aspects 146 (1–3), 63–74.
Wang, N.C.Y., Venkatapathy, R., Bruce, R.M., Moudgal, C., 2011. Development of quantitative structure–activity relationship (QSAR) models to predict the carcinogenic potency of chemicals. II. Using oral slope factor as a measure of carcinogenic potency. Regul. Toxicol. Pharmacol. 59, 215–226.
Yangali-Quintanilla, V., Sadmani, A., McConville, M., Kennedy, M., Amy, G., 2010. A QSAR model for predicting rejection of emerging contaminants (pharmaceuticals, endocrine disruptors) by nanofiltration membranes. Water Res. 44, 373–384.