J Incl Phenom Macrocycl Chem (2011) 70:37–47
DOI 10.1007/s10847-010-9854-9

ORIGINAL ARTICLE

A quantitative structure–property relationships study of the stability constant of crown ethers by molecular modelling: new descriptors for lariat effect

Jahan B. Ghasemi · M. K. Rofouei · M. Salahinejad

Received: 19 May 2010 / Accepted: 27 August 2010 / Published online: 14 September 2010
© Springer Science+Business Media B.V. 2010

Abstract Quantitative structure–property relationships (QSPR) modeling of stability constant (log K) for complexes of 15-Crown-5 ethers with sodium cation (Na^+) at 25 °C in methanol solution as a function of the new and Dragon generated descriptors was established by multiple linear regression and partial least squares regression. The possibility of three-dimensional cation encapsulation by the crown ether unit and the side arms and cooperative interaction between ring and side-arm in lariat crown ethers are investigated. These side arms can alter the cation-binding constant and the relative cation selectivity significantly when comparing with their parent crown ethers. In order to consider lariat and side arm effect and conformation changing in the crown ether rings on complexation of reactions, we defined and calculated some novel descriptors to address these effects. Three new descriptors are involved in the QSPR model and had significant effect on stability constants. The quality of the model was evaluated using leave-one-out cross-validation and randomization test method. We showed that a combination of new defined and Dragon based descriptors could lead to relevant QSPR models, that in comparison with model solely generated by Dragon descriptors has improved and showed more robustness and prediction ability.

Keywords 15-Crown-5 · Crown ethers · Stability constant · QSPR · Lariat effect · New descriptor

J. B. Ghasemi (✉)
Chemistry department , Faculty of Science,
K. N. Toosi University of Technology, Tehran, Iran
e-mail: jahan.ghasemi@gmail.com

M. K. Rofouei · M. Salahinejad
Chemistry department , Faculty of Science,
Tarbiat Moallem University, Tehran, Iran

Introduction

Since Pedersen reported the synthesis and cation-binding characteristics of the crown ethers (macrocyclic polyethers) [1, 2], many studies about their complexing ability, phase transfer catalysis, metal cation transport and metal cation analysis have been reported [3–5]. They are also widely applied in chemical technology and analytical chemistry as ion pair extractants, membrane transfer, ion carriers, masking agents and sensors [6–10]. Crown ethers continue to be one of the most useful parts of supramolecular (host/guest) chemistry [11].

Many different modifications of the crown ethers, such as changing the ring size, the kinds of substituents and the types of donor atoms have been made to enhance their complexation properties. Lariat crown ethers (LCE) are modified crown ethers having one or more donor-group-bearing side arms [12]. They were designed to add dimensionality to the essentially flat crown ether and have acceptable complexation–decomplexation dynamics coupled with the possibility of three-dimensional cation encapsulation by the crown ether unit and the side arm [13]. It was shown that the introduction of both nondonating alkyl and donating oxyethylene groups to the carbon-pivot crown ethers (side chain attached through a carbon atom) alter not only the cation-binding constant but also the relative cation selectivity significantly when comparing with their parent crown ether [14–16]. In general, side arms especially those containing donor groups; enhance the binding strength of lariat crown ethers toward cations in comparison with their crown ether analogues without side arms by cooperative ring and side-arm interaction [17]. Therefore the number of reported lariat crown ethers bearing one, two, or more side arms has increased dramatically in the past few years [18, 19]. The

complexation phenomena exhibited by lariat crown ethers are influenced by the hole-size relationship, ligand flexibility and conformation, shape and topology, total number of donor atoms and solvation energies for the cation, macrocycle, and complex [20].

The computer models are useful for the inspection of trends in complexation phenomena. They offer a fast measure of predictability in the absence of extensive experimental or computed data on crown ether properties. In our previous paper, we reported on the application of quantitative structure–property/activity relationships (QSPR/QSAR) technique to build a QSPR best multiple regression model, which correlated and predicted the stability constants (log K) of complexation between 15-crown-5 ethers and potassium ion (K^+) in methanol solution [21]. The aim of present study is to define and calculate some novel descriptors to estimate lariat effect, the effects of different pendant-arm substituents and also changing in the conformation of the crown ether ring on complexation process of sodium ion with 15-Crown-5 derivatives. These new descriptors combined with Dragon descriptors were used to build a new QSPR model, which could correlate and predict the stability constant of complexation between 15-Crown-5 derivatives and sodium ion (Na^+) in methanol solution. We will show that a combination of novel descriptors and Dragon descriptors could lead to relevant QSPR models, which in comparison with Dragon based descriptors, may allow us to improve the robustness of prediction.

Materials and methods

Data set

A set of 88 15-crown-5 ether derivatives with one or more side arms were selected. Their chemical structures are listed in Table 1. The experimental values of stability constants (log K) of 15-crown-5 ether derivatives with sodium cation (Na^+) were taken from the literature [22] and listed in Table 2.

The data set was divided into training and prediction set. The training set of 70 molecules was used to adjust the parameters of the models, and the rest of molecules (18 about 20% of all molecules) were used to evaluate model prediction ability. The test set compounds were selected by considering both the distribution of stability constants and structural diversity (Table 1). It can be seen that except the highest and lowest-activity compounds were included in the training set, the stability constants of test set is similar to that of the training set. Since the temperature and solvent also affect the stability constants, we used only data obtained at standard temperature (25 °C) in methanol solution.

Molecular optimization and descriptor calculation

The structure of molecules was drawn in ChemBioOffice 11.0 [23] and the mol files format exported in HyperChem 8.0.5 [24]. The geometry optimization was performed on a Dell Vostro 1520 Laptop computer with Intel (R) Core ™ 2 Duo CPU with windows XP as operating system. The optimization program includes two stages of molecular orbital (MO) calculation: first using MM^+ force field and then using semi-empirical method AM1. In addition we applied a more rigorous optimization method like density functional theory (DFT) with and without solvation effect for optimization of the molecules. As the final results did not show considerable differences we used AM1 to keep parsimony of the model making in regards to time and cost of calculation method. The gradient norm criterion 0.01 kcal/Å was applied in the geometry optimization for all structures. The HyperChem hin output files were used by the Dragon program [25] to compute molecular descriptors. We define some new descriptors to consider lariat effect, the effects of different pendant-arm substituents and also changing in cavity shape of the crown ether ring on complexation of cations with crown ethers. Because the generation of the descriptors is based on the gas-phase geometry calculation of AM1, the solvation of the ligands molecules didn't taking into account. The SPSS software (version 16.0, SPSS, Inc.) was employed for the MLR analysis and other calculations were performed in the MATLAB (version 7.6.0.324, Math Works, Inc.) environment.

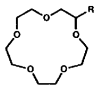
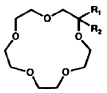
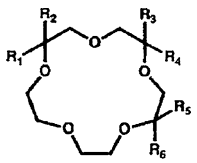
Variable selection

The generated descriptors were first analyzed for the existence of constant or near constant variables and those detected were removed. In addition, to reduce redundancy in the descriptor data matrix, correlation of the descriptors with each other and with the dependent variable (log K) of the molecules was examined and the collinear descriptors ($r > 0.9$) were detected. Among the collinear descriptors, that with the highest correlation with log K was retained and the others were removed from the data matrix. Subsequently, the method of stepwise multiple linear regression (MLR) was used to select the most important descriptors and to calculate the coefficients relating the descriptors to stability constant.

Results and discussion

The molecular descriptor is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some

Table 1 Chemical Structure of 15C5 derivatives

No		No	
1	C_2H_5	48	R_1: CH_3 R_2: CH_3
2	C_6H_{13}	49	R_1: CH_3 R_2: CH_2Br
3	C_8H_{17}	50	R_1: CH_3 R_2: $CH_2OCH_2CH_2OH$
4	$C_{10}H_{21}$	51	R_1: CH_3 R_2: $CH_2(OCH_2CH_2)_2OH$
5	$C_{12}H_{25}$	52	R_1: CH_3 R_2: $CH_2(OCH_2CH_2)_3OH$
6	CH_2OH	53	R_1: CH_3 R_2: $CH_2OCH_2CH_2OCH_3$
7	CH_2OCH_3	54	R_1: CH_3 R_2: $CH_2O(CH_2CH_2O)_2CH_3$
8	$CH_2OCH_2CH{=}CH_2$	55	R_1: CH_3 R_2: $CH_2O(CH_2CH_2O)_3CH_3$
9	$CH_2COC(CH_3)_3$	56	R_1: CH_3 R_2: $CH_2O(CH_2)_3OCH_3$
10	$CH_2OC_3H_7$	57	R_1: CH_3 R_2: $CH_2OC_6H_{13}$
11	$CH_2CO_2CH_3$	58	R_1: CH_3 R_2: $CH_2SC_6H_{13}$
12	$CH_2OC_8H_{17}$	59	R_1: CH_3 R_2: $CH_2NHC_6H_{13}$
13	C_6H_5	60	R_1: CH_3 R_2: $CH_2OC_8H_{17}$
14	$CH_2OC_6H_4$	61	R_1: CH_3 R_2: $CH_2OCH_2CH_2OC_8H_{17}$
15	$CH_2OCH_2C_6H_4$	62	R_1: CH_3 R_2: $CH_2(OCH_2CH_2)_2OC_8H_{17}$
16	CH_2O[2-$CH_3OC_6H_4$]	63	R_1: CH_3 R_2: $CH_2OC_{12}H_{25}$
17	CH_2O[3-$CH_3OC_6H_4$]	64	R_1: CH_3 R_2: $CH_2OCH_2CH_2OC_{12}H_{25}$
18	CH_2O[4-$CH_3OC_6H_4$]	65	R_1: CH_3 R_2: $CH_2(OCH_2CH_2)_2C_{12}H_{25}$
19	CH_2OCH_2–[2-$CH_3OC_6H_4$]	66	R_1: CH_3 R_2: CH_2OCH_2–[2-C_5H_4N]
20	CH_2O–[2-$NO_2C_6H_4$]	67	R_1: CH_3 R_2: CH_2OCH_2–[2-THF]
21	CH2O–[4-NO2C6H4]	68	R_1: CH_3 R_2: CH_2O–[2-$CH_3OC_6H_4$]
22	CH_2O–[2-CH_2=$CHCH_2OC_6H_4$]	69	R_1: CH_3 R_2: CH_2O–[8-Quinoline]
23	CH_2O–[2-CH_2=$C(CH_3)CH_2OC_6H_4$]	70	R_1: CH_3 R_2: CH_2O–[2-CH_3-8-Quinoline]
24	CH_2O–[4-CH_3O–$4C_3H_7C_6H_3$]	71	R_1: C_6H_{13} R_2: CH_2Br
25	CH_2O–[2-CH_3O-4-$CH_2CH{=}CH_2C_6H_3$]	72	R_1: C_6H_{13} R_2: $CH_2OCH_2CH_2OCH_3$
26	CH_2O–[2-(CH_3O-4-$CH_3CH(OH)CH2C_6H_3$]	73	R_1: C_6H_{13} R_2: $CH_2(OCH_2CH_2)_2OCH_3$
27	CH_2O–[2-(Oxirane)–$CH_2OC_6H_4$]	74	R_1: C_6H_{13} R_2: $CH_2(OCH_2CH_2)_3OCH_3$
28	CH_2O–[2-(2-methyloxirane)–$CH_2OC_6H_4$]	75	R_1: C_6H_{13} R_2: $CH_2OC_6H_{13}$
29	$CH_2OCH_2CH(OH)CH_3$	76	R_1: C_6H_{13} R_2: $CH_2OC_8H_{17}$
30	$CH_2OCH_2CH(OH)CH_2OH$	77	R_1: C_6H_{13} R_2: $CH_2OCH_2CH_2OC_8H_{17}$
31	CH_2OCH_2–(2,2-dimethyl-1,3dioxolane)	78	R_1: C_6H_{13} R_2: $CH_2(OCH_2CH_2)_2OC_8H_{17}$
32	$CH_2OCH_2CH(OC_6H_{13})$–$CH_2OC_6H_{13}$	79	R_1: C_6H_{13} R_2: CH_2O–[8-Quinoline]
33	CH_2OCH_2–(Oxirane)	80	R_1: C_8H_{17} R_2: CH_2Br
34	$CH_2OCH_2CH_2OCH_3$	81	R_1: C_8H_{17} R_2: $CH_2OCH_2CH_2OCH_3$
35	$CH_2OCH_2CH_2OC_4H_9$	82	R_1: C_8H_{17} R_2: $CH_2(OCH_2CH_2)_2OCH_3$
36	$CH_2OCH_2CH_2OC_8H_{17}$	83	R_1: C_8H_{17} R_2: $CH_2(OCH_2CH_2)_3OCH_3$
37	$CH_2O(CH_2CH_2O)_2CH_3$		
38	$CH_2O(CH_2CH_2O)_2C_8H_{17}$		
39	$CH_2O(CH_2CH_2O)_3H$		
40	$CH_2O(CH_2CH_2O)_3CH_3$		
41	CH_2NH_2		
42	$CH_2NHC(CH_3)_3$		
43	$CH_2NHC_6H_{13}$	84	R_1,R_3,R_4,R_6: H R_2, R_5: $CH_2OCH_2CH_2OCH_3$
44	$CH_2NHC_6H_5$	85	R_1, R_6: CH_3; R_3,R_4: H; R_2, R_5: $CH_2OCH_2CH_2OCH_3$ (Trans)
45	$CH_2NHCH_2CH_2NH_2$	86	R_1, R_6: CH_3; R_3,R_4: H; R_2, R_5: $CH_2OCH_2CH_2OCH_3$ (Mixture of isomers)
46	CH_2O–[1-Naphtalene]	87	R_1,R_3: CH_3; R_2,R_4: $CH_2OCH_2CH_2OCH_3$; R_5, R_6: H (Trans)
47	CH_2O–[8-Quinoline]	88	R_1,R_3: CH_3; R_2,R_4: $CH_2OCH_2CH_2OCH_3$; R_5, R_6: H (Mixture of Isomers)

standardized experiment [26, 27]. Once appropriate molecular descriptors are chosen, QSAR models can be constructed to predict the target property of untested chemicals.

Dragon software computes more than 1490 structural descriptors including: constitutional, topological, geometrical, radial distribution function (RDF), charge, WHIM (Weighted Holistic Invariant Molecular), GETAWAY (GEometry, Topology and Atoms-Weighted AssemblY) and 3D-MoRSE (3D-Molecular Representation of Structure based on Electron diffraction), molecular walk counts, Galvez topological charge indices, Randic molecular profiles, functional groups and atom-centred fragments [26]. At the first stage Dragon descriptors were used to build QSPR model on stability constant of complexation between 15-Crown-5 derivatives and sodium ion. Then in order to consider lariat and side arm effects and also the changing in the conformation of the crown ether ring, we defined and calculated more than fifteen descriptors to estimate of these effects. These new descriptors can be classified as follows: (i) Descriptors to estimate lariat and side arm effects i.e. number of lariat unit (O–CH_2CH_2–O), number of oxygen, donor atoms and withdrawing group in the side arms, length of side arms (number of bond), number of side arms on the crown ether ring (ii) Conformation changing (shape and size) of crown ring by side arms i.e. distances between the nearest oxygen to the side arm (O_{13}, Fig. 1) with other oxygen's of the crown ring, distances between adjacent and facing oxygens, the longest and shortest distance of two oxygen of the ring and the ratio of these distances after AM1 optimization and (iii) Hybrid descriptors, the combination of some of these descriptors as a coded descriptor, includes: number of arms–length of the longest arm, number of lariat unit–length of lariat, number of donor group–distance of donor group and number of withdrawing group–distance of withdrawing group to the crown ring. We used Dragon based descriptors and combination of new descriptors with Dragon descriptors to build QSPR models using MLR and PLS methods to correlate and predict the stability constants of formed complexes between 15-Crown-5 derivatives and sodium ion (Na^+) in methanol solution.

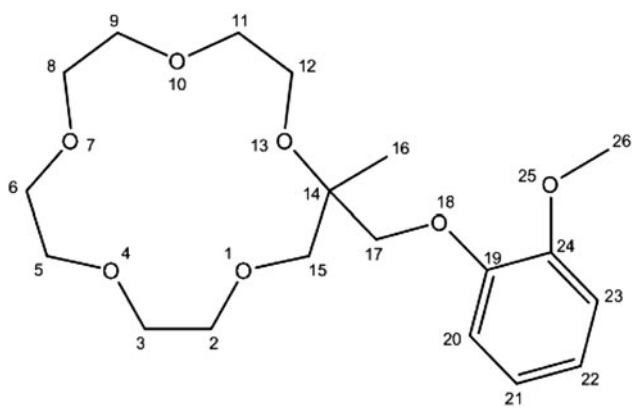


Fig. 1 Chemical structure of a typical molecule (molecule no. 68)

Construction of models

Both MLR and partial least square (PLS) techniques were used for modelling of stability constants of 15-crown-5 derivatives. After regression analysis, a few suitable models were obtained among which the best model was selected and a five molecular descriptors proposed were used to establish a MLR models. Experimental and predicted stability constants and relative error of MLR and PLS models are shown in Table 2. Statistical results of comparing regression models of MLR and PLS methods for model based on the combination of new defined descriptors along with Dragon descriptors (Model I) and descriptors solely obtained by Dragon (Model II) are summarized in Table 3. The values of five descriptors involved in MLR model I represented in Table 4. The plots of predicted log K versus experimental log K and the residuals (experimental log K − predicted log K) versus experimental log K value and the random distribution of residuals about zero mean, obtained by the MLR and PLS modeling are shown in Figs. 2 and 3, respectively. The regression results of the MLR model I are summarized in Table 5. In Table 5, b and S_b are the non-standardized coefficient of descriptors and standard error of coefficient, respectively, and b_s is the standardized regression coefficient. As there is more than one variable presented in the correlations, it is necessary to examine the stability of the regression. The variance inflation factors (VIF) are indices, which assess the quality of the estimations of the model coefficients. VIF values greater than five indicates that information of descriptors can be hidden by correlation of descriptors [28]. As one can see from Table 5, the VIF values are all less than 2.3, indicating the stability of the equations constructed.

Validation of the models

The obtained models were cross-validated using the leave-one-out (LOO) technique and internally validated by a randomization test. In the LOO cross-validation method a data point is removed from the set and the model is recalculated. The predicted property for that point is then compared with its actual value. This is repeated until each data point is omitted once. The LOO cross-validation correlation coefficient (q^2) and root mean square error of the cross validation (RMSECV) of models are presented in Table 4.

Table 2 Experimental and predicted stability constants of 15-crown-5 ether complexes by PLS and MLR and resulted relative errors of Models I and II

No	MLR PLS												
	Model I				Model II			Model I			Model II		
	Exp	Pred	residual	R.E(%)	Pred	residual	R.E(%)	Pred	residual	R.E(%)	Pred	residual	R.E(%)
1	2.7	2.72	0.02	0.9	3.06	0.36	13.4	2.76	0.06	2.1	2.95	0.25	9.4
2	3.2	3.03	−0.17	−5.3	2.90	−0.30	−9.2	3.02	−0.18	−5.8	2.78	−0.42	−13.0
3	3.2	3.14	−0.06	−2.0	2.94	−0.26	−8.3	2.90	−0.30	−9.4	2.89	−0.31	−9.6
4	3.2	2.96	−0.24	−7.5	3.57	0.37	11.4	2.76	−0.44	−13.6	3.49	0.29	8.9
5	3.13	3.18	0.05	1.6	3.19	0.06	1.8	3.01	−0.12	−3.9	3.25	0.12	3.9
6	3.1	2.76	−0.34	−11.1	2.70	−0.40	−12.8	2.92	−0.18	−5.8	2.87	−0.23	−7.4
7	3.03	2.67	−0.36	−11.9	2.86	−0.17	−5.5	2.82	−0.21	−6.9	3.93	0.90	29.7
8	3.12	2.91	−0.21	−6.7	3.24	0.12	3.8	2.91	−0.21	−6.9	3.27	0.15	5.0
9	2.95	2.66	−0.29	−9.9	3.22	0.27	9.0	2.65	−0.30	−10.3	2.99	0.04	1.2
10	3.05	3.18	0.13	4.4	3.26	0.21	6.9	3.25	0.20	6.5	3.29	0.24	7.7
11[a]	2.72	2.77	0.05	1.7	3.14	0.42	15.4	3.12	0.40	14.6	2.80	0.08	2.9
12	3.18	3.24	0.06	2.0	3.02	−0.16	−5.2	3.25	0.07	2.1	2.94	−0.24	−7.5
13	3.3	3.34	0.04	1.3	2.94	−0.36	−10.8	3.28	−0.02	−0.7	3.09	−0.21	−6.4
14	2.51	2.97	0.46	18.3	2.77	0.26	10.2	2.89	0.38	15.0	2.78	0.27	10.7
15[a]	2.97	2.93	−0.04	−1.2	2.96	−0.01	−0.3	2.95	−0.02	−0.6	3.10	0.13	4.4
16	3.24	2.98	−0.26	−8.1	2.86	−0.38	−11.6	3.00	−0.24	−7.4	2.89	−0.35	−10.7
17	2.89	3.03	0.14	4.8	2.79	−0.10	−3.5	2.94	0.05	1.8	2.73	−0.16	−5.4
18	3	3.09	0.09	3.2	2.88	−0.12	−4.1	3.13	0.13	4.5	2.79	−0.21	−7.1
19[a]	3.04	3.33	0.29	9.5	3.10	0.06	2.0	3.26	0.22	7.2	3.12	0.08	2.7
20	2.83	2.82	−0.01	−0.3	2.97	0.14	5.1	3.28	0.45	15.8	2.94	0.11	3.9
21	2.72	2.82	0.10	3.7	2.98	0.26	9.5	3.06	0.34	12.4	2.74	0.02	0.8
22	3.07	3.05	−0.02	−0.5	3.16	0.09	2.8	3.00	−0.07	−2.3	3.11	0.04	1.4
23	3.04	3.16	0.12	4.1	3.78	0.74	24.2	3.00	−0.04	−1.5	3.63	0.59	19.5
24	2.94	2.96	0.02	0.6	3.06	0.12	4.0	3.10	0.16	5.5	3.05	0.11	3.8
25	2.86	3.01	0.15	5.4	3.23	0.37	13.0	3.07	0.21	7.3	3.18	0.32	11.1
26[a]	2.93	2.24	−0.69	−23.7	3.05	0.12	4.0	2.56	−0.37	−12.6	2.81	−0.12	−4.2
27	3.03	3.15	0.12	3.8	3.04	0.01	0.5	3.10	0.07	2.3	2.96	−0.07	−2.5
28	3.02	3.16	0.14	4.6	3.66	0.64	21.2	3.34	0.32	10.6	3.32	0.30	9.8
29	3.9	3.05	−0.85	−21.9	3.41	−0.49	−12.5	3.40	−0.50	−12.9	3.29	−0.61	−15.7
30[a]	3	2.17	−0.83	−27.7	2.65	−0.35	−11.6	2.74	−0.26	−8.6	2.38	−0.62	−20.6
31[a]	3.03	2.69	−0.34	−11.3	3.36	0.33	10.8	3.05	0.02	0.5	3.30	0.27	9.0
32	2.97	2.86	−0.11	−3.6	3.46	0.49	16.5	3.27	0.30	10.2	3.21	0.24	8.0
33	3.04	3.21	0.17	5.8	3.18	0.14	4.8	3.21	0.17	5.6	3.24	0.20	6.6
34	3.15	3.18	0.03	1.1	3.13	−0.02	−0.6	3.23	0.08	2.6	3.19	0.04	1.3
35	3.09	3.23	0.14	4.5	3.17	0.08	2.5	2.29	−0.80	−25.9	3.09	0.00	0.1
36	3.22	3.26	0.04	1.1	3.19	−0.03	−0.8	3.17	−0.05	−1.4	3.07	−0.15	−4.6
37[a]	3.13	3.10	−0.03	−0.9	3.03	−0.10	−3.1	3.16	0.03	1.0	3.16	0.03	0.8
38	3.23	3.45	0.22	6.9	3.25	0.02	0.6	3.28	0.05	1.5	3.13	−0.10	−3.0
39	3.04	3.20	0.16	5.2	2.96	−0.08	−2.7	3.16	0.12	3.8	3.20	0.16	5.1
40	3.09	3.23	0.14	4.7	2.98	−0.11	−3.4	3.20	0.11	3.6	3.15	0.06	2.0
41[a]	2.82	3.03	0.21	7.4	2.57	−0.25	−8.8	2.93	0.11	4.0	2.61	−0.21	−7.4
42	2.79	2.93	0.14	4.9	3.60	0.81	29.2	2.81	0.02	0.8	3.15	0.36	13.0
43	2.82	3.38	0.56	19.9	3.00	0.18	6.4	3.00	0.18	6.2	2.63	−0.19	−6.7
44	2.91	2.86	−0.05	−1.7	2.90	−0.01	−0.3	2.83	−0.08	−2.9	2.71	−0.20	−7.0
45[a]	2.92	3.20	0.28	9.5	2.77	−0.15	−5.0	3.13	0.21	7.3	2.57	−0.35	−11.9

Table 2 continued

No	MLR PLS												
	Model I				Model II			Model I			Model II		
	Exp	Pred	residual	R.E(%)	Pred	residual	R.E(%)	Pred	residual	R.E(%)	Pred	residual	R.E(%)
46	2.74	3.06	0.32	11.6	3.25	0.51	18.6	2.99	0.25	9.0	3.45	0.71	25.9
47[a]	3.72	3.37	−0.35	−9.4	3.02	−0.70	−18.7	3.39	−0.33	−8.9	3.12	−0.60	−16.1
48	2.99	3.44	0.45	14.9	3.32	0.33	10.9	3.27	0.28	9.4	3.32	0.33	11.1
49	2.86	3.32	0.46	16.1	3.03	0.17	6.0	2.29	−0.57	−19.9	3.00	0.14	5.0
50	3.88	3.57	−0.31	−7.9	3.56	−0.32	−8.3	3.72	−0.16	−4.1	3.78	−0.10	−2.6
51	3.88	3.72	−0.16	−4.1	3.69	−0.19	−4.9	3.76	−0.12	−3.1	3.90	0.02	0.6
52	3.73	3.62	−0.11	−2.8	3.30	−0.43	−11.5	3.49	−0.24	−6.5	3.61	−0.12	−3.3
53	3.87	3.51	−0.36	−9.2	3.65	−0.22	−5.7	3.57	−0.30	−7.7	3.75	−0.12	−3.1
54[a]	3.89	3.53	−0.36	−9.1	3.68	−0.21	−5.4	3.68	−0.21	−5.3	3.79	−0.10	−2.6
55	3.87	3.60	−0.27	−6.9	3.75	−0.12	−3.0	3.79	−0.08	−2.1	3.86	−0.01	−0.2
56	3.48	3.56	0.08	2.4	3.58	0.10	2.9	3.59	0.11	3.2	3.60	0.12	3.5
57	3.57	3.51	−0.06	−1.5	3.57	0.00	0.1	3.55	−0.02	−0.4	3.55	−0.02	−0.6
58	3.08	3.44	0.36	11.6	3.31	0.23	7.3	3.20	0.12	3.8	3.12	0.04	1.2
59	3.08	3.64	0.56	18.0	3.27	0.19	6.1	3.62	0.54	17.4	3.04	−0.04	−1.5
60[a]	3.54	3.73	0.19	5.4	3.42	−0.12	−3.4	3.62	0.08	2.3	3.36	−0.18	−5.1
61	3.75	3.70	−0.05	−1.5	3.48	−0.27	−7.2	3.84	0.09	2.3	3.47	−0.28	−7.5
62	3.88	3.60	−0.28	−7.1	3.78	−0.10	−2.5	3.81	−0.07	−1.9	3.68	−0.20	−5.2
63	3.42	3.56	0.14	4.0	3.61	0.19	5.6	3.54	0.12	3.4	3.55	0.13	3.9
64	3.75	3.86	0.11	3.0	3.44	−0.31	−8.3	3.82	0.07	1.9	3.56	−0.19	−5.2
65	3.89	3.81	−0.08	−2.1	3.84	−0.05	−1.2	3.87	−0.02	−0.4	3.92	0.03	0.7
66[a]	3.58	3.87	0.29	8.1	3.21	−0.37	−10.3	3.75	0.17	4.9	3.32	−0.26	−7.2
67	4.02	3.61	−0.41	−10.1	3.41	−0.61	−15.1	3.74	−0.28	−7.0	3.42	−0.60	−14.8
68	3.79	3.86	0.07	1.8	4.07	0.28	7.4	3.85	0.06	1.6	4.05	0.26	6.9
69	4.87	4.53	−0.34	−7.1	4.12	−0.75	−15.4	4.64	−0.23	−4.7	4.16	−0.71	−14.5
70	4.31	3.76	−0.55	−12.8	3.93	−0.38	−8.9	3.78	−0.53	−12.3	3.89	−0.42	−9.6
71	2.74	3.17	0.43	15.5	3.01	0.27	9.7	2.72	−0.02	−0.9	3.04	0.30	10.8
72[a]	3.9	3.89	−0.01	−0.2	4.08	0.18	4.7	3.80	−0.10	−2.5	4.18	0.28	7.2
73	3.91	3.93	0.02	0.5	4.08	0.17	4.2	4.09	0.18	4.6	4.09	0.18	4.7
74	3.71	3.64	−0.07	−2.0	3.40	−0.31	−8.4	3.94	0.23	6.1	3.59	−0.12	−3.3
75[a]	3.56	3.68	0.12	3.4	3.03	−0.53	−14.8	3.76	0.20	5.5	3.30	−0.26	−7.2
76	3.39	3.52	0.13	3.7	3.19	−0.20	−6.0	3.59	0.20	6.0	3.47	0.08	2.5
77	3.62	3.44	−0.18	−5.0	3.35	−0.27	−7.4	3.55	−0.07	−1.8	3.48	−0.14	−3.8
78	3.75	3.38	−0.37	−10.0	3.20	−0.55	−14.6	3.44	−0.31	−8.3	3.48	−0.27	−7.1
79[a]	4.85	4.19	−0.66	−13.6	4.04	−0.81	−16.7	4.14	−0.71	−14.7	4.04	−0.81	−16.6
80	2.79	2.78	−0.01	−0.2	2.90	0.11	3.9	2.82	0.03	0.9	2.94	0.15	5.2
81[a]	3.82	3.48	−0.34	−8.8	3.81	−0.01	−0.1	3.37	−0.45	−11.8	3.99	0.17	4.5
82	3.86	3.56	−0.30	−7.8	3.82	−0.04	−0.9	3.26	−0.60	−15.5	3.93	0.07	1.9
83	3.75	3.66	−0.09	−2.3	3.84	0.09	2.3	3.52	−0.23	−6.2	3.91	0.16	4.2
84	3.09	3.36	0.27	8.6	3.26	0.17	5.6	3.30	0.21	6.7	3.40	0.31	10.1
85[a]	3.89	4.20	0.31	7.9	3.97	0.08	2.1	4.06	0.17	4.3	4.03	0.14	3.7
86	4.11	4.34	0.23	5.5	4.02	−0.09	−2.1	4.21	0.10	2.5	4.13	0.02	0.4
87	4.15	4.19	0.04	1.1	4.02	−0.13	−3.2	4.17	0.02	0.6	4.07	−0.08	−1.8
88	4.36	4.41	0.05	1.1	4.13	−0.23	−5.3	4.35	−0.01	−0.2	4.19	−0.17	−3.8

[a] Prediction set

Table 3 Statistical quality parameters of MLR and PLS for Model I and II

Model	RMSEC	RMSEP	RMSECV	R^2_{cal}	Q^2	R^2_{Pred}
MLR						
Model I	0.2593	0.3745	0.2847	0.71	0.652	0.631
Model II	0.3054	0.3467	0.3357	0.598	0.517	0.642
PLS[a]						
Model I	0.1748	0.2010	0.2441	0.860	0.720	0.885
Model II	0.2101	0.2834	0.2821	0.809	0.661	0.752

[a] PLS models with five latent variables

The randomization technique consists of giving random values to the dependent variable and constructing a model with the real input descriptors. This randomization is repeated several times, and the resulting data are trained against real independent variables [29]. The **y**-randomizations performed imply that acceptable QSAR models were obtained for the given data sets by the current modeling method and they did not show any chance correlation. In this work, the LOO q^2 and r^2 values resulting from **y**-randomization for the MLR model were 0.04 ± 0.05 and 0.07 ± 0.04 and for PLS model were 0.03 ± 0.04 and 0.24 ± 0.08, respectively. The results of these internal validation methods for model I are presented in Fig. 4. To show that the selected variables in the model are not a random process (especially the new defined descriptors) we defined a random or dummy variable and insert it in the pool of initial descriptors. We repeat several times the variable selection process by stepwise regression and the selection of the dummy variable was not observed.

Effect of descriptors on the model

Concerning the interpretability of the descriptors, or effect of each individual descriptors on the response variable, it is important to take into account that modeled response is the result of a series of complex electrostatic and solvating phenomenon and mechanisms, thus it is very difficult to attribute too much importance to the mechanistic meaning of the molecular descriptors used in a QSPR model. Moreover, it must also be highlighted that in multivariate models such as MLR models, even though the interpretation of the singular molecular descriptor can be certainly useful, it is only the combination of the selected set of descriptors that is able to model the target property. On the other hand, although the explanation of the involved descriptors in the model is not a simple task, by definition of chemical and physical interpretations of the selected molecular descriptors in final model, we could get a better understanding of the relationship between the structure of

Table 4 The descriptors values involved in model I for 88 compounds

NO	NALA	Mor05 m	SPI	D_4	D_1
1	102	−1.333	9.381	5.936	3.232
2	106	−1.807	12.728	4.635	3.271
3	108	−3.178	14.318	4.041	2.965
4	110	−3.042	15.875	5.888	3.13
5	112	−3.426	17.407	5.245	3.210
6	102	−1.493	9.381	5.909	3.233
7	103	−1.687	10.247	5.960	3.074
8	105	−2.321	11.916	4.535	2.961
9	104	−3.168	92.865	4.435	3.148
10	105	−2.361	11.916	3.940	3.213
11	104	−2.29	31.401	4.364	2.922
12	110	−2.843	15.875	6.073	3.614
13	104	−3.041	4.583	4.327	3.299
14	106	−2.91	4.796	5.818	3.095
15	107	−2.911	4.899	5.957	3.068
16	106	−3.538	13.342	5.958	3.069
17	107	−3.678	13.964	5.884	3.086
18	108	−4.061	14.56	5.885	3.085
19	107	−3.741	14.107	5.793	3.477
20	106	−3.779	40.212	5.919	3.032
21	108	−3.644	48.01	5.325	2.992
22	108	−3.846	15.133	5.314	2.966
23	108	−5.469	49.254	5.939	3.183
24	109	−4.126	46.109	5.612	3.105
25	109	−4.462	46.109	5.619	3.105
26	109	−4.689	165.042	5.734	3.086
27	108	−4.165	5.292	5.950	3.060
28	108	−4.372	15.264	5.809	3.093
29	105	−2.769	35.341	4.349	3.224
30	107	−2.536	120.818	5.031	2.963
31	107	−2.656	43.669	5.681	3.083
32	112	−3.981	66.408	5.851	3.216
33	105	−2.951	4.583	4.111	3.084
34	106	−3.367	12.728	4.040	2.993
35	109	−3.663	15.1	4.045	2.980
36	113	−3.747	18.166	4.044	2.988
37	109	−3.72	15.1	4.322	2.846
38	116	−4.831	20.421	4.037	2.997
39	111	−3.403	16.643	4.043	2.991
40	112	−3.64	17.407	4.044	2.985
41	102	−2.66	9.381	5.818	3.319
42	104	−2.847	92.865	4.147	3.541
43	108	−3.313	14.318	4.771	3.431
44	106	−2.33	4.796	5.982	3.114
45	105	−2.124	11.916	4.646	3.437
46	106	−3.72	5.196	5.957	3.063
47	106	−4.033	5.196	4.199	3.058
48	201	−3.029	20.273	5.690	2.926

Table 4 continued

NO	NALA	Mor05 m	SPI	D_4	D_1
49	202	−2.231	22.045	4.962	2.809
50	205	−2.459	27.074	4.751	3.084
51	208	−4.265	32.156	4.989	2.932
52	211	−3.722	37.63	5.359	3.023
53	206	−3.208	28.74	5.955	3.080
54	209	−3.525	33.926	5.332	2.922
55	212	−4.053	39.573	5.340	2.920
56	206	−3.988	30.43	5.582	2.901
57	208	−3.386	32.156	5.181	2.889
58	208	−3.096	32.156	5.426	2.900
59	208	−3.617	32.156	4.371	2.835
60	210	−3.626	35.749	4.486	3.005
61	213	−4.214	41.581	4.464	2.836
62	216	−4.2	48.021	5.335	2.923
63	214	−3.204	43.658	3.542	2.698
64	217	−4.847	50.309	4.417	2.955
65	219	−5.363	55.1	4.611	2.818
66	207	−4.038	10.954	5.009	3.008
67	207	−2.949	10.583	4.235	2.743
68	206	−5.198	30.364	5.950	3.105
69	206	−4.806	11.747	4.393	3.638
70	206	−4.739	32.909	4.217	2.737
71	206	−1.898	31.559	4.487	2.619
72	206	−4.629	42.024	4.752	3.147
73	209	−4.948	49.659	4.762	3.165
74	212	−3.977	57.454	5.124	3.105
75	208	−4.497	47.106	4.753	2.906
76	210	−3.908	52.23	4.969	2.889
77	213	−3.667	60.117	5.232	2.936
78	216	−3.279	68.345	4.925	2.925
79	206	−6.456	16.971	4.122	2.766
80	206	−1.284	31.559	6.337	2.616
81	208	−4.555	47.106	5.595	2.794
82	209	−5.314	55.588	5.597	2.790
83	212	−5.872	64.14	5.769	2.889
84	206	−2.732	43.012	5.775	3.066
85	406	−4.534	217.449	5.016	3.354
86	406	−4.203	217.449	5.148	3.656
87	406	−5.257	224.9	4.305	3.099
88	406	−4.765	224.9	4.502	3.554

the compounds and their stability constants with crown ethers.

The five descriptors that involved in the QSPR model I are, three new defined descriptors, one 3-D MoRSE descriptor (Mor05 m) and one topological descriptor (SPI). The first and most significant descriptor involved in model I is a new descriptor. It's a hybrid descriptor that calculated as the number of arms–length of longest arms (NALA),

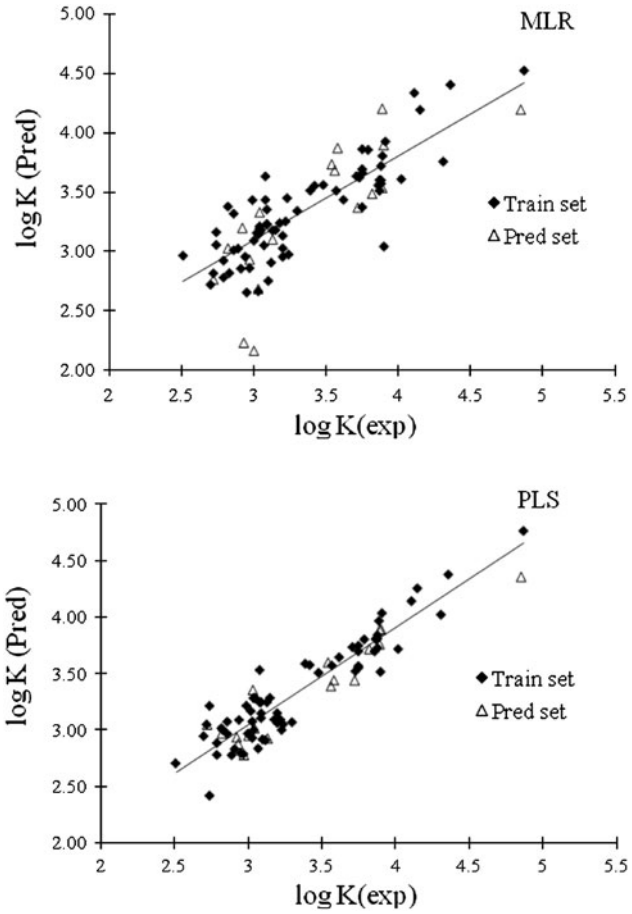


Fig. 2 Dispersion plot of the predicted versus experimental log K values for train and prediction sets for MLR and PLS model

which the first number of this descriptor denoted the number of arms and the second number denoted as length of longest arm on the crown ether ring. For example in Fig. 1 NALA descriptor was equal to 206, where 2 refers to two side arms and 6 refers to 6 bonds, from atom number 14 to 26, (as length of longest arm). This descriptor simply implies the effect of side arm and the length of side arms on stability constant. As shown in Table 5, this variable has a positive value in the regression coefficients, this indicate that this descriptor contributes positively to the value of log K. As discussed earlier, the possibility of three-dimensional cation encapsulation by the crown ether unit and the side arms and cooperative interaction between ring and side-arm, enhance the binding strength of lariat crown ethers toward cations in comparison with their crown ether analogues without side arms and thus increase the stability constant.

The second variable is 3-D MoRSE-signal 05/weighted by atomic masses (Mor05 m). The appearance of 3D-MoRSE descriptor that encodes the 3D structure of a molecule reveals the role of steric interactions of crown

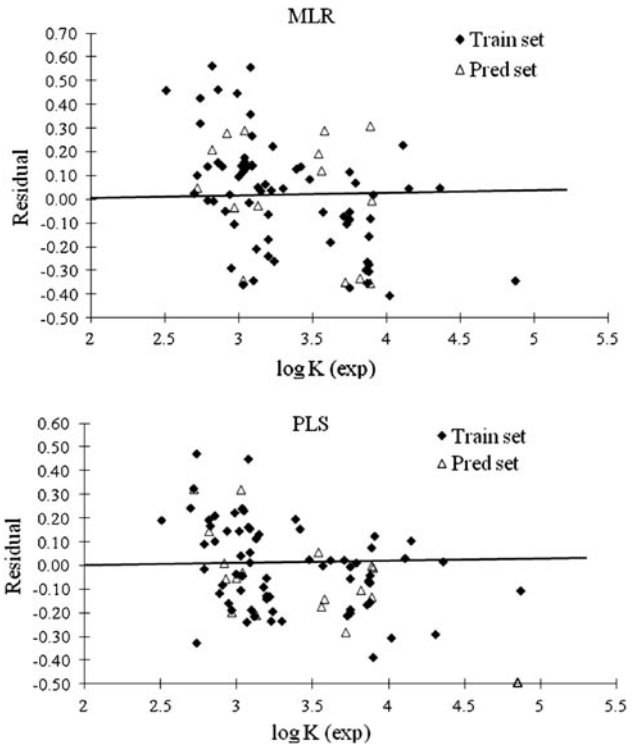


Fig. 3 Dispersion plot of the experimental log K values versus residuals for MLR and PLS models

ether in log K and the fact that this descriptor weighted by atomic masses, represent the role of bulkiness of a molecule in log K.

The third descriptor involved in QSPR model is superpendentic index (SPI). Superpendentic index takes into consideration all pendent vertexes. The index value changes significantly with a small change in the branching of a molecule [30]. This index is also a measure of molecular size, but focuses on the branched character of a molecular graph [31] and thus characterizes molecular bulk. Both of Mor05 m and SPI descriptor have negative value in the regression coefficients of the model.

The fourth and fifth significant descriptors in model I are D_4 and D_1, new defined descriptor, respectively. D_1 is distance of two adjacent oxygens, distance between O_{13} and O_1 in Fig. 1, and D4 is distance between two facing oxygens, O_{13} and O_7, in crown ether ring. In general, the efficiency and selectivity in host/guest binding strongly depends on shape and preorganization within the host molecule, the size-match of the host cavity to the guest [32]. The size-fit effect appears to play a subsidiary role in the complexation of metal ions by crown ethers. Alkali metal cations are spherical; therefore, they prefer a spherical donor atoms array in the host compound. In order to achieve this geometry some degree of preorganization in

Table 5 Molecular descriptors, their coefficients appeared in MLR equation for model I

Variable	Description of molecular descriptor	b	S_b	b_s	VIF
Intercept	–	1.472	0.539		–
NALA	number of arms–length of the longest arm	0.005	0.001	0.795	2.230
Mor05 m	3D-MoRSE-signal 32/weigthed by atomic masses	−0.192	0.032	−0.399	1.206
SPI	Superpendicet index	−0.004	0.001	−0.417	2.293
D_4	Distance between two faced oxygens	−0.144	0.043	−0.206	1.031
D_1	Distance between two adjacent oxygens	0.395	0.154	0.171	1.206

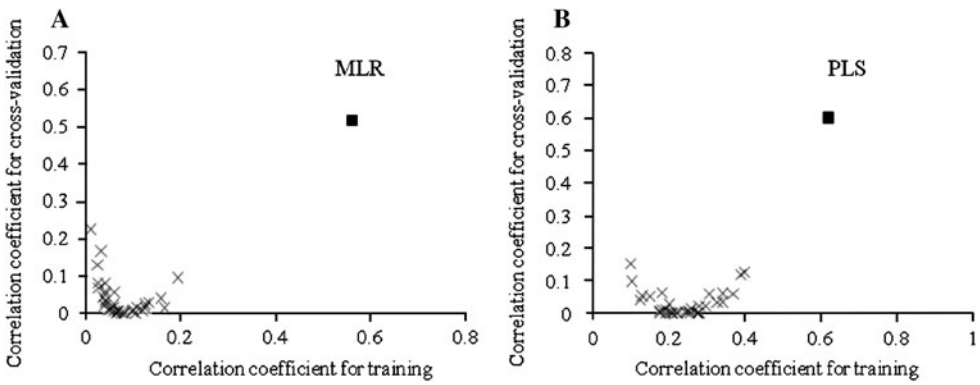


Fig. 4 Scatter plot for R^2 versus q^2 for the real models (*plain square*) and randomized models (*star*) **a**: MLR and **b**: PLS

the host molecule is required. These two descriptors, D_1 and D_4, indicate the shape and symmetry of cavity. D_4 has a negative value and D_1 has positive value in the regression coefficients, by increasing ring width (D_1) and decreasing ring length (D_4) crown cavity can achieved more symmetric shape and thus the stability constant was enhanced.

The five descriptors that appeared in the QSPR model II includes: four 3-D MoRSE descriptors involved 3-D MoRSE-signal 05/weighted by atomic masses (Mor05 m), 3-D MoRSE-signal 24/weighted by atomic van der Waals volumes (Mor24v), 3-D MoRSE-signal 32/unweighted (Mor32u) and 3-D MoRSE-signal 30/weighted by atomic van der Waals volumes (Mor30v) and one 2D autocorrelation descriptor, Moran autocorrelation–lag 4/weighted by atomic van der Waals volumes (MATS4v). As mentioned earlier, 3D-MoRSE descriptors have great potential for representation of molecular structure. They reflects the three-dimensional arrangement of the atoms of a molecule and do not care about chemical bonds.

Conclusions

We hereby report the novel descriptors to estimate lariat effect, the effects of different pendant-arm substituents and also conformation changing (shape and symmetry) of the crown ring of lariat and armed crown ethers. These new descriptors in combination with Dragon generated descriptors were used to build QSPR regression models, which correlated and predicted the stability constants (log K) of complexation between 15-crown-5 ethers and sodium ion (Na^+) in methanol solution. MLR and PLS techniques were used to construct QSPR models. We have showed that a combination of new defined descriptor and Dragon descriptors could lead to relevant QSPR models, in comparison with Dragon based descriptors, that allow us to improve the robustness of predictions.

References

1. Pedersen, C.J.: Cyclic polyethers and their complexes with metal salts. J. Am. Chem. Soc. **89**, 7017–7036 (1967)
2. Pedersen, C.J.: The discovery of crown ethers. Angew. Chem. Int. Ed. **27**, 1021–1027 (1988)
3. Chandler, K., Culp, C.W., Lamb, D.R., Liotta, C.L., Eckert, C.A.: Phase-transfer catalysis in supercritical carbon dioxide: kinetic and mechanistic investigations of cyanide displacement on benzyl chloride. Ind. Eng. Chem. Res. **37**, 3252–3259 (1998)
4. Leevy, W.M., Weber, M.E., Gokel, M.R., Hughes-Strange, G.B., Daranciang, D.D., Ferdani, R., Gokel, G.W.: Correlation of bilayer memebrane cation transport and biological activity in alkyl-substituted lariat ethers. Org. Biomol. Chem. **3**, 1647–1652 (2005)
5. Takayanage, T.: Analysis of ion-association reaction in aqueous solution and its utilization by capillary zone electrophoresis. Anal. Sci. **20**, 255–265 (2004)
6. Bartsch, R.A., Ivy, S.N., Lu, J., Huber, V.J., Talanov, V.S.: Metal ion extraction by lariat ethers with 'tunable' proton-ionizable groups. Pure Appl. Chem. **70**, 2393–2400 (1998)
7. Walkowiaka, W., Kozlowski, C.A.: Macrocycle carriers for separation of metal ions in liquid membrane processes—a review. Desalination **240**, 186–197 (2009)
8. Richens, D.A., Simpson, D., Peterson, S., McGinn, A., Lamb, J.D.: Use of mobile phase 18-crown-6 to improve peak resolution between mono- and divalent metal and amine cations in ion chromatography. J. Chromatogr. A **1016**, 155–164 (2003)
9. Bartsch, R.A., Eley, M.D., Utterback, M.D., McDonough, J.A., Jang, Y.: New multidentate ligand carriers for macrocycle-facilitated metal ion transport across liquid membranes. In: Bartsch, R.A., Way, J.D. (eds.) Chemical separation with liquid membranes, pp. 155–166. Am. Chem. Soc, Washington (1996)
10. Gokel, G. W., Leevy, W. M., Weber, M. E.: Crown ethers: sensors for ions and molecular scaffolds for materials and biological models. Chem. Rev. **104**, 2723–2750 (2004)
11. Lehn, J.-M.: Supramolecular chemistry: concepts and perspectives. VCH, Weinheim (1995)
12. Gokel, W., Dishong, D.M., Diamond, C.J.: Lariat ethers: synthesis and cation binding of macrocyclic polyethers possessing axially disposed secondary donor groups. J. Chem. Soc. Chem. Commun. **22**, 1053–1054 (1980)
13. Bartsch, R.A., Walkowiak, W., Charewicz, W.A.: Selective transport of alkali metal cations in solvent extraction by proton-ionizable dibenzocrown ethers. J. Coord. Chem. **27**, 75–85 (1992)
14. Inokuma, S., Yasuda, T.I., Araki, S., Sakai, S., Nishimura, J.: Synthesis and specific binding property of a new doubly armed crownophane. Chem. Lett. **23**, 201–208 (1994)
15. Goli, D.M., Dishong, D.M., Diamond, C.J., Gokel, G.W.: Lariat ethers. VI. evidence for intramolecular chelation in sodium and potassium cation binding by 15-crown-5, carbon-pivot lariat ethers. Tetrahedron Lett. **23**, 5243–5246 (1982)
16. Liu, Y., Han, B., Inoue, Y., Ouchi, M.: Complexation thermodynamics of crown ethers, calorimetric titration of cation complexation with some azacrown ethers. J. Org. Chem. **63**, 2144–2147 (1998)
17. Gatto, V.J., Arnold, K.A., Viscariello, A.M., Miller, S.R., Morgan, C.R., Gokel, G.W.: Syntheses and binding properties of bibracchial lariat ethers (BiBLEs): survey of synthetic methods and cation selectivities. J. Org. Chem. **51**, 5373–5384 (1986)
18. Schultz, R.A., White, B.D., Dishong, D.M., Arnold, K.A., Gokel, G.W.: 12-, 15-, and 18-membered-ring nitrogen-pivot lariat ethers: syntheses, properties, and sodium and ammonium cation binding properties. J. Am. Chem. Soc. **107**, 6659–6668 (1985)
19. Gokel, G.W.: Lariat ethers: from simple sidearms to supramolecular systems. Chem. Soc. Rev. **1**, 39–47 (1992)
20. Arnold, K.A., Mallen, J., Trafton, J.E., White, B.D., Fronczek, F.R., Gehrig, L.M., Gandour, R.D., Gokel, G.W.: Ring-sidearm cooperativity in cation inclusion complexes of 12-membered ring lariat ethers: effect of sidearm chain length and a clarification of long-sidearm binding strengths. J. Org. Chem. **53**, 5652–5657 (1988)
21. Ghasemi, J.B., Saaidpour, S.: QSPR modeling of stability constants of diverse 15-crown-5 ethers complexes using best multiple linear regression. J. Incl. Phenom. Macrocycl. Chem. **60**, 339–351 (2008)
22. Izatt, R.M., Pawlak, K., Bradshaw, J.S.: Thermodynamic and kinetic data for macrocycle interaction with cations and anions. Chem. Rev. **91**, 1721–2085 (1991)
23. ChemBioDraw, version 11.0, www.Cambridgesoft.com
24. Hyperchem, v.8.0.5: Hypercube Inc. http://www.hyper.com

25. Dragon web version, software version 3., www.disat.unimib.it/chm/, (2003)
26. Todeschini, R., Consonni, V.: Handbook of molecular descriptors. Wiley-VCH, Weinheim (2000)
27. Sumathy, K.: "Chemical descriptors" from QSAR world—A Strand Life Sciences Web Resource. http://www.qsarworld.com/insilico-chemistry-chemical-descriptors.php
28. Cho, D.H., Lee, S.K., Kim, B.T.: Quantitative structure–activity relationship (QSAR) study of new fluorovinyloxyacetamides. Bull. Korean Chem. Soc. **22**, 388–394 (2001)
29. van der Voet, H.: Comparing the predictive accuracy of models using a simple randomization test. Chemomet. Intell. Lab. **25**, 313–323 (1994)
30. Gupta, S., Singh, M., Madan, A.K.: Superpendentic index: a novel topological descriptor for predicting biological activity. J. Chem. Inf. Comput. Sci. **39**, 272–277 (1999)
31. Kauffman, G., Jurs, P.: QSAR and *k*-nearest neighbor classification analysis of selective cyclooxygenase-inhibitors using topologically-based numerical descriptors. J. Chem. Inf. Comput. Sci. **41**, 1553–1560 (2001)
32. Canceill, J., Lacombe, L., Collet, A.: Water-soluble cryptophane binding lipophilic guests in aqueous solution. J. Chem. Soc., Chem. Commun., **3**, 219–221 (1987)