

## QSPR Modeling of Stability Constants of the Li–Hemispherands Complexes Using MLR: A Theoretical Host–Guest Study

Jahan B. Ghasemi,<sup>a@</sup> Shahin Ahmadi,<sup>b</sup> and Mahnaz Ayati<sup>a</sup>

<sup>a</sup>Chemistry Department, Faculty of Science, K.N. Toosi University of Technology, Tehran, Iran

<sup>b</sup>Chemistry Department, Faculty of Science, Razi University, Kermanshah, Iran

@Corresponding author E-mail: Jahan.ghasemi@gmail.com

*A multiple linear regression (MLR) model in quantitative structure property relationship (QSPR) was developed for predicting stability constant of 56 complexes of hemispherands with lithium at standard temperature (25°C) and just in CDCl<sub>3</sub> saturated with D<sub>2</sub>O. A large number of descriptors were calculated with Dragon software and a subset of calculated descriptors was selected from 6 classes of Dragon descriptors with a forward stepwise regression as a feature selection technique. Five calculated descriptors were selected as the most feasible descriptors. The data was randomly divided to the training and prediction. The predictive ability of the model was evaluated using leave-one-out (LOO) cross-validation method. The obtained model showed high prediction ability with root mean square error of prediction (RMSEP), 0.65, and square correlation coefficient (R<sup>2</sup>) of 0.920.*

**Keywords:** Quantitative structure property relationship (QSPR), molecular descriptor, hemispherands, host-guest complex, stability constant.

### Introduction

Supramolecular chemistry and the quantification of non-covalent interaction (electrostatic forces, hydrogen bonding and van der Waals interactions) offer the basis for novel approaches in medicine, host-guest chemistry,<sup>[1-6]</sup> chromatography, and biocatalysis.<sup>[8]</sup> The development of artificial receptors capable of forming complexes with specific guests<sup>[9-13]</sup> is important to the progress of supramolecular host-guest chemistry.

In host-guest complexation, the concave surface of a host complements the convex surface of a guest. Similarly, the receptor sites of enzymes frequently contain rigid cavities whose internal surfaces complement the convex surfaces of substrates or inhibitors.

The conformational ambiguity of the crown ethers may be reduced by restricting rotation about single bonds in the crown ether macrocycle either by fusion of an aromatic ring. The latter strategy has been used most effectively by Cram and his co-workers in the increasingly pre-organized hemispherand and spherand systems.<sup>[14]</sup>

The principle of preorganization is that the logK for host-guest complex formation is increased significantly, if the host and guest are organized for binding and have low solvation prior to complexation. This principle was experimentally demonstrated by the synthesis of spherands designed to complex selectively with Li<sup>+</sup> and Na<sup>+</sup> cations.<sup>[15]</sup> Hemispherands can be rigidified further by incorporating extra bridges which sometimes contain additional binding sites.<sup>[16-18]</sup> The majority of preorganized macrocycles form very stable complexes with target cations in at least some solvents, and part of them show significant selectivity.<sup>[19]</sup>

The efficiency and selectivity, with which these compounds can complex a cation, relies on a number of factors, *e.g.*: (i) shape and preorganization within the host molecule, (ii) the size-match of the host cavity to the guest,<sup>[20]</sup> (iii) cation charge and type, and (iv) donor atom charge and type. Each of these factors is discussed more detailed below.

The principle of preorganization states that host-guest binding is the strongest when only very small changes in organization of host, guest and solvent are required for complexation.<sup>[21]</sup> Alkali metal cations are spherical; therefore, they prefer a spherical donor atom array in the host compound. In order to achieve this geometry some degree of preorganization in the host molecule is required.

Pedersen was the first who suggested the "size-fit" principle in his seminal paper of 1968.<sup>[20]</sup> He has found that the crown-4 hosts selectively bind Li<sup>+</sup>, crown-5 hosts selectively bind Na<sup>+</sup>, and crown-6 systems complex selectively with K<sup>+</sup>. He has concluded that these results reflect the relative size of each cation *vis a vis* the size and shape of the host cavity, in each case. The size-fit principle does have its limitations; cations too large or too small to fit within a cavity may still be complexed but not necessarily completely within the cavity or on a 1:1 host:guest basis. More complex host-guest relationships are known, for example 2:1 "sandwich-type" complexes. In such cases, other factors such as cation charge and type, ligand donor atom type, ligand substituents and solvent become important.<sup>[22]</sup>

Pearson's hard-soft acid-base (HSAB) principle<sup>[23]</sup> provides a useful starting point for donor atom selection when designing a host for a particular guest. The principle states that hard acids prefer to bind to hard bases and soft

acids prefer to bind to soft bases, where hard acids have small highly charged and nonpolarisable acceptor atoms and soft acids are larger and not so highly charged. Correspondingly hard bases have small, highly electronegative donor atoms and soft bases are larger and more polarisable.<sup>[24]</sup> The bonding between hard acids and hard bases is dominated by electrostatic interactions, whereas the bonding between soft acids and soft bases is primarily covalent in nature. Crown ethers utilizing ether oxygens as donor atoms, which are "hard base" species, coordinate well with the hard acid, alkali metal cations.

Solovev and Varnek reported the QSPR model of substructural molecular fragments based on splitting of a molecular graph into a limited number of topological fragments and calculation of their contributions to a given property, stability constants of complexes of crown ethers with alkali metal cations in methanol was calculated.<sup>[25]</sup> Svetlitski *et al.* developed QSPR models for the stability constants of complexes between 63 different organic ligands and 14 lanthanides.<sup>[26]</sup> In our previous molecular modeling works, we used the application of QSPR techniques in the development of a new, simplified approach to prediction of compounds properties with different techniques.<sup>[27-31]</sup> Recently, for the first time, we have established the QSPR models for stability constants of 58 complexes of 15C5 crowns with potassium ion.<sup>[32]</sup> The aim of the present study, is the development of predictive QSPR model of stability constant for lithium complexes with hemispherands. This model enables to make reliable predictions of the stability constant for previously unknown complexes and to elucidate the structural factors determining the stability constants.

## Materials and Methods

### Data Set

The chemical structures and experimental values for the stability constants of fifty nine sphere derivatives taken from the literature<sup>[33]</sup> are presented in Tables 1 and 4, respectively. The data set was split into a training set and a testing set randomly. The training set of 47 complexes was used to adjust the parameters of the models, and the test set of 10 complexes 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 CDCl<sub>3</sub> saturated with D<sub>2</sub>O.

### Molecular Optimization and Descriptor Calculation

All calculations were run on a Pentium IV personal computer with windows XP as operating system. The CS ChemOffice 2005 molecular modeling software ver. 9, supplied by Cambridge Software Company, was employed for optimization of the structure of the molecules and calculation of descriptors. The molecular structures of data set were sketched using ChemDraw Ultra module of this software. The sketched structures were exported to Chem3D module in order to create their 3D structures. Each molecule was "cleaned up" and energy minimization was performed using Allinger's MM2 force field and further geometry optimization was done using semi-empirical quantum method Austin Method 1 (AM1)<sup>[34]</sup> using the Polack-Rabiere algorithm until the root mean square gradient was 0.01. More than 350 molecular descriptors is derived to properly characterize the chemical structure of the 56 macrocycles, 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. These variables are calculated by means of the software Dragon version 3.0 available in the Web.<sup>[35]</sup>

### Forward Stepwise Regression

The forward stepwise regression procedure<sup>[36]</sup> 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:

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

with *d* being the number of descriptors of the model, *N* is the number of molecules of the training set, and *res<sub>i</sub>* 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. According to minimal *S* the optimized number of descriptors was selected.

## Results and Discussion

### MLR Model

The multiple linear regression (MLR) was performed on the macrocycles of the training set shown in Table 4. The five descriptors values for 56 hemispherands are indicated in Table 3. A five parameter model was obtained for prediction of stability constant of macrocycle compounds. The equation with its usual statistical parameters is as follows;

$$y = -1.04 \times 10^2 (\pm 11) + 2.38 \times 10^{-02} (\pm 1.16 \times 10^{-02}) T(N \dots O) + 44 (\pm 5) LP1 - 1.9 (\pm 0.4) Mor32m - 1.6 (\pm 0.3) Mor30v - 5.20 \times 10^{-04} (\pm 1.31 \times 10^{-04}) piPC07 \quad (2)$$

*n*=45, *S*<sup>2</sup>=0.421, *R*<sup>2</sup>=0.953, *F*=161 (at 95% level)

The results of the model are summarized in Table 2. The predicted values of log*K* by using regression model are represented in Table 4. In Table 2 and Equation 2, *b* and *S<sub>b</sub>* are the nonstandardized coefficient of descriptors and standard error of coefficient, respectively, and *b<sub>s</sub>* is the standardized regression coefficient. The predicted and experimental log*K* and the residuals (experimental log*K* vs predicted log*K*), obtained by the BMLR modeling, are shown in Table 4.

The standardized regression coefficients reveal the significance of an individual descriptor presented in the regression model. Obviously, in Table 4, sum of topological distance N...O on the stability constant of the macrocycles is more significant than that of the other descriptors. The order of significance of the other descriptors is LP1 > Mor32m > piPC07 > Mor30v.

As there is more than one variable presented in the correlations, it is necessary to examine the stability of our

regression. Upon investigating the collinearity of variables in, we obtain the variance inflation factor (VIF) for each descriptor which is summarized in Table 2.

$$\text{VIF} = 1/(1 - R_{xi}^2) \quad (3)$$

where  $R_{xi}^2$  is the multiple correlation of the equation when dependent variable is replaced with one independent variable  $x_i$ . This indicator reflects the extent of the collinearity of the independent variables or descriptors in QSAR studies.

As one can see, all the VIF values are less than 3.0, 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).

### Validation of the Model

Validation is a crucial aspect of any QSAR/QSPR model to prove the predictability of the model obtained. Golbraikh and Tropsha<sup>[37]</sup> suggested that for a QSAR/QSPR model to have a high predictive power the following conditions should be fulfilled: (i) a high cross-validated  $Q^2$  value; (ii) correlation coefficient  $R$  between the predicted and observed activities/properties of compounds from an external test set close to 1; (iii) at least one (but better both) of the correlation coefficients for regression through the origin (predicted versus observed activities, or observed versus predicted activities) should be close to  $R^2$ ; (iv) at least one slope of the regression lines through the origin should be close to 1. In this paper, the cross-validated  $R^2$  (i.e.  $Q^2$ ) was assessed and the high value of  $Q^2$  can be regarded as the proof of the high predictive ability of the model.  $Q^2 = [1 - \text{PRESS}/\text{SS}]$ , where PRESS is the predicted residual sum of squares of deleted data in the cross-validation, calculated as  $[\text{predicted value} - \text{experimental value}]^2$ , and SS is the sum of squares of  $Y$  corrected for the mean, calculated as  $[\text{experimented value} - \text{mean experimental value}]^2$ . To assess the robustness of the model we use many-leave-out and the result has good alignment with leave-one-out cross validation  $Q^2$ , proves good statistical quality of the final model, Table 5. However, the high value of  $Q^2$  appears to be a necessary but not sufficient condition for the models to have a high predictive power. So an external set, i.e. test set, is necessary besides  $Q^2$ . In all cases, the  $Q^2$ , root mean square error of the calibration set (RMSEC), were calculated to assess the quality of the models. An external test set was chosen to verify the MLR model, and the root mean square error of prediction set (RMSEP) was also examined. Table 5 indicates the statistical parameters of the model.

### Interpretation of Descriptors

As we discussed in introduction section, the process of cation-macrocycle association depends on several factors including shape and preorganization within the host molecule, the size-match of the host cavity to the guest, cation charge and type, donor atom charge and type, the nature of the solvent and *etc.* Here to eliminate the effect of temperature and solvent on the stability constants, we used only data at

25 °C and just in  $\text{CDCl}_3$  saturated with  $\text{D}_2\text{O}$ .

The cavity size of hemispherands that used in this study is large (more than 4 Å),<sup>[38]</sup> but the ionic diameter of the lithium cation (1.20 Å) is smaller than the cavity of the macrocycles, so as might be expected,  $\text{Li}^+$  is waterionized and bound inside the cavity. The stoichiometry of the binding of  $\text{Li}^+$  (M) with macrocycles (L) studied in the present work is 1:1 (ML). The electron pairs of donor groups in these macrocycles can be preorganized to bind the smaller alkali metal cation such as  $\text{Li}^+$ . So the flexibility of these compounds is very important for complexation process.<sup>[38]</sup> It is noted that the other parameter such as size of the macrocycle cavity, charge density of ions (i.e., coulombic interactions) as well as hydrophobic interaction play a major role in governing the occurrence and stability of the complexed species.

The descriptors involved in the QSPR model are:

- (i)  $T(\text{N}\cdots\text{O})$ , sum of topological distance between  $\text{N}\cdots\text{O}$ ,
- (ii) LP1, Lovasz-Pelkin index (leading eigenvalue),
- (iii) Mor32m, 3D-MoRSE-signal 32/weighted by atomic masses,
- (iv) Mor30v, 3D-MoRSE-signal 30/weighted by atomic van der Waals volumes,
- (v) piPC07, Molecular path count of order 7.

In this QSPR model according to the selected descriptors, complexation phenomenon is mainly related to: (i) topological ( $T(\text{N}\cdots\text{O})$ , LP1, and piPC07), and (ii) 3D-Morse descriptors (Mor32m and Mor30v). As you can see in Table 2, the positive sign of regression coefficients of  $T(\text{N}\cdots\text{O})$  and LP1 in model show that with  $T(\text{N}\cdots\text{O})$  and LP1 increasing the  $\log K$  will be increased, and the negative sign of regression coefficients of Mor32m, Mor30v, and piPC07 indicate that increasing these descriptors will decrease the extent of  $\log K$ .

The most significant descriptor involved in the model is “sum of topological distance between  $\text{N}\cdots\text{O}$ ”,<sup>[39]</sup> the topological indices is calculated by either the adjacency matrix or the topological distance matrix. The topological distance between two vertices is the number of edges in the shortest path between these. The molecule structures (Table 1) indicate that with increasing of nitrogen atom size the oxygen with double bond is increased and oxygen with double bond is harder than oxygen with single bond,<sup>[40]</sup> and, based on HSAB principle, it can strongly bind to lithium. The positive sign of this descriptor confirms our concluding, and with increasing of  $\text{N}\cdots\text{O}$  the stability constant of macrocycle increases.

The second descriptor used in model is the largest eigenvalue of adjacency matrix  $A$ , it is among the most popular graph invariants and known as the Lovasz-Pelkin index, also called leading eigenvalue.<sup>[41]</sup> 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.<sup>[38]</sup>

The 3D-MoRSE descriptors<sup>[42,43]</sup> provide three-dimensional information from the three-dimensional coordinates by using the same transform as in electron

**Table 1.** Chemical structures of 56 hemispherands.

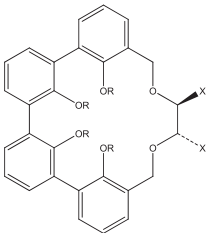
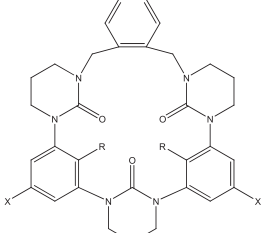
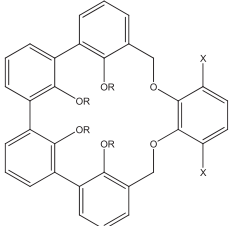
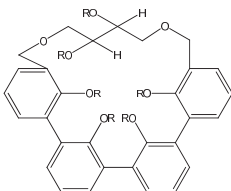
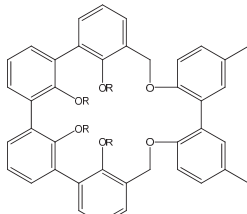
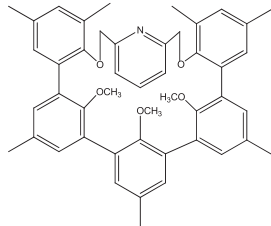
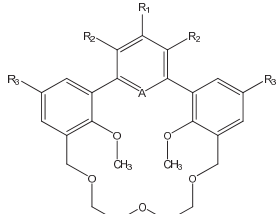
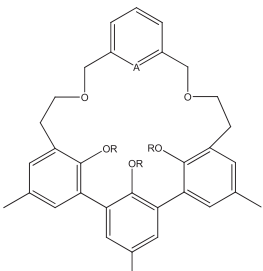
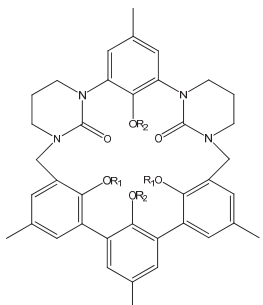
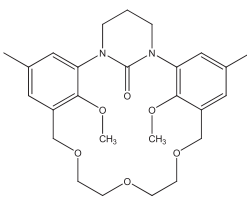
No.	Structure	No.	Structure
1	 R=CH <sub>3</sub> , X=CH <sub>2</sub> OCH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	17	 R=H, X=H
2	 R=CH <sub>3</sub> , X=H	18	 R=CH <sub>3</sub>
3	R=C <sub>2</sub> H <sub>5</sub> , X=H	19	
4	R=CH <sub>3</sub> , X=CH <sub>3</sub>	20	
5	 A=COC <sub>3</sub> H <sub>7</sub> , R <sub>1</sub> , R <sub>3</sub> =CH <sub>3</sub> , R <sub>2</sub> =H	21	 A=N, R=CH <sub>3</sub>
6	A=COH, R <sub>1</sub> , R <sub>3</sub> =CH <sub>3</sub> , R <sub>2</sub> =H	22	 R <sub>1</sub> =CH <sub>3</sub> , R <sub>2</sub> =CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>
7	A=COCH <sub>3</sub> , R <sub>1</sub> =C(O)CH <sub>3</sub> , R <sub>2</sub> =H, R <sub>3</sub> =CH <sub>3</sub>		
8	A=COCH <sub>3</sub> , R <sub>1</sub> =NO <sub>2</sub> , R <sub>2</sub> =H, R <sub>3</sub> =CH <sub>3</sub>		
9	A=CNO <sub>2</sub> , R <sub>1</sub> =C <sub>6</sub> H <sub>5</sub> , R <sub>2</sub> =H, R <sub>3</sub> =CH <sub>3</sub>		
10	A=CCO <sub>2</sub> CH <sub>3</sub> , R <sub>1</sub> , R <sub>3</sub> =CH <sub>3</sub> , R <sub>2</sub> =H		
11	A=CNH <sub>2</sub> , R <sub>1</sub> , R <sub>3</sub> =CH <sub>3</sub> , R <sub>2</sub> =H		
12	A=CSOCH <sub>3</sub> , R <sub>1</sub> =CH <sub>3</sub> , R <sub>2</sub> =H, R <sub>3</sub> =t-C <sub>4</sub> H <sub>9</sub>		
13	A=CSO <sub>2</sub> CH <sub>3</sub> , R <sub>1</sub> =CH <sub>3</sub> , R <sub>2</sub> =H, R <sub>3</sub> =t-C <sub>4</sub> H <sub>9</sub>		
14			
15	R=CH <sub>3</sub> , X=CH <sub>2</sub> OCH <sub>3</sub>		
16	R=C <sub>2</sub> H <sub>5</sub> , X=H		

Table 1. (continued)

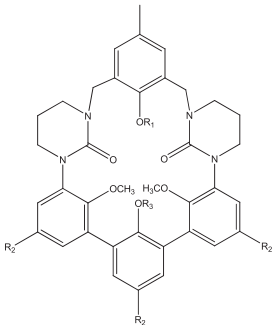
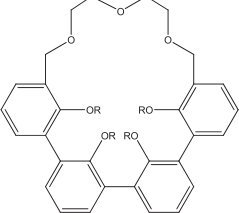
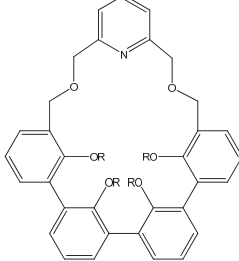
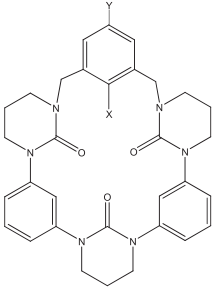
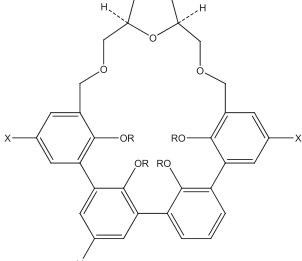
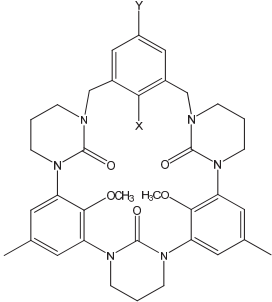
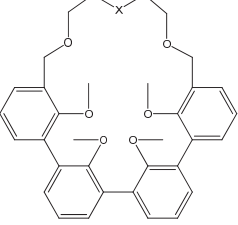
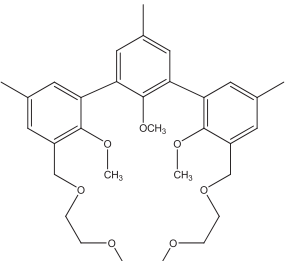
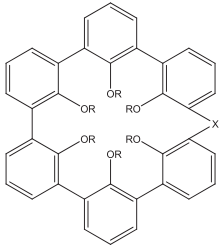
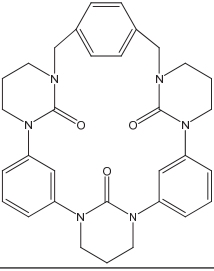
No.	Structure	No.	Structure
23		38	
24	$R_1, R_2=H, R_3=CH_3$	39	$R=CH_3$
25	$R_1=H, R_2=CH_3, R_3=CH_2C_6H_5$		$R=C_2H_5$
26	$R_1=CH_2CHCH_2, R_2=CH_3, R_3=CH_2C_6H_5$	40	
27			$R=CH_3$
28	$X, Y=H$	41	
29	$X=OCH_3, Y=CH_3$		$R=C_2H_5, X, Y=H$
30		42	
31	$X, Y=H$	43	$X=S$
32	$X=Br, Y=H$	44	$X=SO$
33	$X=3,5-(t-C_4H_9)_2-4-CH_3-OC_6H_2, Y=H$		$X=SO_2$
34	$X=9\text{-Anthracenyle}, Y=H$	45	
35	$X=H, Y=t-C_4H_9$		
36	$X=Br, Y=t-C_4H_9$	46	
37	$X=3,5-(t-C_4H_9)_2-4-CH_3-OC_6H_2, Y=t-C_4H_9$		
38	$X=3-HOC_6H_4, Y=H$		
39	$R=CH_3, X=CH_2SO_2CH_2$		
40	$R=CH_3, X=CH_2SCH_2$		

Table 1. (continued)

No.	Structure	No.	Structure
47		53	
48		54	
49	R=CH <sub>3</sub>	55	R <sub>1</sub> =OCH <sub>3</sub> , R <sub>2</sub> =CH <sub>3</sub> , R <sub>3</sub> =H
50		56	R <sub>1</sub> =3,5-(CH <sub>3</sub> ) <sub>2</sub> C <sub>6</sub> H <sub>3</sub> , R <sub>2</sub> , R <sub>3</sub> =H
51	R=H, X=CH <sub>2</sub> OCH <sub>2</sub> , Y=CH <sub>2</sub> CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub>		
52	R=C <sub>3</sub> H <sub>7</sub> , X=CH <sub>2</sub> OCH <sub>2</sub> , Y=CH <sub>2</sub> CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub>		R=CH <sub>3</sub> , X=CH <sub>2</sub> CH <sub>2</sub> (OCH <sub>2</sub> CH <sub>2</sub> ) <sub>4</sub>
	R=C <sub>3</sub> H <sub>7</sub> , X=2,6-C <sub>5</sub> H <sub>3</sub> N		

Table 2. The best MLR model results.

Variable	Description of molecular descriptor	<i>b</i>	<i>S<sub>b</sub></i>	<i>b<sub>s</sub></i>	VIF
Intercept	–	-1.04·10 <sup>2</sup>	11	–	–
T(N··O)	Sum of topological distance between N··O	2.38·10 <sup>-2</sup>	1.16·10 <sup>-2</sup>	0.92	2.061
LP1	Lovasz-Pelkin index (leading eigenvalue)132	44	5	0.45	2.19
Mor32m	3D-MorSE-signal 32/weighted by atomic masses	-1.9	0.4	-0.33	1.95
Mor30v	3D-MorSE-signal 30/weighted by atomic van der Waals volumes	-1.6	0.3	-0.18	1.33
piPC07	Molecular path count of order 7 345	-5.20·10 <sup>-4</sup>	1.31·10 <sup>-4</sup>	-0.20	2.62

diffraction (which uses it to prepare theoretical scattering curves). Various atomic properties can be taken into account giving high flexibility to this representation of a molecule. The descriptor equation is based on the general molecular transform and is given by:

$$I(s) = \sum_{i=2}^N \sum_{j=1}^{i-1} A_i A_j \frac{\sin sr_{ij}}{sr_{ij}}$$

$$s = 0, \dots, 31.0 \text{ \AA}^{-1}$$

where  $I(s)$  is the scattered electron intensity,  $A$  is an atomic property chosen as atomic number and  $r_{ij}$  is the interatomic distances between  $i^{\text{th}}$  and  $j^{\text{th}}$  atoms. The values of this function were calculated at 32 evenly distributed values of  $s$  in the range of 0-31.0  $\text{\AA}^{-1}$  from the three-dimensional atomic coordinates of a molecule as obtained by the 3D structure generator CORINA.<sup>[44]</sup>

Mor32m and Mor30v are a representation of the three-dimensional structure of organic molecules that can be probably indicate cavity size and shape of molecule that

**Table 3.** The descriptors values for 56 hemispherands.

No.	T(N···O)	LP1	Mor32m	Mor30v	piPC07
1	0	2.504	-0.891	0.584	2553.930
2	0	2.505	-0.990	0.561	2594.274
3	0	2.509	-0.782	0.580	2718.774
4	0	2.506	-0.684	0.636	2651.086
5	0	2.512	-0.341	0.395	1912.750
6	0	2.503	-0.495	0.409	1776.063
7	0	2.518	-0.670	0.362	2263.750
8	44	2.518	-0.114	0.160	2081.500
9	34	2.535	-0.139	0.302	2993.594
10	0	2.525	-0.698	0.320	2207.641
11	34	2.503	-0.52	0.428	1776.063
12	0	2.541	-0.198	0.392	2565.391
13	0	2.541	-0.275	0.560	2595.766
14	48	2.490	-0.169	0.164	546.875
15	0	2.504	-0.488	0.530	2426.555
16	0	2.507	-0.842	0.711	2511.055
17	88	2.468	-0.340	-0.143	846.469
18	0	2.503	-0.705	0.656	2413.555
19	0	2.506	-0.832	0.518	3218.836
20	38	2.531	-0.464	0.036	3730.922
21	33	2.484	-0.677	0.391	1678.594
22	136	2.528	-0.553	0.623	2746.625
23	140	2.516	-0.246	0.562	2176.906
24	140	2.538	-0.211	0.434	2629.813
25	140	2.538	-0.785	0.507	2699.375
26	92	2.464	-0.107	0.227	849.000
27	132	2.473	0.151	0.284	928.500
28	164	2.521	-0.072	0.173	1019.250
29	164	2.522	0.003	0.322	1043.438
30	228	2.535	0.071	0.727	2134.688
31	164	2.550	0.114	0.843	3400.313
32	164	2.522	0.034	0.267	1150.313
33	164	2.524	-0.208	0.334	1174.500
34	228	2.545	-0.091	1.130	2402.438
35	222	2.525	0.027	0.202	1623.375
36	0	2.524	-0.921	0.829	4184.977
37	0	2.521	-0.755	0.520	4022.977
38	0	2.503	-0.577	0.665	2388.555
39	0	2.507	-0.838	0.484	2513.055
40	40	2.503	-0.740	0.473	2512.805
41	0	2.507	-1.189	0.651	2542.055
42	0	2.503	-0.773	0.655	2388.555
43	0	2.503	-0.468	0.366	2400.555
44	0	2.503	-0.789	0.538	2412.555
45	0	2.510	-0.276	0.487	1851.625
46	96	2.463	0.017	0.463	845.625
47	0	2.517	-1.213	0.667	4105.008
48	0	2.516	-0.808	0.624	3725.250
49	0	2.506	-0.694	0.353	1894.313
50	0	2.513	-0.561	0.314	1966.875
51	0	2.515	-0.693	0.406	2031.000
52	40	2.516	-0.670	0.525	2155.250
53	156	2.518	-0.623	0.390	1702.313
54	196	2.516	-0.03	0.365	1122.000
55	156	2.518	-0.311	0.351	1702.313
56	0	2.512	-0.897	0.720	1883.000

**Table 4.** Experimental and predicted stability constants of spheres complexes for Li<sup>+</sup> in CDCl<sub>3</sub> saturated with D<sub>2</sub>O at 25 °C for train and test sets.

No.	Exp. logK <sub>Li<sup>+</sup></sub> <sup>a</sup>	Pred. logK <sub>Li<sup>+</sup></sub> <sup>b</sup>	RE (%) <sup>c</sup>	No.	Exp. logK <sub>Li<sup>+</sup></sub> <sup>a</sup>	Pred. logK <sub>Li<sup>+</sup></sub> <sup>b</sup>	RE (%) <sup>c</sup>
1	4.61	4.65	-0.87	29	9.59	8.89	7.30
2 <sup>d</sup>	5.23	4.89	6.50	30 <sup>d</sup>	9.40	9.65	-2.66
3	5.2	4.59	11.73	31	7.91	7.87	0.51
4	4.26	4.22	0.94	32	9.23	8.87	3.90
5	5.04	4.61	8.53	33	9.40	9.28	1.28
6	4.78	4.55	4.81	34	9.38	9.62	-2.56
7	5.08	5.36	-5.51	35 <sup>d</sup>	8.95	10.25	-14.53
8	5.11	5.79	-13.31	36 <sup>d</sup>	4.11	4.35	-5.84
9	5.2	5.65	-8.65	37 <sup>d</sup>	4.40	4.48	-1.82
10	5.28	5.81	-10.04	38	4.18	3.98	4.78
11	5.2	5.38	-3.46	39	4.74	4.86	-2.53
12	5.64	5.29	6.21	40	5.34	5.47	-2.43
13	4.91	5.15	-4.89	41	4.76	5.23	-9.87
14	4.91	5.55	-13.03	42	4.45	4.36	2.02
15	4.23	4.05	4.26	43	4.53	4.24	6.40
16	4.08	4.51	-10.54	44 <sup>d</sup>	4.04	4.56	-12.87
17	6.23	6.18	0.80	45	4.93	4.29	12.98
18	4.38	4.22	3.65	46	4.40	4.53	-2.95
19	4.85	4.39	9.48	47 <sup>d</sup>	4.40	4.88	-10.91
20	6.28	6.21	1.11	48 <sup>d</sup>	4.54	4.35	4.19
21	4.83	4.92	-1.86	49	4.96	5.08	-2.42
22	8.11	8.16	-0.62	50	5.33	5.16	3.19
23 <sup>d</sup>	8.79	7.55	14.11	51	5.51	5.32	3.45
24	7.84	8.42	-7.40	52	6.23	6.02	3.37
25	9.23	9.33	-1.08	53	9.96	9.24	7.23
26 <sup>d</sup>	5.58	5.08	8.96	54	9.08	9.34	-2.86
27	5.51	5.82	-5.63	55	8.86	8.72	1.58
28 <sup>d</sup>	9.84	9.24	6.10	56	4.74	5.14	-8.44

<sup>a</sup> Experimental logK<sub>Li<sup>+</sup></sub>.<sup>b</sup> Predicted logK<sub>Li<sup>+</sup></sub>.<sup>c</sup> Percent of relative error of prediction.<sup>d</sup> Test set.**Table 5.** Statistical results of comparing regression models.

Model	RMSEC	RMSEP	RMSECV	Q <sup>2</sup>	R <sub>cal</sub> <sup>2</sup>	R <sub>test</sub> <sup>2</sup>
MLR	0.35	0.65	0.40	0.953 <sup>a</sup> 0.931 <sup>b</sup> 0.930 <sup>c</sup>	0.960	0.920

<sup>a</sup>LOO, <sup>b</sup>Leave-5-out, <sup>c</sup>Leave-10-out

are important for stability constant of macrocycle-metal complexes. As we know the size of lithium is smaller than the cavity size of macrocycles thereby with increasing the size of cavity the interaction between donors and lithium ion decreases, as well as the stability constant.

The molecular path count is the total number of length paths  $m$  in the graph and denoted by  ${}^mP$  ( $m = 0, 1, \dots, L$ ), where  $L$  is the length of the longest path in the graph. The molecular path count may be obtained by the following equation:

$${}^mP = \frac{1}{2} \sum_{i=1}^A {}^mP_i \quad m \neq 0$$

piPC07 increasing leads to wide cavity of macrocycles, lowering the attraction with cation, weak complexation and the decrease the stability constants (logK).

## Conclusions

We have demonstrated that the theoretical molecular descriptors can be successfully applied in the development of predictive QSPR models for the stability constants of 56 lithium macrocycle compounds at standard temperature (25°C) and just in CDCl<sub>3</sub> saturated with D<sub>2</sub>O. A five descriptors equation was developed with these statistical



parameters:  $R^2_{cal} = 0.960$ ,  $Q^2 = 0.953$ ,  $R^2_{test} = 0.920$ , RMSEC = 0.35, RMSECV=0.40 RMSEP=0.65. The strategy applied in this study is in some aspects different from previous works on QSPR modeling of macrocycles.<sup>[32]</sup> A set of 56 hemispherands with very diverse chemical structures was used, and the initial number of molecular descriptors was only five. From the selected parameters, it can be seen that the complexation characteristic of the macrocycle compounds are mainly determined by topological and 3D-morse descriptors. At last, this study can improve the understanding for complexation from the molecular level and then would be of considerable use in prediction of stability constant of hemispherand-metal complexes.

## References

1. Plenio H., Diodane R. *J. Am. Chem. Soc.* **1996**, *118*, 356-367.
2. Zhang X.X., Bordunov A.V., Bradshaw J.S., Dalley N.K., Kou X., Izatt R.M. *J. Am. Chem. Soc.* **1995**, *117*, 11507-11511.
3. Pedersen C.J. *Science* **1988**, *241*, 536-540.
4. Cram D.J. *Science* **1983**, *219*, 1177-1183.
5. Lehn J.M. *Acc. Chem. Res.* **1978**, *11*, 49-57.
6. Dougherty D.A., Stauffer D.A. *Science (Washington, D.C.)* **1990**, *250*, 1558-1560.
7. Pedersen C.J. In: *Synthetic Multidentate Macrocylic Compounds* (Izatt R.M. and Christensen J.J., Eds.) New York: Academic Press, **1978**.
8. Lehn J.M. *Struct. Bond.* **1973**, *76*, 1-69.
9. Pressman B.C. *Annu. Rev. Biochem.* **1976**, *45*, 501-530.
10. Simon W., Morf W.E., Meier P.Ch. *Struct. Bond.* **1973**, *76*, 113-160.
11. Lindenbaum S., Rytting J.H., Stenson L.A. In: *Progress in Macrocylic Chemistry* (Izatt R.M. and Christensen J.J., Eds.) New York: Wiley & Sons, **1979**.
12. Dalley N.K. In: *Synthetic Multidentate Macrocylic Compounds* (Izatt R.M. and Christensen J.J., Eds.)
13. Goldberg I. *Acta Crystallogr., Sect. B.* **1975**, *31*, 754-762.
14. Cram D.J. *Angew. Chem., Int. Ed. Engl.* **1988**, *27*, 1009-1020.
15. Cram D.J., Kaneda T., Helgeson R.C., Brown S.B., Knobler C.B., Maverick E., Rueblood K.N.T. *J. Am. Chem. Soc.* **1985**, *107*, 3645-3657.
16. Tucker J.A., Knobler C.B., Goldberg I., Cram D.J. *J. Org. Chem.* **1989**, *54*, 5460-5482.
17. Lein G.M., Cram D.J. *J. Am. Chem. Soc.* **1985**, *107*, 448-455.
18. Doxsee K.M., Feigel M., Stewart K.D., Canary J.W., Knobler C.B., Cram D.J. *J. Am. Chem. Soc.* **1987**, *109*, 3098-3107.
19. Cram D.J., Ho S.P. *J. Am. Chem. Soc.* **1986**, *108*, 2998-3005.
20. Pedersen C.J. *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **1968**, *27*, 1305-1309.
21. Izatt R.M., Pawlak K., Bradshaw J.S. *Chem. Rev.* **1995**, *95*, 2529-2586.
22. Lamb J.D., Izatt R.M., Christensen J.J. In: *Progress in Macrocylic Chemistry* (Izatt R.M. and Christensen J.J., Eds.) New York: Wiley, **1981**.
23. Pearson R.G. *J. Am. Chem. Soc.* **1963**, *85*, 3533-3539.
24. Lowry T.H., Richardson K.S. *Mechanism and Theory in Organic Chemistry*. 3d ed., New York: Harper & Row, **1987**.
25. Solovev V.P., Varnek A.A. *Russ. Chem. Bull.* **2004**, *53*, 1434-1445.
26. Svetlitski R., Lomaka A., Karelso M.N. *Sep. Sci. Technol.* **2006**, *41*, 197-216.
27. Ghasemi J., Ahmadi Sh. *Annali di Chimica* **2007**, *97*, 69-83.
28. Ghasemi J., Saaipour S., Brown S.D. *J. Mol. Struct.* **2007**, *805*, 27-32.
29. Ghasemi J., Saaipour S. *Chem. Pharm. Bull.* **2007**, *55*, 669-674.
30. Ghasemi J., Shahmirani S., Farahani E.V. *Annali di Chimica* **2006**, *96*, 327-337.
31. Ghasemi J., Asadpour S., Abdolmaleki A. *Anal. Chim. Acta.* **2007**, *588*, 200-206.
32. Ghasemi J., Saaipour S. *J. Inclusion Phenom. Macrocylic Chem.* **2008**, *60*, 339-351.
33. Izatt R.M., Pawlak K., Bradshaw J.S. *Chem. Rev.* **1991**, *91*, 1721-2085.
34. Dewar M.J.S., Zoebisch E.G., Healy E.F., Stewart J.J.P. *J. Am. Chem. Soc.* **1985**, *107*, 3902-3909.
35. DRAGON, Web 3.0 available from <<http://www.disat.unimib.it/chm>>.
36. Draper N.R., Smith H. *Applied Regression Analysis*. New York: John Wiley & Sons, **1981**.
37. Golbraikh A., Tropsha A. *J. Mol. Graphics Modell.* **2002**, *20*, 269-276.
38. Elroby S.A.K., Lee K.H., Cho S.J., Hinchliffe A. *Can. J. Chem.* **2006**, *84*, 1045-1049.
39. Todeschini R., Consonni V. *Handbook of Molecular Descriptors*. New York: John Wiley & Sons, **2000**.
40. Sanders J.K.M., Williams D.H. *J. Am. Chem. Soc.* **1971**, *93*, 641-645.
41. Lovasz L., Pelkin J. *Periodica Mathematica Hungarica* **1973**, *3*, 175-182.
42. Schuur J.H., Selzer P., Gasteiger J. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 334-344.
43. Gasteiger J., Sasowski J., Selzer P., Steinhauer L., Steinhauer V. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1030-1037.
44. Sadowski J., Gasteiger J. *Chem. Rev.* **1993**, *93*, 2567-2581.

Received 06.11.2010

Accepted 29.11.2010