Short communication

Quantitative Structure-Retention Relationship Analysis of Some Xylofuranose Derivatives by Linear Multivariate Method

Strahinja Z. Kovačević,* Lidija R. Jevrić, Sanja O. Podunavac Kuzmanović, Nataša D. Kalajdžija and Eva S. Lončar

[1] *Department of Applied and Engineering Chemistry, Faculty of Technology, University of Novi Sad, Bulevar cara Lazara 1, 21000 Novi Sad, Serbia*

** Corresponding author: E-mail: strahinjakovacevic@hotmail.com; phone:+381642839686; fax:+38121450413*

Received: 17-11-2012

Abstract

The relationship between retention behavior of eight 1,2-*O*-cyclohexylidene xylofuranose derivatives and their molecular characteristics was studied using chemometric Quantitative Structure–Retention Relationships (QSRR) approach. QSRR analysis was carried out on the retention parameter R_M^0, obtained by normal-phase thin-layer chromatography, by using molecular descriptors, as well as partition coefficient for *n*-octanol/water bi-phase system (log*P*). Molecular descriptors were calculated from the optimized structures. Principal component analysis (PCA) followed by hierarchical cluster analysis (HCA) and multiple linear regression (MLR) was performed in order to select molecular descriptors that best describe the retention behavior of the compounds investigated, and to determine the similarities between molecules. MLR equations, that represent the retention measure R_M^0 as a function of the *in silico* molecular descriptors were established. The statistical quality of the generated mathematical models was determined by standard statistical measures and *cross*-validation parameters. Obtained results indicate that previously mentioned mathematical models are statistically significant and can successfully predict retention behavior of examined xylofuranose derivatives.

Keywords: 1,2-*O*-cyclohexylidene xylofuranose derivatives; QSRR; Molecular descriptors; Multivariate data analysis; TLC.

1. Introduction

Cyclohexylidene acetals of monosaccharides belong to the group of the most common cyclic acetals used in carbohydrate chemistry.[1] Examined 1,2-*O*-cyclohexylidene xylofuranose derivatives are often used as key intermediates and starting compounds in organic synthesis of various biomolecules.[2–8] These derivatives are also of great interest for chromatography due to variety of their functional groups.[9]

It is well known that mechanisms of chromatographic separation are very complex and depend on many factors such as type of chromatographic system, physicochemical characteristics of analytes, experimental conditions, etc. Therefore, in order to understand chromatographic processes, it is very useful to establish mathematical models which can predict the retention behavior of analytes on the basis of their structural characteristics in applied chromatographic system. Determination of the correlations between molecular structure and retention behavior of molecules in different chromatographic systems is the main task of Quantitative Structure–Retention Relationships (QSRR) chemometric method.[10] Chemometric processing of chromatographic data can reveal systematic information both about the analytes (retention, physicochemical properties, etc.) and about the stationary phases studied (the molecular mechanism of separation).[11–14] In QSRR models, the retention (e.g. the retention parameter R_M^0) of solutes in specific chromatographic system is presented as a function of molecular descriptors of the analytes.[13,15] The main parameters used in QSRR studies are physicochemical parameters, non-specific parameters and topological indices.[16]

QSRR analysis is also applicable for prediction of the retention behavior of newly synthesized molecules[13,15,17] and quantitative comparison of separation pro-

perties of individual types of chromatographic layers.[18] QSRR studies are widely applied in high-performance liquid chromatography (HPLC), gas chromatography (GC) and thin-layer chromatography (TLC).[19] QSRRs in TLC are used for prediction of retention and determination of lipophilicity and other physicochemical constants.[20,21]

In the case of TLC, retention of an analyte is described by the R_M value defined by the Bate-Smith equation[15]:

$$R_M = \log[(1/R_f) - 1] \quad (1)$$

where R_f is the so-called retardation factor, defined as the ratio of the single zone distance and the solvent front. The value of R_M depends linearly on the logarithm of the concentration of the organic modifier in the mobile phase (φ) according to the following relation:

$$R_M = R_M^0 + S \cdot \varphi \quad (2)$$

where R_M^0 is the intercept and S is the slope. In this paper, R_M^0 factors of 1,2-*O*-cyclohexylidene xylofuranose derivatives, obtained by normal-phase (NP) TLC in four diferent mobile phases, were correlated with several molecular descriptors. For the QSRR models it is very important to select most suitable molecular descriptors for predicting retention. Hence, PCA was performed on molecular descriptors and retention factors (R_M^0) to reveal some similarities among studied compounds and to select adequate descriptors. HCA has been carried out in order to confirm the grouping of compounds already obtained by the PCA. Descriptors of analyzed molecules were calculated using suitable software for molecular design. Two molecular descriptors as independent variables, that have low value of intercorrelation coefficient, were used for constructing each statistically valid MLR model.

The objectives of the conducted QSRR analysis were to evaluate the retention data by multivariate statistical methods and to find the possible relationship between retention characteristics and the physicochemical parameters of the investigated 1,2-*O*-cyclohexylidene xylofuranose derivatives in order to understand the separation mechanism in the given chromatographic systems.

2. Materials and Methods

The QSRR analysis was performed in the following several steps: molecular structure optimization by computer software, molecular descriptors computation, molecular descriptors selection, structure-retention model generation using MLR method, and statistical validation.

2. 1. Thin-Layer Chromatography

The procedure of the TLC separation of the studied molecules and obtained retention parameters (R_M^0) are presented in literature.[9] For QSRR analysis R_M^0 values, obtained by using four different mobile phases (cyclohexane as diluent; ethyl acetate (EA), acetone (AC), dioxane (DI), and tetrahydrofuran (THF) as modifiers; φ = 0.3 for all modifiers) and silica gel as stationary phase, were chosen.

2. 2. Studied Compounds

The names of the compounds investigated are listed in Table 1, and their chemical structures are presented in Figure 1.

2. 3. Molecular Modeling and *in silico* Molecular Descriptors

The derivation of *in silico* molecular descriptors proceeds from the chemical structure of the compounds. In order to calculate the molecular descriptors, all molecules were drawn into ChemBioDraw Ultra version 12.0 program.[22] The 3D modeling of examined molecules was carried out using ChemBio3D Ultra version 12.0 software[22] running on AMD Sempron Processor 3000+. The obtained 3D models were subjected to energy minimization using molecular mechanics force field method (MM2). The cutoff for structure optimization was set at a gradient of 0.1 kcal/Åmol. The Austin Model 1 (AM1) was used for full geometry optimization of all structures until the root mean square (RMS) gradient reached a value smaller than 0.0001 kcal/Åmol using MOPAC.[23]

The values of molecular descriptors (Table 2) for each molecule in the data set were calculated using the

Table 1. The names of the examined molecules

No.	Name
1	1,2-*O*-cyclohexylidene-α-D-xylofuranose
2	1,2-*O*-cyclohexylidene-3-*O*-*p*-toluenesulfonyl- α-D-xylofuranose
3	1,2-*O*-cyclohexylidene-5-*O*-*p*-toluenesulfonyl- α-D-xylofuranose
4	1,2-*O*-cyclohexylidene-3,5-di-*O*-*p*-toluenesulfonyl-α-D-xylofuranose
5	5-*O*-benzoyl-1,2-*O*-cyclohexylidene-α-D-xylofuranose
6	5-*O*-benzoyl-1,2-*O*-cyclohexylidene-3-*O*-*p*-toluenesulfonyl-α-D-xylofuranose
7	3,5-di-*O*-acetyl-1,2-*O*-cyclohexylidene-α-D-xylofuranose
8	1,2-*O*-cyclohexylidene-3,5-di-*O*-methanesulfonyl-α-D-xylofuranose

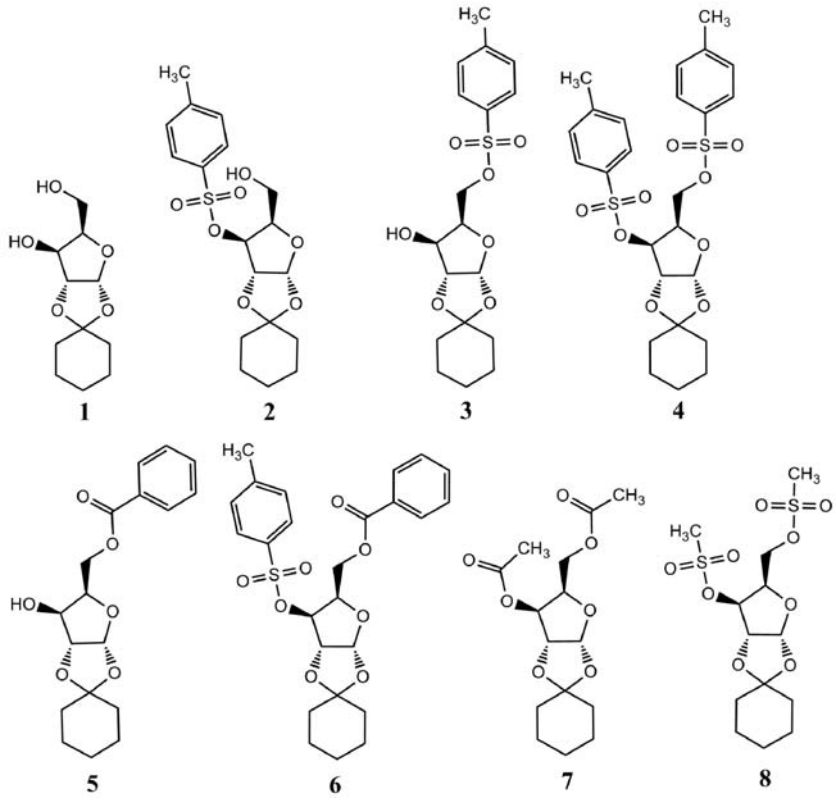

Figure 1. The chemical structures of the derivatives investigated

software ChemBio3D Ultra version 12.0, ALOGPS 2.1,[24] and MarvinSketch version 5.7.[25] Determined descriptors of examined compounds were topological descriptors (Wiener index – WI, molecular topological index – MTI), physicochemical descriptors (boiling point – BP, melting point – MP, critical pressure – CP, critical temperature – CT, critical volume – CV, ideal gas thermal capacity – IGTC, Gibbs energy distribution – GE, partition coefficients for *n*-octanol/water bi-phase system – logP_{ChDr}, Alog*P*, AClog*P*), molecular bulkiness descriptors (molar

Table 2. The values of the molecular descriptors for eight 1,2-*O*-cyclohexylidene xylofuranose derivatives

Molecule	MR [cm^3/mol]	TE [kcal/mol]	MTI	PSA [Å^2]	vdWSA [Å^2]	GE [kJ/mol]	BP [K]	CP [bar]	CT [K]
1	53.904	35.8094	2910	68.15	345.81	–377.32	627.848	36.820	801.525
2	91.847	41.5998	11808	91.29	551.70	–183.78	695.952	20.493	881.346
3	91.847	44.9932	13152	91.29	552.24	–183.78	692.201	20.493	876.595
4	129.790	49.1641	29962	114.43	759.63	9.76	760.305	13.033	955.980
5	83.229	58.5437	10654	74.22	484.93	–379.52	706.325	23.203	890.495
6	121.171	65.1244	25736	97.36	691.26	–185.98	774.429	14.381	969.920
7	72.207	53.8738	7252	80.29	473.46	–690.74	643.453	20.549	831.291
8	79.702	47.1155	9146	114.43	542.19	–296.84	589.667	19.753	792.319

Molecule	CV [cm^3/mol]	IGTC [J/mol · K]	MP [K]	WI	logP_{ChDr}	Alog*P*	AClog*P*
1	587.5	252.742	469.74	413	0.582	0.120	0.190
2	870.5	357.302	548.98	1625	2.797	2.170	1.510
3	870.5	357.302	548.98	1805	2.800	2.170	1.510
4	1153.5	461.862	628.22	4057	5.010	4.220	2.830
5	878.5	356.282	556.56	1469	2.710	2.170	2.150
6	1161.5	460.842	635.80	3497	4.920	4.210	3.480
7	825.5	350.260	477.84	1049	1.040	0.880	1.160
8	697.5	302.716	415.10	1311	–0.330	0.090	–0.540

refractivity – MR, total energy – TE, van der Waals surface area – vdWSA), and polarity descriptor (polar surface area – PSA).

2. 4. Multivariate Statistical Analysis and Model Validation

In QSRR analysis the main problem is how to reduce the number of variables and how to detect structure in the relationships between variables, that is to classify variables.[26] This can be done by various statistical methods of explorative analysis, classification methods and regression methods.[26,27] PCA and HCA are most often used explorative statistical methods.[26,28] Also, MLR is the most widely used regression method in QSRR.[15]

PCA is a technique for reducing the amount of data when there is correlation present. It is worth stressing that it is not a useful technique if the variables are uncorrelated.[29] PCA calculates latent, new variables by a combination of the original variables, representing the multidimensional data structure in an optimal way.[30] In a multidimensional space, where the variables define the axes, the data are projected into a few principal components (PCs) that are linear combinations of the original variables and describe the maximum variation within the data. Each PC is characterized by scores and loadings. Scores are the new coordinates of the projected objects, and loadings reflect the direction with respect to the original variables.[19] The loadings plot displays relationships between variables and can be used to identify variables (molecular descriptors in this study) which contribute to the positioning of the objects on the scores plot. The scores plot provides a data overview displaying patterns or groupings within the data.

HCA is a method for dividing a group of objects into classes so that similar objects are in the same class (cluster). As in PCA, the groups are not known prior to the mathematical analysis and no assumptions are made about the distribution of the variables. Cluster analysis searches for objects which are close together in the variable space. The data in each cluster share some common trait, often proximity according to some defined distance measure.[26]

The general purpose of MLR analysis is to quantitate the relationship between several independent or predictor variables and a dependent variable.[12] MLR model is built with descriptive variables using the least squares methods to minimize the residuals.[19] General MLR model is:

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \cdots + b_n \cdot x_n \quad (3)$$

where y is the quantitative property to predict (dependent variable), x_n an independent (descriptive) variable, a the intercept, and b_n the regression coefficient for x_n. The main restriction of MLR analysis is the case of large descriptors-to-compounds ratio or multicollinear descriptors in general.[27] For construction of MLR models it is very important to avoid multicollinearity. Variance Inflation Factor (*VIF*) is a diagnostic tool used to check the impact of multicollinearity in the MLR models. The *VIF* was calculated for indepenedent variables in each established MLR model according to equation[31,32]:

$$VIF_i = (1 - R_i^2)^{-1} \quad (4)$$

where R_i^2 is the coefficient of determination in a regression of the x_i independent variable on all other indepenedent variables in MLR model. The literature suggests that *VIF* greater than 10 indicates multicollinearity.[32–36]

Model validation is very important aspect of any QSRR analysis. The statistical quality of the generated MLR equations was measured by use of the standard statistical parameters (Pearson's correlation coefficient (r), F-test (Fisher's value), and the standard error of estimation (s)), and *cross*-validation parameters (*cross*-validated coefficient of determination (r^2_{cv}), adjusted coefficient of determination (r^2_{adj}), predicted residual sum of squares (*PRESS*), total sum of squares (*TSS*), and standard deviation based on predicted residual sum of squares (S_{PRESS})).[37,38] The correlation coefficient values closer to 1.0 represent the better fit of the regression, and high values of the F-test indicate that the model is statistically significant.[38] Standard deviation expresses the variation of the residuals or the variation about the regression line, and should have a low value for the regression to be significant.[37,38] The lower *PRESS* value is, the better the predictability of the model.[17,39] If *PRESS* value is less than *TSS* value, the model predicts better and can be considered statistically significant. *TSS* values are in terms of the dependent variable y. In many cases, r^2_{cv} and r^2_{adj} are taken as a proof of the high predictive ability of estimated mathematical models in QSRR. High values of these statistical characteristics (r^2_{cv}, $r^2_{adj} > 0.5$) indicate high predictivity of the equations.[31] Unlike r^2, r^2_{cv} may be negative, indicative of a very poor mathematical model, also unlike r^2, which tends to increase upon the addition of any descriptor, r^2_{cv} will decrease upon the addition of irrelevant descriptors.[40]

3. Results and Discussion

3. 1. PCA

In order to overview the data for similarities and dissimilarities, PCA has been carried out on the set of calculated molecular descriptors and retention data using Statistica v. 8 software.[41] Therefore, PCA can cluster compounds based on their structural and chromatographic features.

PCA was first performed on chromatographic data (R_M^0 values) and resulted in a two-component model that explains 99.44% of the data variation. The first principal

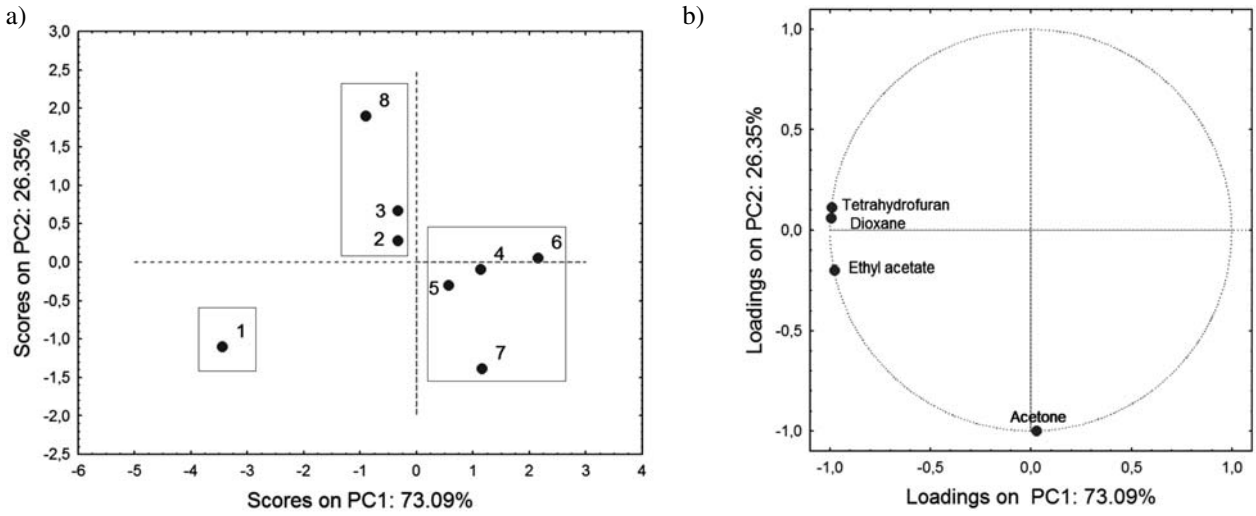

Figure 2. Score values (a) and factor loadings (b) of retention parameters for the first two PCs.

component explains up to 73.09% of the variability, and the second accounts for up to 26.35%. Figure 2 shows score values and the mutual projections of the loading vectors for the first two PCs.

The loading graph indicates the highest negative impact of systems with tetrahydrofuran, dioxane, and ethyl acetate along the PC1 direction, and acetone along the PC2 direction. The obtained results show that PC1 separate examined compounds according to their retention which is caused by the polarity and solubility of the substituents in applied mobile phases.[9] Along the PC1 direction, retention of the examined compounds decreases. Loading plot highlights the most influential chromatographic systems responsible for such retention order. In this case the loading graph does not reveal any significant influence of the mobile-phase composition along the PC2 direction.

The PCA performed on descriptors resulted in a three-component model that explains 98.33% of total variance. It reveals a quite different classification of compounds. First PC comprises 79.72% of the total data variability, and the second 11.35%. Scores graph (Figure 3a) revealed that the classification of examined molecules was achieved based on the structural characteristic: the presence of the voluminous aromatic substituents (*p*-toluenesulfonyl and benzoyl groups). Going along the PC1 axis from its negative end towards positive values, compounds which contain two aromatic substituents (4 and 6) are positioned very close to each other, and well separated from the rest of the compounds. Compounds 2, 3 and 5 contain one aromatic substituent, and compounds 7 and 8 have two small non-aromatic substituents (acetyl and methanesulfonyl groups). Compound 1 is unsubstituted and is positioned at the positive end of the PC1 axis.

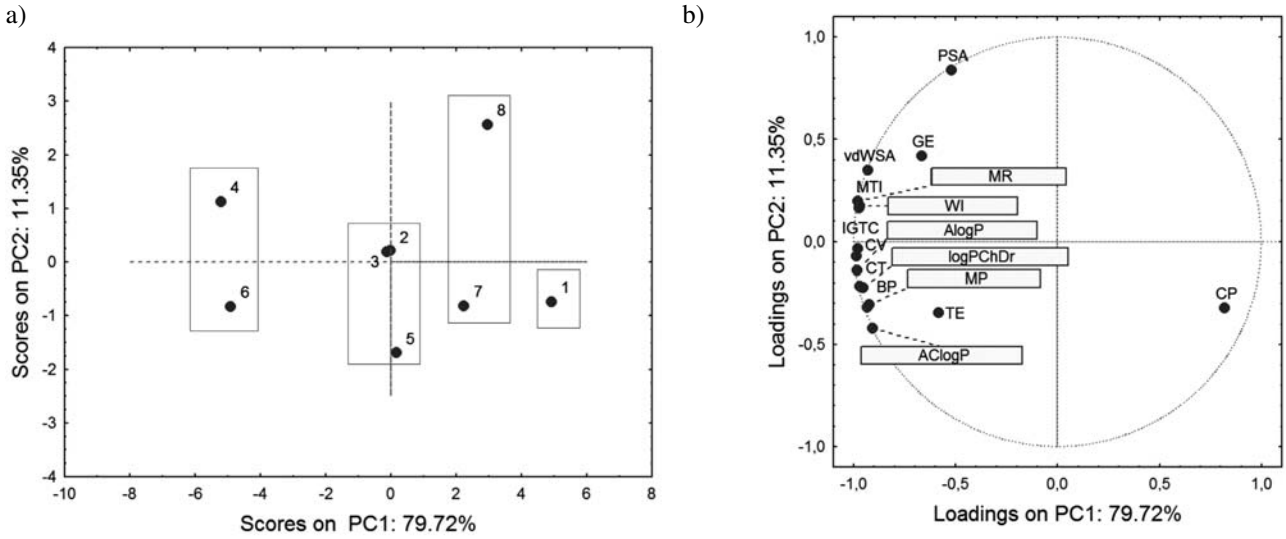

Figure 3. Score values (a) and factor loadings (b) of molecular descriptors for the first two PCs.

As it can be seen from the loading graph (Figure 3b), the majority of descriptors have a significant negative impact on PC1, while only CP has a positive influence. On the basis of the obtained plots (Figure 3) and molecular structures (Figure 1) of analyzed compounds, it can be concluded that the molecular volume is discriminating factor between compounds, because the majority of the calculated molecular descriptors mainly depends on molecular volume (molecular size).[42–45]

3. 2. HCA

HCA has been performed using NCSS 2007 and GESS 2006 Statistical Software[46] in order to confirm the grouping of compounds already obtained by the PCA. Clustering is based on the Euclidean distance and Ward's linkage algorithm.

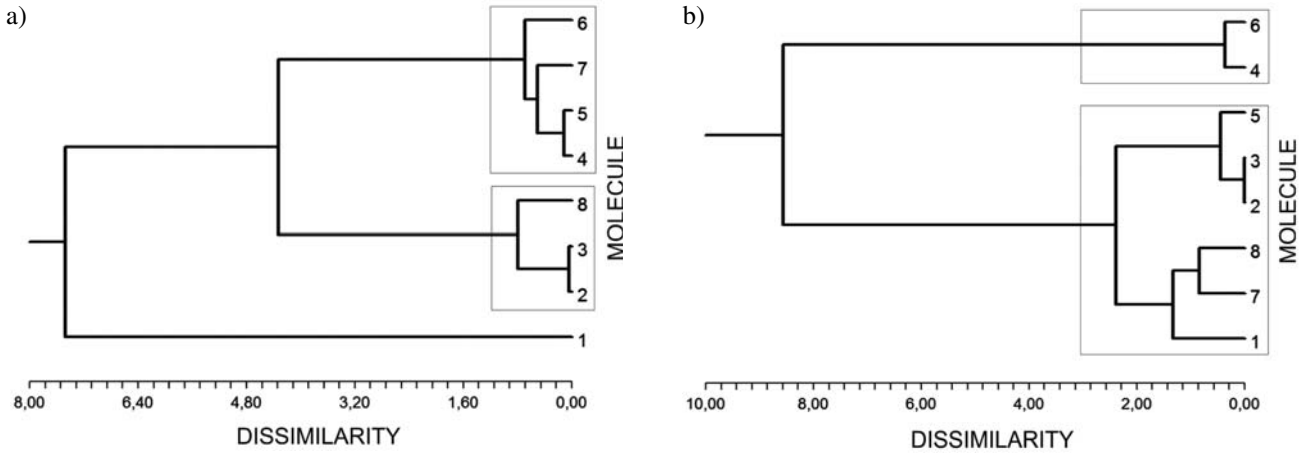

Figure 4. Dendograms of 8 examined compounds in the space of 4 chromatographic systems (a) and 16 molecular descriptors (b).

As it can be observed from Figure 4a, dendogram based on the retention parameters shows two well-separated clusters and compound 1 out of clusters. Clustering of the compounds on the obtained dendogram is based on their retention characteristics and it is the same as on the PC1-PC2 score plot (Figure 2a).

The cluster analysis performed on descriptors resulted in two main clusters (Figure 4b). The first cluster is made of compounds 4 and 6, that have two aromatic substituents, while the second cluster with substructures contains unsubstituted compound 1, and compounds with one aromatic substituent (2, 3 and 5) and two non-aromatic substituents (7 and 8). It is obvious that compounds in HCA are grouped in the same way as in PCA (Figure 3a).

3. 3. MLR

MLR analysis has been carried out to derive the best QSRR models which can predict retention behavior of investigated molecules based on selected molecular descriptors. It is very important to define the number of independent variables in the model equation, because in this way the over-parameterization of the mathematical model as well as the chance correlation between the descriptors is avoided.[47] In this study as independent variables two descriptors were selected according to number of molecules investigated. The stepwise regression routine showed which two-descriptor combinations form the MLR models characterized by the highest correlation coefficient.

The software package used for conducting MLR analysis was NCSS 2007 and GESS 2006. The descriptors obtained by stepwise regression routine served as the input data for MLR analysis. The correlation coefficients among selected descriptors are presented in Table 3.

As a result of MLR analysis, three statistically significant equations, free of multicollinearity (*VIF* < 10), were obtained (Table 4). The statistical validity of the established models, as depicted in Table 4, was determined by *r*, *F*, and *s*. The *F*-value is found statistically significant at 99% level since all the calculated *F* values are higher as compared to tabulated values.

Table 3. Correlation matrix for molecular descriptors used in MLR analysis

	CP	TE	GE	IGTC	PSA
CP	1				
TE	–0.6097	1			
GE	–0.4380	–0.0840	1		
IGTC	–0.8648	0.6618	0.5280	1	
PSA	–0.7534	0.1294	0.6193	0.5033	1

Positive values in regression coefficient indicate that observed descriptor contributes positively to the value of $R_M{}^0$, whereas negative values indicate that the greater the value of descriptor, the lower the value of $R_M{}^0$. Based on the

Table 4. Best MLR models for prediction of retention behavior of 1,2-*O*-cyclohexylidene xylofuranose derivatives

Modifier	Variables			Multiple Linear Regression: $y = a \cdot x_1 + b \cdot x_2 + c$							Eq.
	y	x_1	x_2	a	b	c	r	F	s	VIF	
EA	R_M^0	CP	TE	0.0391	–0.0137	–1.1295	0.9895	116.90	0.0648	1.6	(5)
EA	R_M^0	CP	PSA	0.0669	0.0095	–3.2617	0.9928	172.18	0.0536	2.3	(6)
THF	R_M^0	GE	IGTC	0.0008	–0.0050	0.6460	0.9788	57.10	0.0755	1.4	(7)

Table 5. *Cross*-validation parameters for models 5–7

Eq.	r^2_{cv}	r^2_{adj}	*PRESS*	*TSS*	*PRESS/TSS*	S_{PRESS}
5	0.9417	0.9707	0.0586	1.0042	0.0584	0.0856
6	0.9658	0.9800	0.0343	1.0042	0.0342	0.0655
7	0.8719	0.9413	0.0871	0.6801	0.1281	0.1043

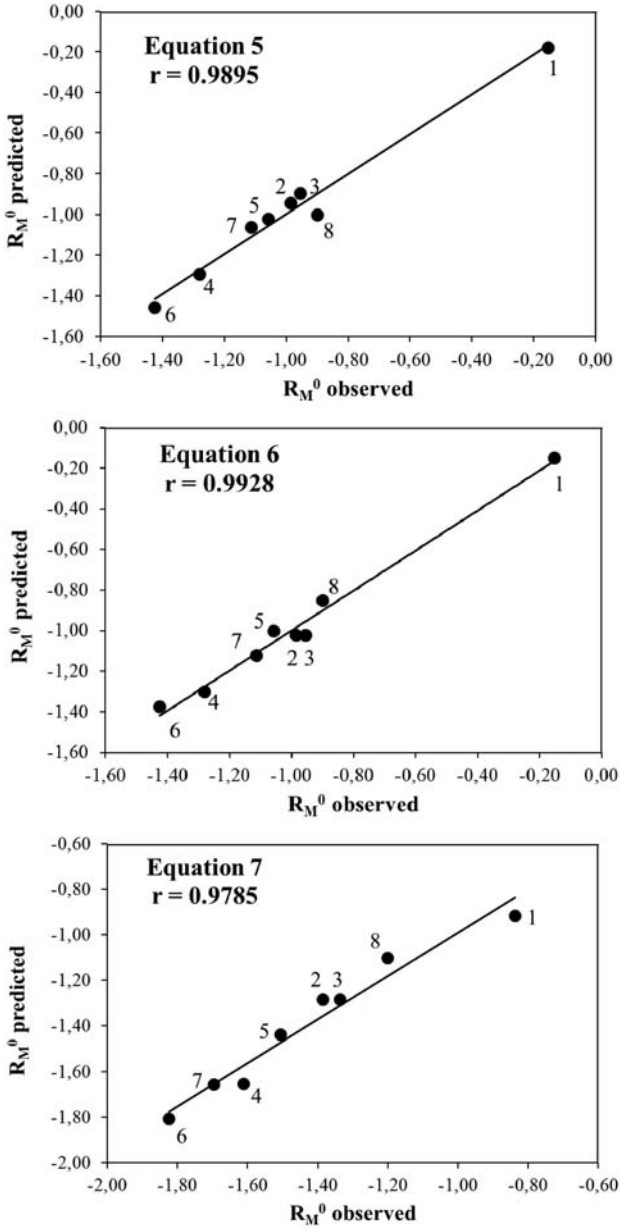

Figure 5. Graphs of experimental vs. predicted R_M^0 values according to equations 5–7.

chosen descriptors and formed MLR models, it can be observed that retention of derivatives examined by adsorption chromatography is best described by physicochemical (thermodynamic) descriptors (CP, GE, IGTC) and molecular bulkiness descriptor (TE), including polarity parameter (PSA).

Equations 5–7 were *cross*-validated by the leave-one-out (LOO) method (Table 5). High values of r^2_{cv} and r^2_{adj}, and low *PRESS* value (significantly less than the *TSS*), were obtained for all the models, indicating that these models have outstanding predictive power.

To confirm our finding, R_M^0 values were calculated from the established models 5–7, and graphically compared with experimental data (Figure 5). Low scattering of points around the linear relationship, significant slope (>0.95), and intercept close to zero (<0.0407), indicate very good concurrence between experimental values of retention parameters and values obtained by defined mathematical models.

Also, on the basis of the magnitude of the residues there is close agreement between observed and calculated retention constants (Figure 6).

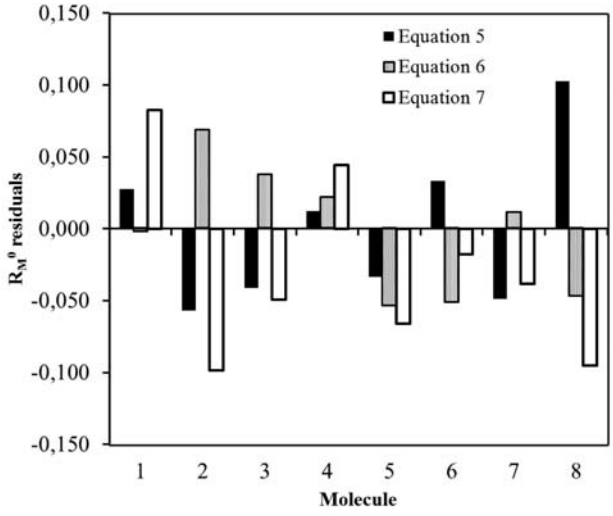

Figure 6. Plot of the residual values against the experimentally observed R_M^0 values for each molecule.

The presented results indicate that MLR analysis combined with a successful variable-selection procedure enables forming of efficient QSRR models for predicting the retention constants of 1,2-*O*-cyclohexylidene xylofuranose derivatives. All these results suggest that the chromatographic behavior of examined molecules depends on molecular descriptors, and the retention constants can be accurately predicted.

4. Conclusion

In this study the focus of QSRR analysis was to identify the most important descriptors affecting normal-phase chromatographic behavior of 1,2-*O*-cyclohexylidene xylofuranose derivatives on silica gel thin layer. For this purpose PCA and HCA followed by MLR were performed. These multivariate statistical methods revealed that analytes can be classified according to their structural characteristics. Established MLR models are statistically significant and free of relevant multicollinearity. CP, TE, GE, IGTC, and PSA are most appropriate molecular descriptors for prediction of the chromatographic retention constant R_M^0. The best statistical results were obtained with ethyl acetate as the modifier. Predictive ability of formed mathematical models based on physically meaningful molecular parameters allows us to estimate and understand retention behavior of structurally similar compounds.

5. Acknowledgement

These results are the part of the project No.172012 and No.172014, supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia and the project No. 114-451-2373/2011, financially supported by the Provincial Secretariat for Science and Technological Development of Vojvodina.

6. References

1. S. Hanessian, Preparative Carbohydrate Chemistry, Marcel Dekker Inc, New York, USA, **1996**.
2. M. Popsavin, V. Popsavin, N. Vukojević, D. Miljković, *Collect. Czech. Chem. Commun.* **1994**, *59*, 1884–1888.
3. V. Popsavin, I. Krstić, M. Popsavin, *Tetrahedron Lett.* **2003**, *44*, 8897–8900.
4. P. Hadžić, N. Vukojević, M. Popsavin, J. Čanadi, *J. Serb. Chem. Soc.* **2001**, *66* (1), 1–8.
5. P. Hadžić, N. Vukojević, *J. Serb. Chem. Soc.* **2001**, *66* (5), 289–295.
6. V. Popsavin, S. Grabež, I. Krstić, M. Popsavin, D. Djoković, *J. Serb. Chem. Soc.* **2003**, *68* (11), 795–804.
7. D. A. Miljković, N. S. Vukojević, P. A. Hadžić, D. J. Minić, N. A. Hughes, M. N. Stuart Hill, J. Harangi, *J. Chem. Soc., Perkin Trans. 2* **1990**, *7*, 1093–1095.
8. P. Hadžić, N. Vukojević, *Carbohydr. Res.* **2003**, *338* (11), 1243–1249.
9. S. M. Petrović, E. Lončar, M. Popsavin, V. Popsavin, *J. Planar Cromatogr.* **1997**, *10*, 406–410.
10. R. Kaliszan, Structure and Retention in Chromatography: A Chemometric Approach, Harwood Academic Publishers, Australia, **1997**.
11. M. A. Al-Haj, R. Kaliszan, A. Nasal, *Anal. Chem.* **1999**, *71*, 2976–2985.
12. J. Ghasemi, S. Saaidpour, *J. Chromatogr. Sci.* **2009**, *47*, 156–163.
13. D. J. Livaja Popović, E. S. Lončar, L. R. Jevrić, R. V. Malbaša, *Hem. Ind.* **2012**, *66* (3), 365–372.
14. L. R. Jevrić, G. B. Koprivica, N. M. Mišljenović, B. Ž. Jovanović, *APTEFF* **2010**, 41, 159–168.
15. L. R. Jevrić, G. B. Koprivica, N. M. Mišljenović, A. N. Tepić, T. A. Kuljanin, B. Ž. Jovanović, *APTEFF* **2011**, *42*, 231–239.
16. S. Šegan, F. Andrić, A. Radoičić, D. Opsenica, B. Šolaja, M. Zlatović, D. Milojković Opsenica, *J. Sep. Sci.* **2011**, *34*, 2659–2667.
17. N. U. Perišić Janjić, S. O. Podunavac Kuzmanović, *J. Planar Chromatogr*. **2008**, *21*, 135–141.
18. R. Kaliszan, *Chem. Rev.* **2007**, *107*, 3212–3246.
19. J. Trifković, F. Andrić, P. Ristivojević, D. Andrić, Ž. Lj. Tešić, D. M. Milojković Opsenica, *J. Sep. Sci.* **2010**, *33*, 2619–2628.
20. K. Héberger, *J. Chromatogr. A*, **2007**, *1158*, 273–305.
21. A. Nasal, D. Siluk, R. Kaliszan, *Curr. Med. Chem.* **2003**, *10*, 381–426.
22. ChemBioOffice 2010, PerkinElmer Informatics, 2175 Mission College Boulevard, Santa Clara, California, USA, http://www.cambridgesoft.com
23. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
24. VCCLAB, Virtual Computational Chemistry Laboratory, http://www.vcclab.org
25. ChemAxon Kft. Záhony u. 7, Building HX 1031 Budapest, Hungary http://www.chemaxon.com/products/marvin/marvinsketch/
26. T. Đaković Sekulić, Z. Lozanov Crvenković, N. Perišić Janjić, *Novi Sad J. Math.* **2008**, *38* (3), 39–46.
27. A. Z. Dudek, T. Arodz, J. Gálvez, *Comb. Chem. High Throughput Screening* **2006**, *9*, 213–228.
28. Z. Királyová, K. Šnuderl, F. Kraic, J. Mocak, *Acta Chim. Slov.* **2008**, *1* (1), 143–152.
29. J. N. Miller, J. C. Miller, Statistics and Chemometrics for Analytical Chemistry, 6th edition, Pearson Education Limited, Harlow, UK, **2010**.
30. K. Šnuderl, M. Simonič, J. Mocak, D. Brodnjak Vončina, *Acta Chim. Slov.* **2007**, *54*, 33–39.
31. D. A. Belsley, *Am. Stat.* **1984**, *38*, 73–82.
32. D. W. Marquardt, R. D. Snee, *Am. Stat.* **1970**, *29*, 3–19.
33. R. M. O'Brien, *Quality & Quantity* **2007**, *41*, 673–690.
34. D. A. Belsley, E. Kuh, R. E. Welsch, Regression Diagnostics:

Identifying Influential Data and Sources of Collinearity, John Wiley, New York, **1980**.
35. G. R. Pasha, M. Akbar Ali Snah, *J. Res. (Sci.)* **2004**, *15* (1), 97–106.
36. T. M. Young, L. B. Shaffer, F. M. Guess, H. Bensmail, R. V. León, *For. Prod. J.* **2008**, *58* (4), 39–48.
37. S. O. Podunavac Kuzmanović, D. D. Cvetković, S. B. Gadžurić, *APTEFF* **2011**, *42*, 251–261.
38. K. K. Sahu, V. Ravichandran, P. K. Jain, S. Sharma, V. K. Mourya, R. K. Agrawal, *Acta Chim. Slov.* **2008**, *55*, 138–145.
39. O. Farkas, J. Jakus, K. Heberger, *Molecules* **2004**, *9*, 1079–1088.
40. S. C. Basak, Q. Zhu, D. Mills, *Acta Chim. Slov.* **2010**, *57*, 541–550.
41. StatSoft Inc., 2300 East 14th Street, Tulsa, Oklahoma, USA: http://www.statsoft.com/
42. M. L. Connolly, *J. Am. Chem. Soc.* **1985**, *107* (5), 1118–1124.
43. R. Mannhold (Ed.), Molecular Drug Properties – Measurement and Prediction, Vol. 37, Wiley-VCH Verlag GmbH & Co., Weinheim, **2008**.
44. R. P. Schwarzenbach, P. M. Gschwend, D. M. Imboden, Environmental Organic Chemistry, Wiley, New York, USA, **2003**.
45. A. G. Saliner, E. Burello, A. Worth, Review of computational approaches for predicting the physicochemical and biological properties of nanoparticles, http://ihcp.jrc.ec.europa.eu/our_labs/predictive_toxicology/doc/EUR_23974 _EN.pdf, (assessed: November 11, 2012)
46. NCSS Statistical Software, http://www.ncss.com/
47. N. Minovski, A. Jezierska-Mazzarello, M. Vračko, T. Šolmajer, *Cent. Eur. J. Chem.* **2011**, *9* (5), 855–866.

Povzetek

QSRR metoda je bila uporabljena s ciljem ugotavljanja odnosa med retencijskim obnašanjem in molekularnimi značilnostmi osmih derivatov 1,2-*O*-cikloheksiliden ksilofuranoze. QSRR analiza retencijskega parametra R_M^0, ki je bil eksperimentalno pridobljen s tankoslojno kromatografijo, je bila izvedena z uporabo molekularnih deskriptorjev in particijskega koeficienta (log*P*). Fizikalno-kemične deskriptorje smo izračunali iz optimiranih struktur. Metoda glavnih osi, metoda hierarhičnega razvrščanja in postopek multiple linearne regresije so bili uporabljeni za določanje molekularnih deskriptorjev, ki najbolje opisujejo retencijske lastnosti raziskanih spojin, ter za določanje podobnosti med molekulami. Dobljene so enačbe, ki predstavljajo retencijski parameter R_M^0 v funkciji *in silico* molekularnih deskriptorjev in parametrov lipofilnosti. Kakovost dobljenih matematičnih modelov je bila določena s standardnimi statističnimi analizami in navzkrižno validacijo parametrov. Rezultati dokazujejo, da so omenjeni matematični modeli statistično značilni, ter da lahko uspešno napovedujejo retencijsko obnašanje raziskovanih derivatov ksilofuranoze.