

## QSAR Modeling of Flotation Collectors Using Principal Components Extracted from Topological Indices

R. Natarajan,<sup>\*,‡</sup> Inderjit Nirdosh,<sup>‡</sup> Subhash C. Basak,<sup>§</sup> and Denise R. Mills<sup>§</sup>

Department of Chemical Engineering, Lakehead University, Thunder Bay, Ontario, Canada P7B 5E1, and  
Natural Resources and Research Institute, University of Minnesota, Duluth, Minnesota 55811

Received March 20, 2002

Several topological indices were calculated for substituted-cupferrons that were tested as collectors for the froth flotation of uranium. The principal component analysis (PCA) was used for data reduction. Seven principal components (PC) were found to account for 98.6% of the variance among the computed indices. The principal components thus extracted were used in stepwise regression analyses to construct regression models for the prediction of separation efficiencies (Es) of the collectors. A two-parameter model with a correlation coefficient of 0.889 and a three-parameter model with a correlation coefficient of 0.913 were formed. PCs were found to be better than partition coefficient to form regression equations, and inclusion of an electronic parameter such as Hammett  $\sigma$  or quantum mechanically derived electronic charges on the chelating atoms did not improve the correlation coefficient significantly. The method was extended to model the separation efficiencies of mercaptobenzothiazoles (MBT) and aminothiophenols (ATP) used in the flotation of lead and zinc ores, respectively. Five principal components were found to explain 99% of the data variability in each series. A three-parameter equation with correlation coefficient of 0.985 and a two-parameter equation with correlation coefficient of 0.926 were obtained for MBT and ATP, respectively. The amenability of separation efficiencies of chelating collectors to QSAR modeling using PCs based on topological indices might lead to the selection of collectors for synthesis and testing from a virtual database.

### INTRODUCTION

Froth flotation is the single most important process used for the beneficiation of low-grade ores. A variety of surface-active reagents are employed in the process. Depending upon their function these reagents are classified into *collectors*, *frothers*, *depressants*, and *activators*.<sup>1</sup> A *collector* imparts hydrophobicity to a mineral by selective adsorption on its surface. A *frother* is used to form a relatively stable froth thus preventing the mineralized air bubbles from coalescence. A *depressant* depresses the flotation of a less desirable mineral either by preventing the adsorption of the collector or by rendering the surface hydrophilic (if it is already hydrophobic). An *activator* makes certain minerals responsive to collectors thus enabling them to float.

A collector adsorbs on a mineral either by physisorption or chemisorption.<sup>2,3</sup> In the absence of strong mineral collector interaction through polar functional groups the adsorption is predominantly through the electrical double layer interaction. Such an interaction is known as the electrostatic bonding and the collector is said to be physisorbed. Chemisorption involves strong collector-mineral interaction through formation of coordinate covalent bonds between the metal ion in the mineral and the atoms of the collectors' polar groups. The adsorption of a physisorbing collector mainly depends on the point of zero charge (pzc) of the mineral. However, it is not so in the case of the chemisorbing collectors.

Chelating agents are organic compounds that are capable of bonding to the metal ions via coordinate bonds resulting in the formation of one or more rings, called the chelate ring(s). Every chelating agent has one characteristic group that contains at least two atoms to form the chelate ring.<sup>4–6</sup> These atoms are known as donor atoms and the most common donor atoms are N, O, S, and P. Flotation involves selective adsorption of surfactants on mineral surfaces. Chelating agents being capable of forming chelates with metal ions in mineral lattice have been tested as collectors. Transition metals form chelates through coordinate bonds more readily than nontransition elements. Due to this difference in bond formation, a chelating collector does not adsorb the gangue minerals such as clay, quartz, and silica and consequently has better selectivity.

The chelating group of a collector may be selected based on the chelating chemistry of the metal ion. Marabini and co-workers<sup>7</sup> suggested a method of selection of the chelating group based on the conditional stability constant. Pradip<sup>8</sup> reviewed the semiempirical methods attempted to select the chelating group. However, quantification of the hydrocarbon chain had not been attempted.

When one thinks about quantification of molecular structures, topological indices are the natural choice. Topological indices point out the mathematical characterization of molecular structure. There are several reviews<sup>9–11</sup> on the listing and basic definitions of topological indices. For the sake of brevity the topological indices are not defined here. There are well over 300 topological indices, some of which have been shown to be related to several physicochemical properties,<sup>12</sup> biological properties,<sup>9</sup> and pharmacological proper-

\* Corresponding author phone: (807)343-8754; fax: (807)343-8928; e-mail: rnat35@hotmail.com.

<sup>‡</sup> Lakehead University.

<sup>§</sup> NRRI, UMN, Duluth.

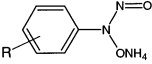
ties.<sup>13,14</sup> To test the applicability of topological indices as a measure of the hydrocarbon chain of collectors, a preliminary attempt was made by authors<sup>15</sup> to correlate the flotation efficiencies of substituted-cupferrons against the Kier and Hall valence connectivity indices. The present study is a continuation of the effort to test the amenability of the flotation efficiency of mineral collectors to QSAR modeling. The success of the approach is of paramount importance because topological indices can be used to select structural analogues to be tested as mineral collectors. The development of a scientific method based on topological indices for the design of mineral collectors is expected to give an impetus to the experiments in the identification and synthesis of mineral-selective collectors. Several topochemical and topostructural indices are computed for the substituted-cupferrons and are used in regression analysis.

Some topological indices are highly intercorrelated. Basak et al.<sup>16</sup> studied the mutual relatedness of a large number of calculated descriptors. Therefore, the intrinsic dimensionality of the data may be much smaller than the number of indices computed. In such a situation Principal Component Analysis (PCA) is a standard tool for data reduction.<sup>17</sup> The first step in principal component analysis is to create a correlation matrix or matrix of variance from the basic data matrix. The data matrix is an  $m \times n$  matrix where the columns of the matrix represent the variables and its rows represent the samples or tests. The matrix is described by a characteristic polynomial equation. The principal components or the factors of this matrix are then determined by computing the characteristic roots of the polynomial. A new matrix is generated in this procedure. The characteristic roots of the polynomial are known as eigenvalues that are a set of positive quantities. The eigenvalues themselves are known as factor loadings. The sum of the squared factor loadings of all features is equal to the eigenvalue of each individual factor. There is an equal number of associated orthogonal vectors called eigenvectors. The factors or principal components are computed in such a way that the first factor describes the largest part of the data variability, the second factor is orthogonal to the first and describes the largest part of the variability left over by the former factor, and so on. Malinowski and Howerly<sup>18</sup> outlined the matrix operations involved in PCA and its application to problems in chemistry in their monograph. Usually principal components with a eigenvalue equal to or greater than one are retained or extracted. Hence, by performing principal component analysis, the majority of the information contained within the original data can be represented in a subspace of lower dimension.

## MATERIALS AND METHODS

**Flotation Data.** Cupferron and substituted-cupferrons were synthesized and tested<sup>19</sup> as collectors to float a Canadian uranium ore. Separation efficiencies of the collectors are taken as the dependent parameter. Separation efficiency of a collector is defined as the difference between the % recovery of the valuable mineral and that of the gangue mineral in the float concentrate. For low-grade ores the separation efficiency (Es) can be expressed as  $Es = \% \text{ metal recovery} - \% \text{ mass recovery}$ . The original data contained 24 compounds. Two of these compounds namely 4-butyl-

**Table 1.** Substituted-Cupferrons Used in the Study



no.	substituent (R)	separation efficiency (Es)		
		exp.	predict1 <sup>a</sup>	predict2 <sup>b</sup>
1	H	27.1	24.5	28.5
2	4-methyl	36.3	33.4	33.9
3	3-methyl	36.4	28.6	27.9
4	4-ethyl	30.6	35.8	35.4
5	4-propyl	33.4	38.2	37.3
6	4-isopropyl	37.4	38.4	37.2
7	4-tert-butyl	37.3	41.8	39.5
8	4-pentyl	42.5	44.0	43.0
9	4-heptyl	52.3	49.6	49.2
10	4-octyl	51.7	53.1	53.2
11	4-nonyl	58.7	56.4	56.8
12	4-fluoro	43.5	41.0	41.9
13	3-fluoro	36.3	36.3	36.1
14	4-chloro	47.1	42.9	43.6
15	3-chloro	31.1	38.2	37.8
16	4-bromo	44.2	44.8	45.4
17	4-methoxy	42.5	39.2	38.7
18	3-methoxy	32.9	35.7	34.2
19	2,6-dimethyl	32.5	37.5	36.5
20	3,5-dimethyl	40.7	40.1	39.5
21	2,4,6-trimethyl	44.8	44.0	42.6
22	4-phenyl	52.4	48.2	53.6

<sup>a</sup> Calculated using eq 1, two-parameter model. <sup>b</sup> Calculated using eq 2, three-parameter model.

cupferron and 4-isobutylcupferron were deleted from the data set because their separation efficiencies were much lower than that of the others. A suitable reason could not be offered for the variation. The substituted-cupferrons used in the study and their separation efficiencies are given in Table 1.

**Computation of Topological Indices.** Topological indices used in the study were Wiener index,<sup>20</sup> Harary index,<sup>21</sup> molecular connectivity indices developed by Randić<sup>22</sup> and Kier and Hall,<sup>9</sup> information theoretic indices based on distance matrices of graphs,<sup>23,24</sup> Balaban's J index,<sup>25,26</sup> and information indices based on neighborhood complexity of vertices in the hydrogen filled as well as hydrogen suppressed molecular graphs.<sup>27-29</sup> They were calculated using POLLY<sup>30</sup> and INDCAL.<sup>31</sup> The symbols and definitions of the topological parameters used in this study are listed in Table 2. The topological indices were calculated for cupferron acid and not for the ammonium salt.

## RESULTS AND DISCUSSION

It has already been stated that PCA is affected by scaling factors. In the present study all topological indices were transformed by the natural logarithm of the index (TI) plus 1 ( $\log_e(TI + 1)$ ). This was done because the magnitudes of some indices were several times greater than those of others. The data matrix was subjected to data reduction using the Principal Component Analysis in SPSS 10.0 for windows.<sup>32</sup>

The first seven principal components (PC) were extracted because they had eigenvalues greater than or equal to one and they account for 98.6% of the data variance. Of the seven PCs the first two alone account for 79.6% of the variance. The eigenvalues and percent variance of the principal components are given in Table 3. The principal components were correlated with the individual topological indices. The

**Table 2.** Symbols and Definitions of Topological Indices Used in the Study

$I^W_D$	information index for the magnitudes of distances between all possible pairs of vertices of a graph
$\bar{I}^W_D$	mean information index for the magnitude of distance
W	Wiener index = half-sum of the off-diagonal elements of the distance matrix of a graph
H	Harary index = half-sum of reciprocal square of the elements of the distance matrix of a graph
J	Balaban index = sum of the reciprocal square root of degree of all adjacent pairs
$I_C^D$	information content of the distance matrix
IC	information content of the distance matrix partitioned by frequency of occurrences of distance h
O	order of neighborhood when $IC_r$ reaches its maximum value for the hydrogen-filled graph
M1	a Zagreb group parameter = sum of square of degree over all vertices
M2	a Zagreb group parameter = sum of cross-product of degrees over all neighboring (connected) vertices $M_2$
$^h\chi$	path connectivity index of order $h = 0-6$
$^h\chi_c$	cluster connectivity index of order $h = 3-6$
$^h\chi_{ch}$	chain connectivity index of order $h = 3-6$
$^h\chi_{pc}$	path-cluster connectivity index of order $h = 4-6$
$P_h$	number of paths of length $h = 0-10$
$I_{ORB}$	information content or complexity of the hydrogen-suppressed graph at its maximum neighborhood of vertices
$O_{ORB}$	order of neighborhood when $IC_r$ reaches its maximum value for the hydrogen suppressed graph
$IC_r$	mean information content or complexity of a graph based on the $r$ th ( $r = 0-6$ ) order neighborhood of vertices in a hydrogen-filled graph
$SIC_r$	structural information content for $r$ th ( $r = 0-6$ ) order neighborhood of vertices in a hydrogen-filled graph
$CIC_r$	complementary information content for $r$ th ( $r = 0-6$ ) order neighborhood of vertices in a hydrogen-filled graph
$TIC_r$	total information content for $r$ th ( $r = 0-6$ ) order neighborhood of vertices in a hydrogen-filled graph
$BIC_r$	bond information content for $r$ th ( $r = 0-6$ ) order neighborhood of vertices in a hydrogen-filled graph
$HsIC_r$	mean information content or complexity of a graph based on the $r$ th ( $r = 0-6$ ) order neighborhood of vertices in a hydrogen-suppressed graph
$HsSIC_r$	structural information content for $r$ th ( $r = 0-6$ ) order neighborhood of vertices in a hydrogen-suppressed graph
$HsCIC_r$	complementary information content for $r$ th ( $r = 0-6$ ) order neighborhood of vertices in a hydrogen-suppressed graph
$HsTIC_r$	total information content for $r$ th ( $r = 0-6$ ) order neighborhood of vertices in a hydrogen-suppressed graph
$HsBIC_r$	bond information content for $r$ th ( $r = 0-6$ ) order neighborhood of vertices in a hydrogen-suppressed graph
$^h\chi^b$	bond path connectivity index of order $h = 0-9$
$^h\chi_c^b$	bond cluster connectivity index of order $h = 3-6$
$^h\chi_{ch}^b$	bond chain connectivity index of order $h = 3-6$
$^h\chi_{pc}^b$	bond path-cluster connectivity index of order $h = 4-6$
$^h\chi^v$	valence path connectivity index of order $h = 0-9$
$^h\chi_c^v$	valence cluster connectivity index of order $h = 3-6$
$^h\chi_{ch}^v$	valence chain connectivity index of order $h = 3-6$
$^h\chi_{pc}^v$	valence path-cluster connectivity index of order $h = 4-6$

10 most highly correlated indices are given in Table 4. The  $PC_1$  was in high correlation, primarily to the path numbers and the connectivity indices. In addition to these the distance-based indices such as W,  $I^W_D$ , and  $\bar{I}^W_D$  were also correlated with the  $PC_1$ . The second PC represented molecular neighborhood symmetry and was highly correlated to  $BIC_6$ ,  $BIC_5$ ,  $HsSIC_4$ ,  $HsSIC_5$ ,  $HsSIC_6$ ,  $BIC_4$ ,  $SIC_6$ , etc. The  $PC_3$  correlated to the same type of indices as the  $PC_2$  but of lower order

**Table 3.** Summary of Principal Component Analysis

PC	eigenvalues	percent of variance	cumulative percent
PCA for Cupferron-Derivatives			
1	77.74	55.14	55.14
2	34.50	24.46	79.60
3	13.72	9.73	89.33
4	5.82	4.13	93.46
5	3.62	2.57	96.03
6	2.30	1.63	97.66
7	1.35	0.96	98.62
PCA for Amino thiophenols			
1	99.04	77.37	77.37
2	16.57	12.95	90.32
3	8.46	6.61	96.93
4	1.96	1.54	98.47
5	1.31	1.02	99.49
PCA for Mercaptobenzothiozole			
1	102.71	77.23	77.23
2	16.89	12.70	89.93
3	6.62	4.98	94.91
4	2.70	2.03	96.94
5	2.32	1.74	98.68

calculated for the hydrogen-suppressed graphs. The  $PC_4$  was predominantly correlated to cluster and path-cluster terms.

The seven principal components were used as independent parameters in constructing regression models. Stepwise regression analysis was performed using the PCs extracted from the topological indices. The results of regression analyses are given in Table 5. The two-parameter model accounts for 79.1% of variance, and the regression equation is

$$Es = (40.69 \pm 1.78) + (5.66 \pm 1.76) PC_1 + (4.23 \pm 1.75) PC_7 \quad (1)$$

$$R^2 = 0.791; SEE = 3.982; n = 22$$

The three-parameter model accounts for 83.4% of the variance, and there is improvement in the predictability. The regression equation is given below:

$$Es = (40.59 \pm 1.64) + (5.61 \pm 1.62) PC_1 + (4.25 \pm 1.65) PC_7 - (1.68 \pm 1.64) PC_3 \quad (2)$$

$$R^2 = 0.834; SEE = 3.647; n = 22$$

In both the models the standard error of the estimate (SEE) is below allowed experimental error. The separation efficiencies (Es) calculated using the regression equations are given in Table 1 along with the experimental values.

In a similar study to form property-property relations,<sup>33</sup> octanol-water partition coefficients/soil-water partition coefficients were used in combination with electronic parameters namely, Hammett substituent constants or potential derived charges on the chelating atoms in the cupferron-derivatives. The correlation coefficient was less than 0.8 for the multiple regression equations formed using various combinations of a partition coefficient and an electronic parameter. The correlation coefficient improved only when five of the data namely, 4-fluoro, 4-chloro, 4-bromo, 4-methoxy, and 4-phenyl substituted compounds were dropped from the regression analyses. The three-parameter regression equation (eq 2) obtained in the present study using PCs has correlation coefficient 0.913 and moreover, neither an indicative variable was included to



**Table 4.** Correlation Coefficients of Variables (10 Most Highly Correlated) with the Principal Component Scores

PC <sub>1</sub>		PC <sub>2</sub>		PC <sub>3</sub>		PC <sub>4</sub>		PC <sub>5</sub>		PC <sub>6</sub>		PC <sub>7</sub>	
TI	R	TI	R	TI	R	TI	R	TI	R	TI	R	TI	R
P <sub>0</sub>	0.996	BIC <sub>6</sub>	0.969	HsIC <sub>1</sub>	0.909	$\chi_{pc}^6$	0.542	IC <sub>2</sub>	0.440	$\chi_{pc}^6$	0.349	IC <sub>0</sub>	0.388
I <sup>W</sup> <sub>D</sub>	0.995	BIC <sub>5</sub>	0.968	HsIC <sub>0</sub>	0.891	$\chi_{pc}^4$	0.472	$\chi^7$	0.420	P <sub>4</sub>	0.311	SIC <sub>0</sub>	0.316
P <sub>1</sub>	0.994	HsSIC <sub>4</sub>	0.942	HsBIC <sub>1</sub>	0.878	$\chi_{pc}^5$	0.460	P <sub>7</sub>	0.403	$\chi_{pc}^6$	0.304	BIC <sub>0</sub>	0.303
H	0.993	HsSIC <sub>5</sub>	0.942	HsSIC <sub>1</sub>	0.800	$\chi_{pc}^6$	0.455	$\chi_{ch}^6$	0.394	$\chi^4$	0.275	IC <sub>1</sub>	0.286
$\chi^0$	0.993	HsSIC <sub>6</sub>	0.925	HsBIC <sub>0</sub>	0.741	$\chi_{ch}^5$	0.428	$\chi_{pc}^6$	0.370	$\chi_{pc}^5$	0.266	SIC <sub>1</sub>	0.248
$\chi^1$	0.991	BIC <sub>4</sub>	0.923	HsBIC <sub>2</sub>	0.670	$\chi_{ch}^6$	0.420	$\chi_{ch}^6$	0.370	P <sub>9</sub>	0.246	BIC <sub>1</sub>	0.228
I <sup>W</sup> <sub>D</sub>	0.990	SIC <sub>6</sub>	0.908	IC <sub>2</sub>	0.665	$\chi_{ch}^6$	0.420	IC <sub>3</sub>	0.313	IC <sub>2</sub>	0.240	$\chi_{ch}^6$	0.159
W	0.988	SIC <sub>5</sub>	0.902	HsSIC <sub>0</sub>	0.646	$\chi^3$	0.414	P <sub>6</sub>	0.297	$\chi_{pc}^5$	0.217	$\chi^6$	0.154
$\chi^3$	0.984	HsSIC <sub>3</sub>	0.895	HsIC <sub>2</sub>	0.581	$\chi_{pc}^4$	0.408	BIC <sub>2</sub>	0.280	P <sub>5</sub>	0.212	$\chi_{ch}^6$	0.141
M <sub>1</sub>	0.984	HsBIC <sub>4</sub>	0.892	HsTIC <sub>0</sub>	0.546	$\chi_{pc}^6$	0.391	SIC <sub>2</sub>	0.257	$\chi_{ch}^4$	0.186	$\chi_{ch}^5$	0.133

**Table 5.** Results of Regression Analysis Using Principal Components

no.	PC included	R	R <sup>2</sup>	adjusted R <sup>2</sup>	SEE	F
Cupferron-Derivatives ( <i>n</i> = 22)						
1	PC <sub>1</sub>	0.713	0.509	0.485	5.946	20.74
2	PC <sub>1</sub> , PC <sub>7</sub>	0.889	0.791	0.769	3.982	35.92
3	PC <sub>1</sub> , PC <sub>7</sub> , PC <sub>3</sub>	0.913	0.834	0.806	3.647	30.10
Aminothiophenols ( <i>n</i> = 9)						
1	PC <sub>1</sub>	0.681	0.463	0.387	17.840	6.04
2	PC <sub>1</sub> , PC <sub>3</sub>	0.926	0.857	0.809	9.956	17.93
Mercaptobenzothiozole ( <i>n</i> = 15)						
1	PC <sub>1</sub>	0.710	0.504	0.465	17.139	13.18
2	PC <sub>1</sub> , PC <sub>2</sub>	0.942	0.887	0.868	8.522	46.96
3	PC <sub>1</sub> , PC <sub>2</sub> , PC <sub>4</sub>	0.971	0.942	0.927	6.353	59.85
Mercaptobenzothiozole ( <i>n</i> = 14; after Deleting 6-HexyloxyMBT)						
1	PC <sub>1</sub>	0.732	0.536	0.497	16.824	13.84
2	PC <sub>1</sub> , PC <sub>2</sub>	0.962	0.926	0.913	7.010	68.90
3	PC <sub>1</sub> , PC <sub>2</sub> , PC <sub>4</sub>	0.985	0.970	0.961	4.691	107.42

offset the deviation of certain data nor was any data dropped from the final list of 22 compounds. It might be concluded that the PCs extracted from topological indices considered in the present study are better than partition coefficients as independent parameters in forming regression equations to model the separation efficiencies of cupferron-derivatives.

Though molecular topology seem to play the key role in determining the extent of hydrophobicity imparted to a mineral by a chelating collector the strength of the bonds formed between the chelating atoms and the metal ion may also have its contribution in determining the efficiency of separation. The chelating ability depends on the electron density (electronic charge) on the chelating atoms and in turn is affected by the electronic effects of the substituents in the collector molecules. Hence, multiple regression analysis was repeated including Hammett substituent constants. Hammett  $\sigma$  values were calculated according to Perrin et al.<sup>34</sup> and reported<sup>33</sup> for the cupferron-derivatives. The correlation coefficient improved from 0.913 to 0.922 and the SEE was 3.562 when  $\sigma$  was included as an additional independent parameter in eq 2. Mullikan charges and potential derived charges on the two oxygen atoms of the substituted-cupferrons calculated and reported in an earlier study<sup>33</sup> were used in the place of Hammett substituent constants. Inclusion of either Mullikan charges or the potential derived charges on the chelating atoms as electronic parameters did not improve the correlation coefficient significantly. The best correlation after including the electronic charge on the chelating atoms was obtained when the Mullikan charge on the oxygen of  $-\text{ONH}_4$  group was used as an independent parameter along with the three PCs in eq 2. The correlation

coefficient for the four-parameter equation was 0.929 with SEE 3.417 and F-score 26.587. As in the case with Hammett constants the improvement in correlation coefficient was only marginal. These results indicate that the principal components based on topological indices are good enough to be used as independent parameters in forming structure activity (flotation) relationships (SAR), and the quantum mechanically derived electronic charges on the chelating atoms was not essential. If at all one wants to include the electronic contribution of the substituents in the regression model, Hammett substituent constants suffice to be used with topological indices or the PCs extracted from them.

To test the applicability of SAR modeling of flotation efficiencies of chelating mineral collectors using topological indices two other series of collectors reported by Marabini<sup>7</sup> were considered. Mercaptobenzothiozoles (MBT) and aminothiophenols (ATP) were tested to float a lead ore and a zinc ore, respectively. The structures and separation efficiencies of MBT and ATP collectors are given in Table 6. PCA was carried out for each series, following the procedure explained for cupferron-derivatives. The summary of PCA for both the ATP and MBT collectors are given in Table 3. Five PCs were found to explain nearly 99% of the variance in each series.

Results of stepwise linear regression analyses using the five PCs are given in Table 5. A two-parameter equation with a correlation coefficient of 0.926 was obtained for ATP. The equation is given below, and the separation efficiencies calculated using eq 3 are given in Table 6:

$$Es = (35.70 \pm 10.12) + (15.50 \pm 8.61) PC_1 - (14.29 \pm 8.41) PC_3 \quad (3)$$

$$R^2 = 0.857; SEE = 9.96; n = 9$$

When soil–water partition coefficient was used correlation coefficient was not significant, and to get a reasonable correlation an electronic parameter and an indicator parameter need to be included<sup>33</sup> in the regression equation. Equation 3 formed with PCs is better than that evolved in the previous study<sup>33</sup> on ATP using partition coefficients, and this is similar to the results explained above for cupferrons.

In the case of MBT collectors the regression equation formed with PC<sub>1</sub>, PC<sub>2</sub>, and PC<sub>4</sub> was found to have the best predictive ability. The equation is given below:

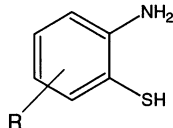
$$Es = (57.70 \pm 3.62) + (16.64 \pm 3.75) PC_1 - (15.40 \pm 3.75) PC_2 - (5.53 \pm 3.75) PC_4 \quad (4)$$

$$R^2 = 0.942; SEE = 6.53; n = 15$$

**Table 6.** Structure of Aminothiophenols and Their Separation Efficiencies (Es)

no.	substituent (R)	separation efficiency (Es)	
		exp.	predicted

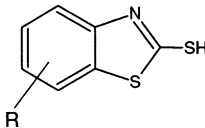
  



Aminothiophenols

1	5-methoxy	14.7	7.0
2	5-ethoxy	9.3	23.2
3	5-propoxy	47.8	36.3
4	5-butoxy	45.1	47.4
5	5-pentoxy	68.7	56.4
6	5-hexyloxy	70.1	61.9
7	5-methyl	9.1	11.4
8	5-hexyl	46.4	57.7
9	6-hexyloxy	20.1	20.0



Mercaptobenzothiazole

1	6-methoxy	63.6	71.2
2	6-ethoxy	67.7	73.3
3	6-propoxy	75.7	76.5
4	6-butoxy	76.6	71.6
5	6-pentoxy	73.1	67.9
6	6-hexyloxy	76.6	
7	6-methyl	73.7	68.5
8	6-ethyl	72.1	70.7
9	6-propyl	73.7	71.0
10	6-hexyl	53.6	56.3
11	5-butoxy	67.3	71.6
12	7-hexyloxy	43.2	44.4
13	2-mercaptobenzo-1,3-thiazole	23.5	21.9
14	2-mercaptobenzodiazole	12.8	15.6
15	2-mercaptobenzoxazole	12.3	10.5

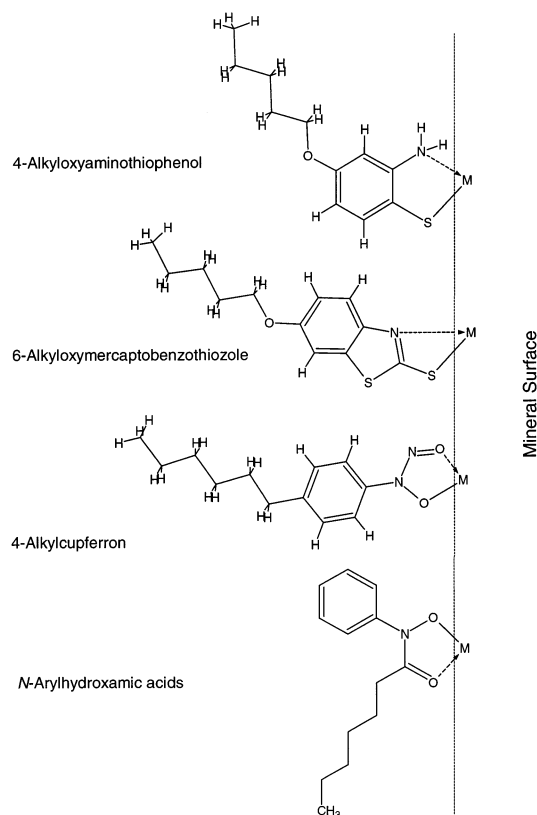
Separation efficiency predicted using eq 4 was found to have residual greater than two times the standard deviation for 6-hexyloxymercaptobenzothiosole (compound # 6 in Table 6). Hence, multiple regression analysis was repeated after excluding this datum and there was significant improvement in the correlations coefficient and F-score (Table 5). Equation 4 was modified as

$$Es = (56.54 \pm 2.81) + (15.30 \pm 2.94) PC_1 - (15.80 \pm 2.93) PC_2 - (4.85 \pm 2.83) PC_4 \quad (5)$$

$$R^2 = 0.970; SEE = 4.69; n = 14$$

Separation efficiencies calculated using eq 5 are given in Table 6. In the case of MBT collectors also the regression equations evolved using PCs were found to be superior to those formed using soil–water partition coefficient.

Perusal of the best regression model obtained for each of the three series of collectors indicates that the SEE is the highest for aminothiophenols. One possible reason for the relatively high SEE for the ATP might be the steric hindrance of certain large substituents that reduce the hydrophobic association of collectors adsorbed on the mineral surface. In the case of cupferrons the steric hindrance would be the least, as the chelating atoms are not directly attached to the

**Figure 1.** Schematic representation of steric hindrance near mineral surface preventing adsorption on adjacent sites. The hindrance increases in the order Cupferron < MBT < ATP < arylhydroxamic acid.

benzene ring. The long alkyl substituents are mostly in the *para* position of the phenyl ring and consequently there would not be hindrance near the mineral site. In the case of MBT collectors too the situation appears to be similar to that of cupferrons. A recent study on *N*-arylhydroxamic acids indicated that increase in the acyl group beyond hexanoyl decreased<sup>35</sup> the efficiency of the collector and the data did not correlate linearly with topological indices.<sup>36</sup> The high steric hindrance near the mineral surface for *N*-arylhydroxamic acids was the reason for this. The steric hindrance near the mineral surface is schematically shown in Figure 1. It may be noted that topological indices are capable of encoding the topology of individual molecules but not associative or intermolecular interactions. This inherent inability of topological indices and the PCs extracted from them to encode the steric hindrance and the consequent reduction in the number of collectors molecules adsorbed on unit area of a collector appears to be one of the possible reasons for the high SEE of the regression equation (eq 4) for ATP collectors.

In general principal components obtained from topological indices were found to form regression equations that could predict the separation efficiencies of chelating collectors reasonably well. In the opinion of the authors this finding is of great importance because similarity space of collectors for their selection can be constructed with PCs extracted from topological indices. This methodology used hitherto in drug discovery may be extended to mineral processing and will avoid intuitive and trial-error methods of selection of mineral collectors. One can decide chelating group from the knowledge of complex equilibria and after making all possible

variations in the architecture of the hydrocarbon part, a virtual database is created. A few compounds may be selected for further synthesis and testing from the database using molecular similarity clustering. Hence, a combination of the knowledge of chelate chemistry and chemical graph theory appears to open new avenues in the synthesis of mineral collectors that are better than those currently in use.

## REFERENCES AND NOTES

- (1) Wills, B. A. *Mineral Processing Technology*; Pergamon Press: 1992.
- (2) Fuerstenau, D. W. Flotation. In *Research Needs in Mineral Processing*; Report of a Workshop held at Arden House, Hasirman Campus, Columbia University, U.S.A., Aug 1–3, 1975.
- (3) Fuerstenau, D. W.; Herra-Urbina, R. Mineral Separation by Froth Flotation. In *Surfactant Based Separation*; Scamhorn, J. F., Ed.; 1988.
- (4) Dwyer, F. P.; Mellor, D. P. *Chelating Agents and Metal Chelates*; Academic Press: New York, 1984.
- (5) Matrel, A. E.; Calvin, M. *Chemistry of Metal Chelate Compounds*; Prentice Hall: Englewood Cliff, NJ, 1952.
- (6) Perrin, D. D. *Organic Complexing Reagents*; Wiley: New York, 1964.
- (7) Marabini, A. M. Criteria for the Design and Synthesis of Chelating Reagents for Flotation. In *Energy Separations Technologies for Metals and Fuels*; Lakshamanan, V. I., Ed.; The Minerals, Metals and Materials Society: 1993; pp 141–152.
- (8) Pradip. On the Design of Selective Reagents for Mineral processing Applications. *Metals Materials Processes* **1991**, 3, 15–30.
- (9) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
- (10) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Analysis*; Research Studies Press: Chichester, England, 1986.
- (11) Basak, S. C.; Gute, B. D. Characterization of Molecular Structures Using Topological Indices. *SAR QSAR Environ. Res.* **1977**, 7, 1–21.
- (12) *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach: The Netherlands, 1999.
- (13) Basak, S. C.; Mills, D.; Gute, B. D.; Grunwald, G. D.; Balaban, A. T. Applications of Topological Indices in Predicting Property/Bioactivity/Toxicity of Chemicals. In *Proceedings of Harry Wiener Symposium*; Rouvray, D. H., King, R. B., Eds.; Ellis Howard: U.S.A., 2001; in press.
- (14) Reinhard, M.; Drefahl, A. *Handbook for Estimating Physicochemical Properties of Organic Compounds*; John Wiley: New York, 1999.
- (15) Natarajan, R.; Nirdosh, I. QSAR Modeling of Flotation Collectors Part 1- Application of Valence Connectivity Indices to the Flotation of a Uranium Ore using Substituted-Cupferrons. *Indian J. Chem.* **2001**, 40(A), 130–134.
- (16) Basak, S. C.; Balaban, A. T.; Gute, B. D. Interrelationship of major Topological Indices: Their nature and Mutual Relatedness. *J. Chem. Inf. Comput. Sci.* **2000**, 40, 891–898.
- (17) Gnanadesikan, R. *Methods for Statistical Data Analysis of Multivariate Observations*; John Wiley: New York, 1977.
- (18) Malinowski, E. R.; Howery, D. *Factor Analysis in Chemistry*; John Wiley: New York, 1980.
- (19) Natarajan, R.; Nirdosh, I.; Muthuswami, S. V.; Jeyaraman, R. Effect of substituent on the performance of Cupferron as a collector for uranium. *Developments Chem. Eng. Mineral Processing* **1994**, 4, 202–217.
- (20) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, 69, 17–20.
- (21) Mihalic, Z.; Nikolic, S.; Trinajstic, N. Comparative Study of Molecular Descriptors Derived from the Distance Matrix. *J. Chem. Inf. Comput. Sci.* **1992**, 32, 28–37.
- (22) Randić, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1995**, 97, 6609–6615.
- (23) Bonchev, D.; Trinajstic, N. Information Theory, Distance Matrix and Molecular Branching. *J. Chem. Phys.* **1977**, 67, 4517–4533.
- (24) Raychaudhury, C.; Ray, S. K.; Ghosh, J. J.; Roy, A. B.; Basak, S. C. Discrimination of Isomeric Structures Using Information Theoretic Topological Indices. *J. Comput. Chem.* **1984**, 5, 581–588.
- (25) Balaban, A. T. Highly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett.* **1982**, 89, 399–404.
- (26) Balaban, A. T. Topological Indices Based on Topological Distances in Molecular Graphs. *Pure Appl. Chem.* **1983**, 55, 199–206.
- (27) Basak, S. C.; Magnuson, V. R. Molecular Topology and Narcosis: A Quantitative Structure–Activity Relationship (QSAR) Study of Alcohols Using Complementary Information Content (CIC). *Arzneim. Forsch.* **1983**, 33, 501–503.
- (28) Basak, S. C.; Roy, A. B.; Ghosh, J. J. In *Proceedings of the Second International Conference on Mathematical Modelling*; Avula, X. J. R., Bellman, R., Luke, Y. L., Rigler, A. K., Eds.; University of Missouri–Rolla, Rolla, MO, 1979; pp 851–856.
- (29) Roy, A. B.; Basak, S. C.; Harriss, D. K.; Magnuson, V. R. Neighborhood Complexities and Symmetry of Chemical Graphs and Their Biological Applications. In *Mathematical Modelling in Science and Technology*; Avula, X. J. R., Kalman, R. E., Liapis, A. I., Rodin, E. Y., Eds.; Pergamon: New York, 1984; pp 745–750.
- (30) Basak, S. C.; Harris, D. K.; Magnuson, V. R. POLLY Version 2.3; Copyright of the University of Minnesota, 1988.
- (31) Natarajan, R.; Nirdosh, I.; Anbazhagan, T. M.; Murali, T. C. Topological Index Calculator (INDCAL): A Computer Program to Calculate Topostructural and Topochemical Indices. In *The International Seminar on Mineral Processing Technology (MPT2002)*, January 3–5, 2002, Bangalore, India.
- (32) SPSS Base 10.0 Applications Guide; SPSS Inc.: U.S.A., 1999.
- (33) Natarajan, R.; Nirdosh, I.; Venuvanalingam, P.; Ramalingam, M. Quantitative Property–Property Relationship (QPPR) approach in predicting flotation efficiency of chelating agents as mineral collectors. *SAR QSAR Environ. Res.* **2002**, 13(5), 499–508.
- (34) Perrin, D. D.; Dempsey, B.; Serjeant, E. P. *pK<sub>a</sub> prediction for organic acids and bases*; Chapman Hall: London, 1981.
- (35) Natarajan, R.; Nirdosh, I. N-Arylhydroxamic acids as mineral collectors for ore-beneficiation. *Can. J. Chem. Eng.* **2001**, 79, 941–945.
- (36) Natarajan, R.; Nirdosh, I. QSAR Modeling of Flotation Collectors Part 2: Application of Topochemical, Topostructural, Physicochemical and Geometrical Parameters to Model the Flotation Efficiencies of N-Arylhydroxamic Acids. Submitted to *Int. J. Miner. Process.*

CI025513Z