

A QSPR Study of Association Constants of Macrocycles toward Sodium Cation

Shahin Ahmadi

Department of Chemistry, Kermanshah Branch, Islamic Azad University, Kermanshah, Iran
E-mail: ahmadi.chemometrics@gmail.com

The association constant ($\log K$) of 53 chelates of new macrocycles (hemispherands, cryptahemispherands, and bridged calix-4) with sodium cation is predicted by a statistically validated QSPR modeling approach. The applied multiple linear regression is based on a variety of theoretical molecular descriptor selected from 6 classes of Dragon software with a forward stepwise multiple linear regression as a feature selection technique. For external validation we applied self organizing maps (SOM) to split the original data set into training and test set. The best four-dimensional model is developed on a training set of 40 macrocycles. The external validation was performed on test set of 13 macrocycles. The QSPR model presented in this study showed good predictions with the leave one out cross validated variance ($Q^2_{\text{loo-cv}} = 0.94$) and the external-validated variance ($Q^2_{\text{ext}} = 0.92$). The applicability domain (AD) of the model is analysed by leverage method.

Keywords: Macrocycles, sodium cation, association constant, QSPR, self organizing maps.

Introduction

Supramolecular chemistry has developed to mimic the weak non-covalent interactions and the phenomenon of the molecular recognition in biological systems.^[1-5] The characterizing feature of the supramolecular chemistry is that carefully designed synthetic structures (hosts) recognize target molecules (guests) forming a supramolecular complex through non-covalent complexes. Supramolecular chemistry and the quantification of non-covalent interactions offer the basis for new approaches in medicine, host-guest chemistry,^[6] chromatography,^[7] and biocatalysis.^[8]

A host-guest relationship involves a complementary stereoelectronic arrangement of binding sites in host and guest, the host component is defined as an organic molecule or ion whose binding sites converge in the complex the guest component is defined as any molecule or ion whose binding sites diverge in the complex. In this study host molecules are hemispherands, cryptahemispherands, and bridged calix-4. "Hemispherands" have been defined as hosts, at least half of whose structures are composed of units unable to fill their own potential cavities by conformational reorganizations. Hemispherands represent an interesting class of host molecules in supramolecular chemistry. Cram and coworkers and Reinhoudt and coworkers have extensively investigated the complexation behavior of hemispherand molecules with various binding sites.^[9-16] Hemispherands are very efficient hosts for alkali metals and ammonium ions. They form mainly 1:1 complexes, and solid-state structural analyses show the ion perching above the macrocyclic cavity of the host. The designation "cryptahemispherand" was given by Cram in 1986 to the class of macrobicyclic compounds which show an extraordinary propensity for complexation of alkali metal cations.^[9] Cryptahemispherands combine the partially preorganized cavity features of the hemispherands, but contain

multiple other ligand-gathering features of the cryptands. The term "calixarene" was introduced by Gutsche to describe a homologous series of macrocyclic phenolformaldehyde condensates.^[17] It originates from the observation that, in molecular models, the tetrameric members of the series have a chalice-like or cup-like appearance. Calixarenes are mainly receptors for small neutral molecules, but they also interact with cations if the solutions are sufficiently basic to permit deprotonation of the phenolic groups.^[18-19] Bridged calixarenes, in which structural features of calixarenes and spherands are combined, have been prepared by Reinhoudt and co-workers.^[20] These novel compounds exhibit high binding ability for alkali cations due to a high degree of preorganization and the highly hydrophobic collar around the molecular cavity which prevents solvent molecules from assisting in the decomplexation process. Cation complexes play a fundamental role in many biological systems; large quantities of sodium, potassium, magnesium, and calcium ions, in particular, are all critical to life. For example, a concentration gradient of K^+ and Na^+ across biological cell membranes is vital to nerve signal transduction.

In the present work, to understand the effect of macrocyclic compounds on the thermodynamic stabilities of the resulting chelates, the association constant of sodium chelate was regarded as a quantitative base. It depends on many factors, including the strength of metal-ligand bonds, the shape of coordination macrocycle, the steric interactions between atomic groups in a molecule, and solvation interactions.^[16] Here the effect of a solvent and temperature on the association constant was eliminated.

According to the reference searching, there were not many papers on predicting association constants of macrocycle chelates. Recently, the author has studied the QSPR modeling of diverse 15C5 ether chelates with sodium cation using GA-MLR method,^[21] Ghasemi and Saaidpour^[22]

reported QSPR modeling of association constants of diverse 15C5 ethers chelates with lithium cation. Zhang *et al.* reported QSPR of association constants of cesium chelates based on uniform design optimized support vector machine.^[23] Lately, we have studied the QSPR modeling of association constants of the Li-hemispherands complexes using MLR.^[24] There were no published reports on molecular modeling studies to obtain QSPR for association constants of sodium chelates with hemispherands, cryptahemispherands, and bridged calix-4. The present work was to build a new model on association constants, and give an insight into sodium chelates with these macrocycles.

The object of this study is the development of a predictive QSPR model of stability constant for sodium and macrocycles. The important aspects of suggested model are the self-organizing maps data splitting, validation of the QSPR model for predictivity (both by internal and external validation), attention to the applicability domain of the model, and the interpretation of selected molecular descriptors. This model enables to make reliable predictions of the association constant for unknown complexes and to elucidate the structural factors determining the association constants.

Experimental

The chemical structures and experimental values for the stability constants of fifty three compounds taken from the literature^[25] are presented in the Tables 1 and 4, respectively. The data set was split into a training set and a test set with self organizing map (SOM). The training set of 40 macrocycles was used to adjust the parameters of the models, and the test set of 13 macrocycles was used to evaluate its prediction ability. Since the temperature and solvent also affect the stability constants, we used only data obtained at standard temperature (25°C) and just in CDCl₃ saturated with D₂O. Experimental log*K* values vary from 4.32 to 12.08, and 5.36 to 10.04 for training and test sets respectively.

Descriptor Generation

All calculations were run on a Pentium IV personal computer with windows XP as operating system. The structures were drawn in HyperChem 7.5^[26] and the geometrical structure of macrocycles was optimized using semi-empirical quantum method Austin Method 1 (AM1)^[27] using the Polack-Rabiere algorithm until the root mean square gradient was 0.01 within the MOPAC^[28] program package. The geometry and other information from the output of quantum chemical calculations were inserted into the Dragon program,^[29] and descriptors for macrocycles were calculated. All these descriptors are derived solely from molecular structure and do not require experimental data to be calculated. More than 350 molecular descriptors is derived to properly characterize the chemical structure of the fifty three macrocycle compounds, involving variables of the type Constitutional, Topological, GETAWAY (GEometry, Topology and Atoms-Weighted Assembly), WHIM (Weighted Holistic Invariant Molecular descriptors), 3D-MoRSE (3D-Molecular Representation of Structure based on Electron diffraction), Aromaticity Indices.

Data Splitting

To build and validate the QSPR model, the studied dataset is divided into a training set and a test set using self-organizing maps (SOM), which takes advantage of the clustering capabilities, ensuring that both the training set and test set separately span the

whole descriptor space and the chemical domains in the two sets are not too dissimilar.^[30-31] SOM package for MATLAB was downloaded and used from <http://www.cis.hut.fi/projects/somtoolbox/>. The 35 of the most significant principal components calculated from each group of DRAGON molecular descriptors were used to describe the relevant structural information of the macrocycles and were used as variables to build a Kohonen map (6×6 neurons, 300 epochs). Figure 1 shows the number of the macrocycles in each cell of the top-map of the Kohonen network after the training. Not all the neurons were occupied. When a cell was occupied by only one sample, this macrocycle was selected for the training set. When more than one macrocycle with very different values of log*K* was present in a cell, the two samples with extreme values of the responses plus, eventually, an intermediate value were selected. The samples not selected for the training set were put into the test set. The training set and the test set consisted of 40 and 13 objects, respectively.

Feature Selection and Model Construction

In this study, a total of 350 descriptors were initially calculated by Dragon software for the entire data set of 53 macrocycles. The total number of descriptors was reduced to 303 descriptors, by eliminating the collinear descriptors (correlation coefficient is less than 0.1). The best set of descriptors was selected by forward stepwise regression procedure. The forward stepwise regression procedure^[32] is an interesting approach both from the didactical point of view and for the simplicity of the algorithm that involves. It consists on a step by step addition of the best molecular descriptors to the model that lead to the smallest value of the standard deviation (*S*), until there is no other variable outside the equation that satisfies the selection criterion. The definition of *S* employed in present analysis is as follows:

$$S = \frac{1}{N-d-1} \left(\sum_{i=1}^N res_i^2 \right) \quad (1)$$

with *d* being the number of descriptors of the model, *N* is the number of molecules of the training set, and *res_i* stands for the residual of molecule *i* (difference between the experimental and predicted property for *i*). The forward stepwise regression technique requires much less linear regressions than a full search of optimal variables.

Model Validation

The internal predictability and validation power of the regression model developed on the selected training set did by leave one out cross validation (LOO-CV). The cross-validated explained variance (*Q_{ext}²*) is calculated by the following equation:

$$Q_{cv}^2 = 1 - \frac{\sum_{i=1}^{cal} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{cal} (y_i - \bar{y})^2} \quad (2)$$

where *y_i*, *ŷ_i* and *ȳ* are, respectively, the measured, predicted, and averaged (over the entire data set) values of the dependent variable, respectively; the summations cover all the compounds in the training set.

The LOO-CV approach is not sufficient to assess robustness and predictability. The QSPR model developed using only training set chemicals is then applied to the external validation set chemicals to verify, more reliably, the predictive ability of the model.

The formula for the calculation of *Q_{ext}²* is:

$$Q_{ext}^2 = 1 - \frac{\sum_{i=1}^{test} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{test} (y_i - \bar{y}_{tr})^2} \quad (3)$$

where y_i and $\hat{y}_i$ are respectively the measured and predicted (over the test set) values of the dependent variable, and $\bar{y}_r$ is the averaged value of the property for the training set; the summations cover all the compounds in the test set.

The Q^2 value is good tests for evenly distributed data, but they are not always reliable for unevenly distributed datasets; instead RMSEs (Root Mean Squared Errors) provide a more reliable indication of the fitness of the model, independently of the applied splitting. Other useful parameters to be considered are the RMSEs calculated on different sets: on calibration or training (RMSEC), and prediction or test (RMSEP). RMSE is calculated as in Equation 4:

$$RMSE = \left(\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \right)^{1/2} \quad (4)$$

where y and $\hat{y}$ are respectively the measured and predicted values of the property; n is the number of compounds in each set of data.

It is important to note that RMSE values must not only be low but also as similar as possible for the training and test sets: this suggests that the proposed model has both predictive ability (low values) as well as sufficient generalization (similar values).^[33]

Furthermore, a variance inflation factor (VIF) analysis was performed to see if multicollinearities existed between the descriptors in the model. The VIF value is calculated from $(1/1-r^2)$, where r^2 is a multi correlation coefficient of one descriptor's effect regressed on the remaining molecular descriptors. Models are not accepted, if they contained descriptors with VIFs over a value of 10. This ensured that the squared multicollinearity coefficient for each descriptor in the model did not exceed 0.90.^[34-35] Finally, the model was validated using the external validation set.

Applicability Domain

A crucial problem of a QSPR model is the applicability domain (AD). Not even a robust, significant, and validated QSPR model can be expected to reliably predict the modeled property for the entire universe of chemicals. In fact, only the predictions for chemicals falling within this domain can be considered reliable and not model extrapolations.

A way of defining the AD of a QSPR model is according to the leverage of a compound. The leverage $h^{[36]}$ of a compound measures its influence on the model. The leverage of a compound in the original variable space is defined as:

$$h_i = x_i^T (X^T X)^{-1} x_i \quad (i=1, \dots, n) \quad (5)$$

where x_i is the descriptor vector of the considered compound and X is the model matrix derived from the calibration set descriptor values. The warning leverage^[37] is defined as follows:

$$h^* = 3 \times \sum_i h_i / n = 3 \times p' / n \quad (i=1, \dots, n) \quad (6)$$

where n is the number of training compounds and p' is the number of model parameters.

To visualize the AD of a QSPR model, the plot of standardized residuals versus leverage values (h) can be used for an immediate and simple graphical detection of both the response outliers (*i.e.*, compounds with cross validated standardized residuals greater than 2.5 standard deviation units, $>2.5 \sigma$) and structurally influential chemicals in a model ($h > h^*$).

Through the leverage approach, it is possible to verify whether a new chemical will lie within the structural model domain or outside the domain. A compound with high leverage in a QSPR model would fortify the model, if the compound is in the training set, but such a compound in the test set could have

unreliable predicted data, the result of substantial extrapolation of the model.^[38]

Results and Discussion

After the descriptors calculation, totally 35 most significant principal components calculated from each group of DRAGON molecular descriptors were used as variables to build a Kohonen map (6×6 neurons, 300 epochs). On the basis of the trained network, the compounds fell into different neuron of the top map (Figure 1). As a result, 40 compounds were included in the training set and 13 compounds fell into the test set. Figure 1 shows the Occupation of the 6×6 Kohonen top-map.

Figure 4 indicates molecular descriptors, experimental and predicted log*K* of macrocycle chelates for Na⁺, in this table the test set indicated with bold. The Forward stepwise regression was performed on the 40 macrocycles of the training set selected by SOM.

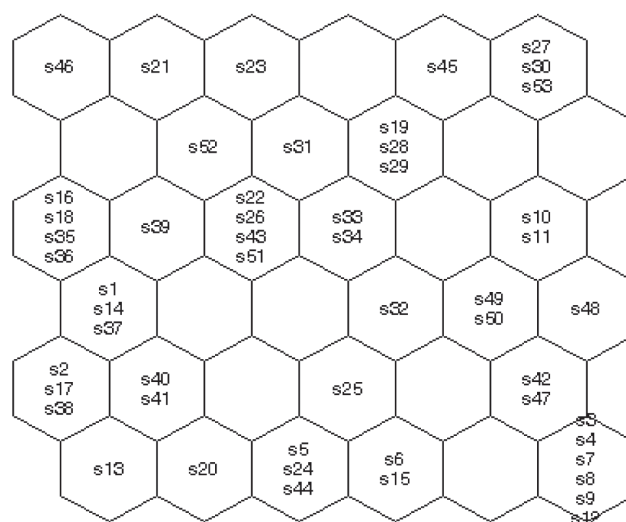


Figure 1. Occupation of the 6×6 Kohonen top-map.

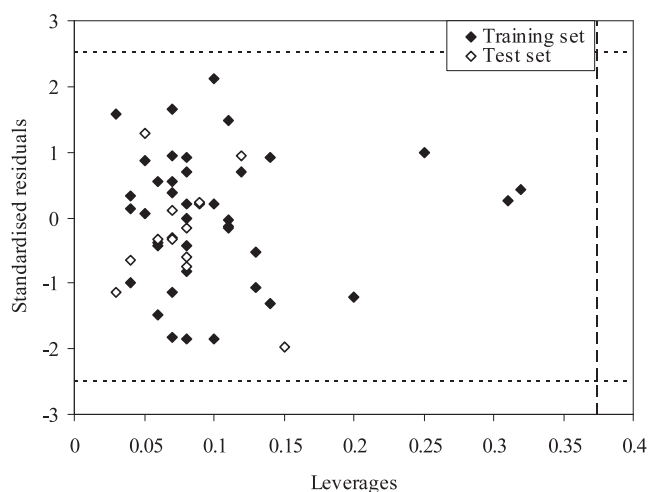


Figure 2. Williams plot of model. The training and test set samples are labeled differently. The dashed lines are the 2.5 limit and the warning value of hat ($h^* = 0.375$).

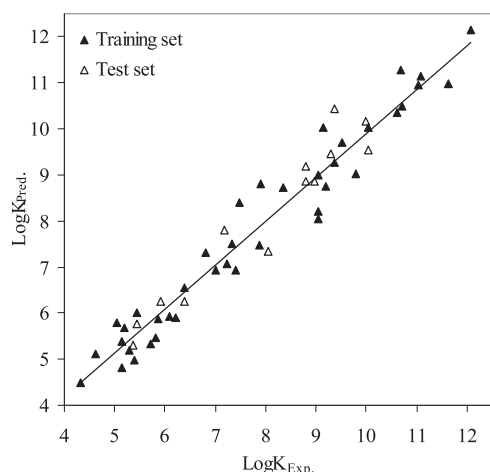


Figure 3. The plot of the predicted versus experimental $\log K$ values for training and test sets of chelates of macrocycles with Na^+ .

MLR Model and Applicability Domain

A four parameter linear model was obtained for prediction of association constant of macrocycle compounds. The best significant relationship for the $\log K$ of macrocycles has been deduced to be

$$\log K = -143.28(\pm 13.46) - 20.59(\pm 1.92)E1v + 3.09(\pm 0.19)IC5 - 57.65(\pm 5.20)LP1 - 0.24(\pm 0.04)RDF125m \quad (7)$$

$n_{\text{Calibration}}=40; n_{\text{Prediction}}=13$
 $Q^2_{\text{Calibration}}=0.95; Q^2_{\text{LOO}}=0.94; Q^2_{\text{Prediction}}=0.92$
 $RMSEC=0.469; RMSECV=0.528; RMSEP=0.455$

The MLR model results are given in Table 2; b and S_b are the nonstandardized coefficient of descriptors and standard error of coefficient, respectively, and b_s is the standardized regression coefficient. The molecular descriptors, descriptor type, definition of descriptors, and coefficient of descriptors are presented in Table 2. Table 3 indicates the linear correlation-coefficient matrix for $\log K$ and five descriptors in the MLR model. In Figure 3, the plot of predicted $\log K$ by the MLR model employed against the experimental $\log K$ is represented.

The inter-correlation of the descriptors used in the MLR model (Table 3) was low (below 0.66), which is in conformity to the study that for a statistically significant model, it is necessary that the descriptors included in the equation should not be inter-correlated with each other.^[39] To further check the inter-correlation of descriptors VIF analysis was performed. The VIF for each descriptor is summarized in Table 2. As one can see, the VIF values are all less than 1.3, indicating the stability of the equations constructed (according to statistics principle, a value of 1.0 is indicative of no correlation, while a value of under 10.0 is statistically satisfactory).^[34-35]

The applicability domain of the reported model was verified by an analysis of the Williams graph of Figure 2, in which the standardized residuals and the leverage value (h) are plotted. On analyzing the model AD in the Williams

plot of MLR model (Figure 2) all of chemicals belong to the applicability domain and there was not any compound in outlier.

Interpretation of the Descriptors

The standardized regression coefficients reveal the significance of an individual descriptor presented in the regression model. Obviously, in Table 2, the effect of IC5 on the association constant of the macrocycles is more significant than that of the other descriptors. The order of significance of the other descriptors is $LP1 > E1v > RDF125m$.

The most significant descriptor used in model is information content index (neighborhood symmetry of 5-order), IC5, describes the connectivity and branching in a molecule and can be related to molecular shape and symmetry.^[40-41] The positive regression coefficients for IC5 reflect the fact that macrocycles with higher branching and symmetry have stronger coordination ability that leads to higher association constant.

LP1 is among the most popular graph invariants and known as the Lovasz-pelkin index, also called leading eigenvalue.^[42] This eigenvalue has been suggested as an index of molecular branching, the smallest value corresponding to chain graphs and the highest to the most branched graphs. The alkyl groups are electron donating, thereby increasing the electron density and basicity of the adjacent donor atoms, these groups can increase the binding strength of macrocycles. Also when the number of alkoxy units increases, the cavity becomes smaller and more preorganized and binding constant increases for the small ions.^[43]

E1v is 1st component accessibility directional WHIM index that weighted by atomic van der Waals volumes, this descriptors can be related to the quantity of unfilled space per projected atom and has been called density (or emptiness); the greater the E1v value, the greater the projected unfilled space.^[44-46]

The less relevant descriptor $RDF125m$, added as the last variable in the QSPR model, is probably useful only to improve model quality in order to adapt some particular macrocycles.^[47-48]

Conclusions

A quantitative structure-property relationship model was derived to study the association constants of 53 sodium-macrocycle compounds at standard temperature (25°C) and just in CDCl_3 saturated with D_2O . Training and test data splitting was done with self-organizing map. The selection of the best variables from among the available descriptors was performed by forward stepwise regression, and resulted in the combination of E1v, IC5, LP1, and $RDF125m$ Dragon descriptors. The predictive ability of this combination of variables was with high Q^2_{ext} (0.92) and low RMSEP (0.45), which highlights the importance of these variables in modeling the studied property. The results from the current work provide a further tool for investigation of the complexation phenomenon.

Acknowledgement. This work is supported by Islamic Azad University, Kermanshah Branch, Kermanshah, Iran.

Table 1. The chemical structures of 53 macrocycles.

No.	Structure	No.	Structure
1		16	
2	$R=C_2H_5, X=H$ $R,X=CH_3$	17	$R=CH_3$
3		18	$R=CH_3$
4	$A=COCH_3, R_1=R_2=CH_3$	19	$R=CH_3$
5	$A=COCH_3, R_1=C(O)CH_3, R_2=CH_3$	20	$R=CH_3$
6	$A=COCH_3, R_1=Br, R_2=CH_3$	21	$R_1=CH_3, R_2=CH_2C_6H_5$
7	$A=COCH_3, R_1=NO_2, R_2=CH_3$		
8	$A=CCO_2CH_3, R_1, R_2=CH_3$		
9	$A=CNO_2, R_1, R_2=CH_3$		
10	$A=CNH_2, R_1, R_2=CH_3$		
11	$A=CSCH_3, R_1=CH_3, R_2=t-C_4H_9$ $A=CSOCH_3, R_1=CH_3, R_2=t-C_4H_9$		
12			
13			
14	$X=H$ $X=CH_2OCH_3$		
15			

Table 1. Continued.

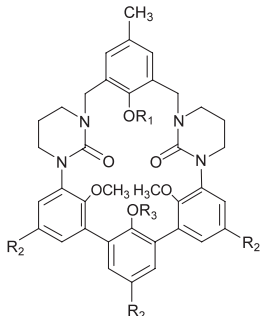
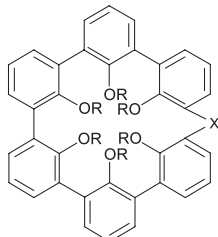
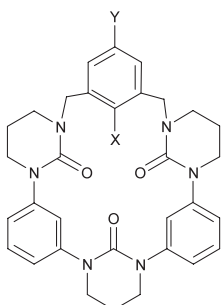
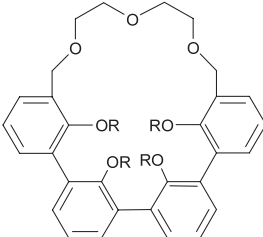
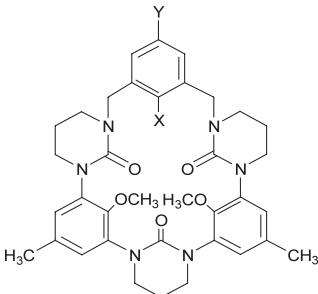
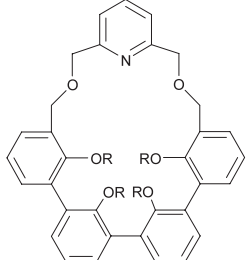
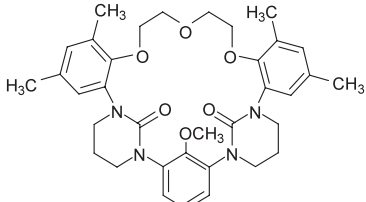
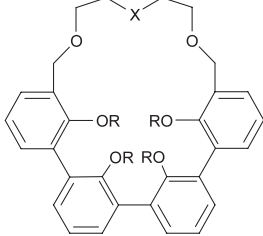
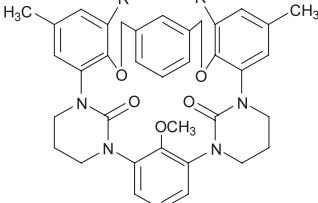
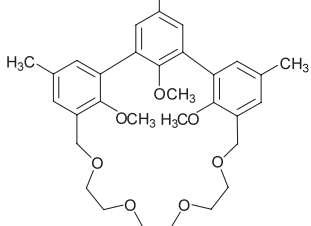
No.	Structure	No.	Structure
22	 $R_1, R_3 = \text{CH}_3$; $R_2 = \text{H}$ $R_1 = \text{CH}_2 = \text{CHCH}_2$; $R_2 = \text{CH}_3$; $R_3 = \text{CH}_2\text{C}_6\text{H}_5$	34	 $R = \text{CH}_3$; $X = \text{CH}_2\text{SO}_2\text{CH}_2$
23		35	
24	 $X, Y = \text{H}$ $X = \text{OCH}_3$; $Y = \text{CH}_3$	36	 $R = \text{CH}_3$ $R = \text{C}_2\text{H}_5$
25		37	
26	 $X = \text{Br}$; $Y = \text{H}$ $X = 3,5-(\text{C}_4\text{H}_9)_2-4-\text{CH}_3-\text{C}_6\text{H}_2$; $Y = \text{H}$ $X = \text{H}$; $Y = \text{C}_4\text{H}_9$ $X = \text{Br}$; $Y = \text{C}_4\text{H}_9$ $X = 3,5-(\text{C}_4\text{H}_9)_2-4-\text{CH}_3-\text{C}_6\text{H}_2$; $Y = \text{C}_4\text{H}_9$ $X = 3-\text{HOC}_6\text{H}_4$; $Y = \text{H}$	38	 $R = \text{CH}_3$ $R = \text{C}_2\text{H}_5$
27		39	
28	 $R = \text{CH}_3$; $X = \text{SO}$ $R = \text{CH}_3$; $X = \text{SO}_2$	40	 $R = \text{CH}_3$; $X = \text{SO}$ $R = \text{CH}_3$; $X = \text{SO}_2$
29		41	
30	 $R = \text{CH}_3$ $R = \text{CH}_2\text{OCH}_3$	42	
31			
32			
33			

Table 1. Continued.

No.	Structure	No.	Structure
43			Cryptahemispherands
44		47 48 49 50	$R = \text{CH}_3$; $X = \text{CH}_2\text{OCH}_2$; $Y = \text{CH}_2\text{CH}_x\text{OCH}_2\text{CH}_2$ $R = \text{C}_3\text{H}_7$; $X = \text{CH}_2\text{OCH}_2$; $Y = \text{CH}_2\text{CH}_2\text{OCH}_2\text{CH}_2$ $R = \text{C}_3\text{H}_7$; $X = 2,6\text{-C}_5\text{H}_3\text{N}$; CH_2OCH_2 ; $Y = \text{CH}_2\text{CH}_2\text{OCH}_2\text{CH}_2$ $R = \text{C}_3\text{H}_7$; $X = \text{CH}_2\text{OCH}_2$; $Y = \text{CH}_2\text{-}[2,6\text{-C}_5\text{H}_3\text{N}]\text{-CH}_2$
45		51 52	 $R_1 = \text{OCH}_3$; $R_2 = \text{CH}_3$ $R_1 = 3,5\text{-(CH}_3)_2\text{C}_6\text{H}_3$; $R_2 = \text{H}$
46		53	Bridged calix-4 $X = \text{CH}_2\text{CH}_2(\text{OCH}_2\text{CH}_2)_4$

Table 2. The MLR model results.

Variable	Descriptor type	Definition	b	S_b	b_s	VIF
Intercept	—	—	-143.2780	13.4651	—	—
E1v	WHIM descriptor	1st component accessibility directional WHIM index/weighted by atomic van der Waals volumes	-20.5893	1.9155	-0.4344	1.25
IC5	topological descriptor	Information content index (neighborhood symmetry of 5-order)	3.0883	0.1901	0.5930	1.02
LP1	topological descriptor	Lovasz-Pelikan index (Leading eigenvalue)	57.6530	5.2036	0.4539	1.28
RDF125m	RDF descriptors	Radial Distribution function-12.5/ weighed by atomic masses	-0.2370	0.0378	-0.2354	1.08

Table 3. Linear correlation-coefficient matrix for the four descriptors and log*K* in the MLR model.

	Elv	IC5	LP1	RDF125m	log <i>K</i>
Elv	1.00				
IC5	-0.13	1.00			
LP1	-0.43	0.05	1.00		
RDF125m	-0.17	0.07	0.26	1.00	
log <i>K</i>	-0.66	0.66	0.61	0.00	1.00

Table 4. Molecular descriptors, experimental and predicted log*K* of macrocycle chelates for Na⁺ in CDCl₃ saturated with D₂O at 25°C; test set are indicated in bold.

No.	Elv	IC5	LP1	RDF125m	Experimental	Predicted	Residual
1	0.365	4.833	2.509	0.112	9.20	8.76	0.44
2	0.425	4.715	2.506	0.334	7.00	6.93	0.07
3	0.370	4.945	2.512	0.030	8.79	9.19	-0.40
4	0.372	4.739	2.518	0.072	8.96	8.85	0.11
5	0.369	4.601	2.510	0.032	9.04	8.04	1.00
6	0.391	4.656	2.518	0.030	9.04	8.21	0.83
7	0.401	4.851	2.525	0.030	9.04	9.01	0.03
8	0.403	4.733	2.503	0.000	8.04	7.35	0.69
9	0.405	4.733	2.503	0.000	6.81	7.31	-0.50
10	0.387	4.578	2.538	1.769	7.91	8.80	-0.89
11	0.392	4.615	2.541	2.875	8.34	8.72	-0.38
12	0.384	5.343	2.490	0.002	8.79	8.87	-0.08
13	0.508	4.614	2.503	0.000	5.15	4.82	0.33
14	0.492	4.805	2.504	0.098	5.45	5.77	-0.32
15	0.453	5.099	2.468	0.365	5.72	5.34	0.38
16	0.421	4.430	2.519	0.161	7.40	6.92	0.48
17	0.483	4.767	2.503	0.104	5.04	5.78	-0.74
18	0.423	4.953	2.506	0.007	7.18	7.79	-0.61
19	0.443	5.121	2.531	1.337	9.80	9.02	0.78
20	0.453	4.959	2.484	0.000	6.08	5.92	0.16
21	0.443	5.736	2.528	0.365	11.64	10.98	0.66
22	0.397	5.433	2.516	0.110	10.62	10.36	0.26
23	0.402	5.758	2.538	1.752	12.08	12.14	-0.06
24	0.463	5.205	2.464	0.080	5.36	5.30	0.06
25	0.440	5.306	2.473	0.264	6.38	6.56	-0.18
26	0.392	5.242	2.522	0.281	9.99	10.17	-0.18
27	0.366	5.311	2.535	5.530	9.38	10.43	-1.05
28	0.373	5.242	2.522	4.642	10.04	9.53	0.51
29	0.374	5.242	2.524	0.943	10.70	10.50	0.20
30	0.385	5.346	2.545	4.58	11.04	10.95	0.09
31	0.367	5.546	2.525	3.009	11.08	11.15	-0.07
32	0.480	5.210	2.509	0.381	7.34	7.49	-0.15
33	0.435	5.360	2.51	8.252	7.23	7.07	0.16
34	0.439	5.395	2.512	7.161	7.88	7.47	0.41
35	0.449	4.302	2.524	0.126	6.38	6.25	0.13
36	0.451	4.243	2.521	0.021	5.86	5.88	-0.02
37	0.523	4.824	2.507	0.007	5.15	5.39	-0.24
38	0.463	4.797	2.503	0.193	5.91	6.26	-0.35
39	0.522	4.915	2.507	0.022	5.20	5.68	-0.48
40	0.488	4.708	2.503	0.17	5.82	5.48	0.34
41	0.470	4.721	2.503	0.067	6.20	5.91	0.29
42	0.477	4.673	2.51	0.084	5.45	6.02	-0.57
43	0.435	5.336	2.512	2.877	7.48	8.39	-0.91
44	0.476	5.045	2.463	0.02	4.32	4.50	-0.18
45	0.457	4.089	2.522	1.58	5.38	4.97	0.41
46	0.399	4.362	2.516	7.784	5.28	5.19	0.09
47	0.367	4.929	2.513	0.012	9.38	9.27	0.11
48	0.359	5.086	2.515	0.006	9.15	10.03	-0.88
49	0.411	5.315	2.516	0.09	9.52	9.71	-0.19
50	0.340	5.315	2.518	0.091	10.69	11.28	-0.59
51	0.439	5.428	2.516	0.147	9.30	9.47	-0.17
52	0.429	5.521	2.518	0.304	10.04	10.04	0.00
53	0.386	4.043	2.512	4.096	4.62	5.11	-0.49

References

1. Steed J.W., Atwood J.L. *Supramolecular Chemistry*. Wiley, Chichester; New York, **2000**.
2. Lehn J.M. *Science* **1985**, 227, 849-856.
3. Muller-Dethlefs K., Hobza P. *Chem. Rev.* **2000**, 100, 143-167.
4. Schneider H.-J., Yatsimirsky A.K. *Principles and Methods in Supramolecular Chemistry*. Wiley, New York, **2000**.
5. Bianchi A., Bowman-James K., García-España E. *Supramolecular Chemistry of Anions*. Wiley-VCH, New York, **1997**.
6. Tzalis D., Tor Y. *Tetrahedron Lett.* **1996**, 37, 8293-8296.
7. Pirkle W.H., Welch C.J. *Tetrahedron-Asymmetr.* **1994**, 5, 777-780.
8. Barak D., Ordentlich A., Segall Y., Velan B., Benschop H.P., DeJong L.P.A., Shafferman A. *J. Am. Chem. Soc.* **1997**, 119, 3157-3158.
9. Cram D.J., Ho S.P. *J. Am. Chem. Soc.* **1986**, 108, 2998-3005.
10. Cram D.J., Ho S.P., Knobler C.B., Maverick E., Trueblood K.N. *J. Am. Chem. Soc.* **1986**, 108, 2989-2998.
11. Cram D.J., Lein G.M., Kaneda T., Helgeson R.C., Knobler C.B., Maverick E., Trueblood K.N. *J. Am. Chem. Soc.* **1981**, 103, 6228-6232.
12. Lein G.M., Cram D.J. *J. Am. Chem. Soc.* **1985**, 107, 448-455.
13. Oude Wolbers M.P., Van Veggel F.C.J.M., Snellink-Ruël B.H.M., Hofstraat J.W., Geurts F.A.J., Reinhoudt D.N. *J. Am. Chem. Soc.* **1997**, 119, 138-144.
14. Pirkle W.H., Welch C.J. *Tetrahedron-Asymmetr.* **1994**, 5, 777-780.
15. Cram D.J., Trueblood K.N. *Top. Curr. Chem.* **1981**, 98, 43-106.
16. Pedersen C.J. *J. Am. Chem. Soc.* **1967**, 89, 2495-2496.
17. Gutsche C.D., Muthukrishnan R. *J. Org. Chem.* **1978**, 43, 4905-4906.
18. Izatt R.M., Lamb J.D., Hawkins R.T., Brown P.R., Izatt S.R., Christensen J.J. *J. Am. Chem. Soc.* **1983**, 105, 1782-1785.
19. Izatt S.R., Hawkins R.T., Christensen J.J., Izatt R.M. *J. Am. Chem. Soc.* **1985**, 107, 63-66.
20. Dijkstra P.J., Brunink J.A., Bugge K.E., Reinhoudt D.N., Harkema S., Ungaro R., Ugozzoli F., Ghidini E. *J. Am. Chem. Soc.* **1989**, 111, 7567-7575.
21. Ahmadi S. *J. Incl. Phenom. Macrocycl. Chem.* (published online 19.10.2010). doi: 10.1007/s10847-010-9881-6.
22. Ghasemi J., Saaidpour S. *J. Incl. Phenom. Macrocycl. Chem.* **2008**, 60, 339-351.
23. Li Y., Su L., Zhang X., Huang X., Zhai H. *Chemometr. Intell. Lab.* **2011**, 105, 106-113.
24. Ghasemi J.B., Ahmadi S., Ayatia M. *Macroheterocycles* **2010**, 3, 234-242.
25. Izatt R.M., Pawlak K., Bradshaw J.S., Bruening R.L. *Chem. Rev.* **1991**, 91, 1721-2085.
26. Hyperchem, v.7.5. Hypercube Inc. **2002**, <http://www.hyper.com>.
27. Dewar M.J.S., Zoebisch E.G., Healy E.F., Stewart J.J.P. *J. Am. Chem. Soc.* **1993**, 115, 5348-5348.
28. Stewart J.J.P. Mopac 6.0, Quantum Chemical Program Exchange 455, **1990**.
29. Talete S. Dragon for Windows (Software for Molecular Descriptor Calculations), Version 5.4, **2006**, <http://www.talete.mi.it/>.
30. Zupan J., Gasteiger J. *Neural Networks in Chemistry and Drug Design*. 2nd ed., Wiley-VCH, Weinheim; New York, **1999**.
31. Zupan J., Novic M., Ruisanchez I. *Chemometr. Intell. Lab.* **1997**, 38, 1-23.
32. Draper N.R., Smith H. *Applied Regression Analysis*. 2d ed., Wiley, New York, **1981**.
33. Guha R., Serra J.R., Jurs P.C. *J. Mol. Graph. Model.* **2004**, 23, 1-14.
34. Jaiswal M., Khadikar P.V., Scozzafava A., Supuran C.T. *Bioorg. Med. Chem. Lett.* **2004**, 14, 3283-3290.
35. Shapiro S., Guggenheim B. *Quant. Struct.-Act. Rel.* **1998**, 17, 327-337.
36. Atkinson A.C. *Plots, Transformations and Regression*. Oxford, Clarendon Press, **1985**.
37. Jaszay Z., Pham T.S., Nemeth G., Bako P., Petnehazy I., Toke L. *Synlett* **2009**, 1429-1432.
38. Gramatica P. *Qsar Comb. Sci.* **2007**, 26, 694-701.
39. Deswal S., Roy N. *Eur. J. Med. Chem.* **2006**, 41, 1339-1346.
40. Kier L.B. *J. Pharm. Sci.* **1980**, 69, 807-810.
41. Bonchev D. *Information Theoretic Indices for Characterization of Chemical Structures*. Research Studies Press, Chichester West Sussex; New York, **1983**.
42. Lovász L., Pelikán J. *Perio. Math. Hung.* **1973**, 3, 175-182.
43. Elroby S.A.K., Lee K.H., Cho S.J., Hinchliffe A. *Can. J. Chem.* **2006**, 84, 1045-1049.
44. Gramatica P., Navas N., Todeschini R. *Chemometr. Intell. Lab.* **1998**, 40, 53-63.
45. Todeschini R., Gramatica P. *Quant. Struct.-Act. Rel.* **1997**, 16, 113-119.
46. Todeschini R., Gramatica P. *Quant. Struct.-Act. Rel.* **1997**, 16, 120-125.
47. Hemmer M.C., Steinhauer V., Gasteiger J. *Vib. Spectrosc.* **1999**, 19, 151-164.
48. Hemmer M.C., Gasteiger J. *Anal. Chim. Acta* **2000**, 420, 145-154.

Received 04.07.2011

Accepted 10.01.2012