Scientific paper

QSRR Modeling of Retention Behavior of Some s-Triazine Derivatives

Lidija R. Jevrić,* Sanja O. Podunavac-Kuzmanović, Jaroslava V. Švarc-Gajić, Aleksandra N. Tepić, Strahinja Z. Kovačević and Nataša D. Kalajdžija

Department of Applied and Engineering Chemistry, Faculty of Technology, University of Novi Sad, Bul. Cara Lazara 1, 21000 Novi Sad, Serbia

** Corresponding author: E-mail: lydija@uns.ac.rs phone:+381214853648; fax:+38121450413*

Received: 02-01-2013

Abstract

The properties relevant to lipophilicity of four series of synthesized s-triazine derivatives have been studied by quantitative structure-retention relationship (QSRR) approach. Examination of chromatographic behavior revealed a linear correlation between R_M values and the volume fraction of mobile phase modifier. Furthermore, a reliable relationship was defined between the retention constants, R_M^0, and theoretically calculated bioactivity descriptors for lipophilicity and solubility. Principal component analysis (PCA) followed by multiple linear regression (MLR) and hierarchical cluster analysis (HCA) was performed to identify the most important factors, to quantify their influences, and to select descriptors that best describe the behavior of the compounds investigated. The best QSRR models were further validated by leave one out technique as well as by the calculation of statistical parameters for the established theoretical models. The R_M^0 values of the investigated s-triazine derivatives have been recommended for description of their lipophilicity and evaluation of pharmacokinetic properties.

Keywords: QSRR, s-triazine derivatives, physico-chemical properties, multiple linear regression.

1. Introduction

1,3,5-Triazine (s-triazine) derivatives belong to very attractive group of chemicals having an important role in agriculture as active components of herbicide formulations, as well as in pharmacology. The effects that are widely exploited in drug industry are their anticonvulsant properties, as well as potent antimalarial and bactericidal activity.[1–4] In addition, their anticancer activity has been demonstrated as well. For their rich biological activity, pharmaceutical industry is each day trying to introduce novel derivatives that could improve drug properties and reduce side effects. Even though having a wide array of therapeutical effects in humans, carcinogenic and mutagenic effects of some s-triazine derivatives on living organisms have been demonstrated.[5–7]

s-Triazine is a weak base with six-membered heterocyclic ring containing three nitrogens replacing carbon-hydrogen units in the benzene ring. The compound, so as its derivatives, has an excellent potential for the formation of non-covalent bonds, such as coordination and H-bonds, via its nitrogen ion-pairs.[8] Non-covalent bonds have a very important role in biological activity of these compounds, but also in understanding of their physiological behavior, namely absorption, metabolism and elimination. Furthermore, such chemical properties of s-triazines are responsible for their characteristic chromatographic behavior.[9]

Molecular lipophilicity is one of the major physicochemical properties. Widespread use of lipophilicity in modeling of biological processes explains the need for rapid and valid procedures for quantification of this physicochemical property.

Chromatographic approach has been shown to be quite successful in modeling physicochemical and biological processes.[10–12] Owing to its simplicity and efficiency, reversed-phase thin-layer chromatography (RP TLC) appears especially attractive for lipophilicity determination.[13,14] Taking into consideration that in reversed-phase chromatography solutes distribute between polar and nonpolar phases, calculated retention parameters can be adopted as indirect designators of compounds lipophilicity.

For the quantitative structure-retention relationships (QSRR) models it is very important to select most suitable structural descriptors for predicting retention. Hence, principal component analysis (PCA) was performed on molecular descriptors and retention factors (R_M^0) to reveal some similarities among studied compounds and to select adequate descriptors. Hierarchical cluster analysis (HCA) has been carried out in order to confirm the grouping of compounds already obtained by the PCA.[15] Descriptors of analyzed molecules were calculated using software for molecular design. Two molecular descriptors, that have low value of intercorrelation coefficient, were used for constructing each statistically valid multiple linear regression (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 s-triazine derivatives in order to understand the separation mechanism in the given chromatographic systems.

2. Experimental

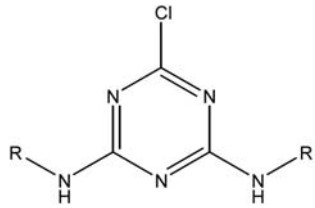
2. 1. Synthesis of Compounds

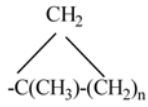
The investigated compounds were 1,3,5 - triazines substituted at positions 4 and 6 by smaller and larger groups with various lipophilic characteristics, chosen for investigation are presented in Table 1. The compounds were synthesized in the laboratory of the Department of Organic Chemistry at the Faculty of Technology and Metallurgy, University of Belgrade. All of investigated s-triazine derivatives were synthesized by the modified procedure of Thurston from cyanuric chloride and corresponding amines.[16, 17] In synthesis commercial cyanuric chloride (2,4,6-trichloro-1,3,5-triazine), was used (Fluka).

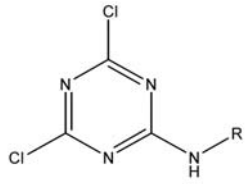
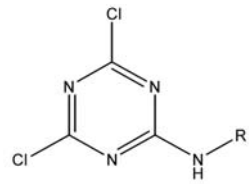
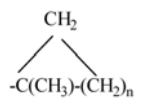
5 g cyanuric-chloride was dissolved in 100 g of pure ether, cooled with cooling mixture on temperature below 0 °C. In this solution, with constant agitation and maintenance of low temperature, dry gaseous ammonia was introduced until its scent appeared in the reaction mixture and after subsequent shaking. Precipitate, which was mainly consisted of ammonium chloride, was immediately separated with filtration, and filtrate was evaporated under reduced pressure. The residual white precipitate, for a better purification, was dissolved in ether, once again. Separation of small amount of insoluble matter was performed with filtration procedure and the filtrate was evaporated, again. All s-triazine derivatives were crystallized from aqueous methanol in a form of small, fluffy needles.[18, 19]

The chemical structures and the purities of the synthesized s-triazine derivatives were confirmed by melting points, mass spectra and UV spectra.

Table 1. The chemical structures of studied s-triazines



Series I		Series II		
Compound	**R**	*Compound*	**R**	***n***
I.1	-CH(CH_3)-C_6H_5	II.1		3
I.2	-CH(CH_3)-C_6H_4-4-CH_3	II.2		4
I.3	-CH(CH_3)-C_6H_4-4-Cl	II.3		5
I.4	-CH(CH_3)-C_6H_4-4-Br			

Series III				Series IV		
						
Compound	**R**	**R**	**R**	*Compound*	**R**	***n***
III.1	C_6H_{11}	H	H	IV.1		3
III.2	C_6H_{11}	CH_3	CH_3	IV.2		4
III.3	C_6H_{11}	C_6H_5	H	IV.3		5
III.4	C_6H_{11}	C_6H_5	C_6H_5			

2. 2. Thin Layer Chromatography

Precoated RP-18W/UV_{254} plates (Macherey-Nagel GmbH and Co., Düren, Germany) were used for HP TLC analysis. Investigated solvent mixtures used as mobile phases according a procedure described earlier.[20]

The investigated compounds were dissolved in an appropriate solvent, methanol, 1 mg ml^{-1}) and the solutions (0.2 µl) were separately spotted into the plates. All the reagents used were of analytical purity. The plates were developed by the ascending technique at room temperature with previous saturation of the chamber with mobile phase. All measurements were carried out at ambient temperature. After drying of the plates, the spots were visualized under UV light at λ = 254 nm. R_F values were calculated as average from three measurements for each solute-mobile phase combination. For subsequent calculations mean R_M values were used; these were calculated by using the formula:

$$R_M = \log(1/R_F - 1) \quad (1)$$

The calculated R_M values for different concentrations of organic solvent were used to check the linearity of their relationship with the volume fraction of organic modifier according to the equation: [21]

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

where φ is the volume fraction of organic solvent in the mobile phase, $R_M{}^0$ is the intercept obtained by extrapolation to $\varphi = 0\%$ of modifier, and S is the slope of the linear plot. Equations (1) and (2) served for deriving data for further QSRR studies.

For the testing the validity of the predictive power of selected MLR models the the leave-one-out technique (LOO) was used. The developed models were validated by the calculation of following statistical parameters: predictive residual error (PRESS), sum of square of deviation (SSY), S_{PRESS}, cross-validated regression coefficient (r^2_{CV}), and adjusted r-squared (r^2_{adj}). These parameters were calculated from the following equations.

$$\mathrm{PRESS} = \sum (Y_{obs} - Y_{calc})^2 \quad (3)$$

$$\mathrm{SSY} = \sum (Y_{obs} - Y_{mean})^2 \quad (4)$$

$$S_{PRESS} = \sqrt{\frac{\mathrm{PRESS}}{n}} \quad (5)$$

$$r^2_{CV} = 1 - \frac{\mathrm{PRESS}}{\mathrm{SSY}} \quad (6)$$

$$r^2_{adj} = 1 - (r^2)\left(\frac{n-1}{n-p-1}\right) \quad (7)$$

where, Y_{obs}, Y_{calc} and Y_{mean} are observed, calculated and mean values; n, number of compounds; p, number of independent parameters.

PRESS is an acronym for prediction sum of squares. It is used to validate a regression model in predictability. To calculate PRESS, each observation is individually omitted. The remaining n-1 observations are used to calculate a regression and estimate the value of the omitted observation. This is done n times, once for each observation. The difference between the actual Y value, y_{obs}, and the predicted Y, y_{calc}, is called the prediction error. The sum of the squared prediction errors is the PRESS value. The smaller PRESS is, the better the predictability of the model. Its value being less than SSY points out that the model predicts better than chance and can be considered statistically significant. SSY are the sums of squares associated with the corresponding sources of variation. These values are in terms of the dependent variable, y.

The PRESS value above can be used to compute an r^2_{CV} statistic, called r^2 cross validated, which reflects the prediction ability of the model. This is a good way to validate the prediction of a regression model without selecting another sample or splitting your data. It is very possible to have a high r^2 and a very low r^2_{CV}. When this occurs, it implies that the fitted model is data dependent. This r^2_{CV} ranges from below zero to above one. When outside the range of zero to one, it is truncated to stay within this range. Adjusted r-squared (r^2_{adj}) is an adjusted version of r^2. The adjustment seeks to remove the distortion due to a small sample size.

In many cases r^2_{CV} and r^2_{adj} are taken as a proof of the high predictive ability of QSRR models. A high value of these statistical characteristic (> 0.5) is considered as a proof of the high predictive ability of the model.

2. 3. Molecular Modeling

Molecular modeling was performed by using CS Chem-Office Software version 7.0 (Cambridge) running on a P-III processor.[22] All molecules were constructed by using Chem Draw Ultra 7.0 and saved as the template structures.[23] For every compound, the template structure was suitably changed considering its structural features, copied to Chem 3D 7.0 to create a 3-D model and, finally, the model was cleaned up and subjected to energy minimization using molecular mechanics (MM2). The minimization was performed until the gradient of conformational potential (RMS) value reached a value less than 0.1 kcal/mol·A. For calculating lipophilicity parameters, the lowest energy structure for each model was used. Partition coefficients were calculated with different theoretical bases (atomic based prediction, fragment based prediction): Alog P, IAlog P, Clog P, log P_{Kowin}, Xlog P, ACDlog P by applying different theoretical procedures.[24–26] ACDlog P was calculated by using commercial software (the commercial physical property calculation software, ACD/Labs Physico-Chemical Laboratory).[27]

2. 4. Stastical Methods

PCA was carried out using Statistica v. 8 software, and HCA using NCSS 2007 and GESS software package.[28] The complete regression analysis were carried out by PASS 2005, GESS 2006, NCSS Statistical Softwares.[28] Physicochemical properties were calculated using the ChemSilico software[29] and Molinspiration online program.[30]

3. Results and Discussion

Results from regression analysis using well-known equation (2) according a procedure described earlier.[20]

The retention behavior of compounds in various chromatographic systems strongly depends on their physicochemical properties due to complex interactions in ternary system analyte/stationary/mobile phase exploiting wide array of different mechanisms. These correlations are known as QSRR.[31,32] Besides practical application in optimization strategies, QSRR studies can significantly contribute to getting some insight into the mechanism of chromatographic separation on a molecular level. The QSRR equations describe R_M^0, determined for organic components of the mobile phase and it is a function of logarithm of octanol/water partition coefficients. Linear relationships between the retention factor, R_M^0, and the standard lipophilicity parameter, log *P*, can be expected because retention of compounds in reversed phase liquid chromatography is principally governed by hydrophobic interactions. Theoretical partition coefficients (*A*log *P*, *IA*log *P*, *C*log *P*, log P_{Kowin}, *X*log *P*, *ACD*log *P*) can be calculated by software packages (Table 2).

While log *P* is substantial for the absorption of the compound, water solubility (log *S*) is another important determinant in biological activity of the substance, since biological systems are, in fact, aqueous systems. Water solubility, or simply solubility, and lipophilicity are closely related. Molecules with low solubility usually have high lipophilicity. The third very important factor is the ionization ability. pH value affects strongly the form in which ionizable compound will be (neutral or ionic). Ionization ability, expressed as pK_a, and solubility, are also related. Solubility of neutral molecules is not pH dependent, while solubility of ionizable molecules depend on pH strongly.

In order to identify the effect of molecular descriptors on the retention constants, QSRR studies of title compounds were performed. Besides the descriptors of lipophilicity (log *P*) the following molecular descriptors were calculated: dissociation constant (pK_a), water solubility (log *S*), molar refractivity (*MR*), total energy (E_t), and Gibbs energy distribution (*GibbsE*) (Table 2).

3. 1. Multivariate Statictical Analysis and Model Validation

3. 1. 1. Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA)

PCA is a very useful statistical technique for reducing the amount of data when there is correlation present, while retaining as much of the information (variation) as possible. Application of PCA to retention data can reveal some similarities among the studied compounds that are governed by both their intrinsic structural properties and specific interactions that occur in different chromatographic systems. Loading plots highlight the most influential chromatographic systems responsible for such clustering. Furthermore, a PCA carried out on the set of calculated molecular descriptors can cluster compounds based on their structural features alone. It is therefore useful to perform a PCA on both retention data and molecular descriptors separately. If a congeneric series of compounds is studied, outliers might be detected by using PCA and removed prior to the final modeling.

Table 2. Molecular descriptors calculated by different theoretical methods

Comp.	*A*log *P*	*IA*log *P*	*C*log *P*	log P_{Kowin}	*X*log *P*	*ACD*log *P*	pK_a	log *S*	*MR*	*E*t	*GibbsE*
I.1	5.250	5.060	4.850	5.070	4.830	3.870	3.019	–4.400	106.361	3.392	820.190
I.2	5.650	5.130	5.850	6.160	5.70	4.790	4.216	–4.400	116.444	3.527	817.770
I.3	5.940	7.080	6.280	6.360	6.070	5.060	3.346	–4.410	115.971	5.063	777.070
I.4	6.180	5.890	6.580	6.850	6.430	5.420	3.162	–4.650	121.607	7.715	829.570
II.1	4.960	5.150	5.320	5.880	3.710	3.70	3.45	–4.000	90.399	33.636	628.690
II.2	5.780	5.960	6.440	6.860	4.850	4.820	3.884	–4.320	99.602	26.337	621.330
II.3	6.550	6.650	7.550	7.850	5.990	5.950	2.001	–3.000	108.804	41.901	613.970
III.1	4.910	4.760	5.40	5.960	4.010	3.740	4.155	–3.330	90.326	21.101	615.470
III.2	5.550	4.810	5.480	6.860	5.140	4.820	6.840	–4.980	98.796	30.761	675.090
III.3	6.070	5.80	7.210	8.170	6.140	5.360	4.691	–4.120	113.516	30.554	799.790
III.4	6.980	7.980	9.020	10.390	8.270	7.070	–	–	136.705	39.737	984.110
IV.1	3.830	3.630	3.120	3.810	2.370	2.850	1.715	–3.510	64.708	21.498	445.790
IV.2	4.210	4.080	3.670	4.30	2.940	3.410	1.778	–3.760	69.309	17.739	442.110
IV.3	4.640	4.480	4.230	4.790	3.510	3.980	0.955	–3.520	73.911	25.401	438.430

The idea behind PCA is to find principal components PC1, PC2,...,PC_n which are linear combinations of the original variables describing each specimen, X_1, X_2,...X_n.

The first principal component (PC1) defines as much of the variation in the molecular descriptors data as possible. The second principal component (PC2) describes the maximum amount of residual variation after the first PC has been taken into consideration, and so on.[33]

HCA is a method which reveals similarities/dissimilarities between objects in the variable space, or similarities/dissimilarities in the object space.[34] Cluster hierarchy is commonly displayed as a tree diagram called a dendogram. The horizontal axis of the dendogram represent the distance or dissimilarity between clusters. The vertical axis represents the objects and clusters.

In this study PCA has been applied in order to overview the data for similarities and dissimilarities. Score values for the first two PCs are presented in Figure 1.

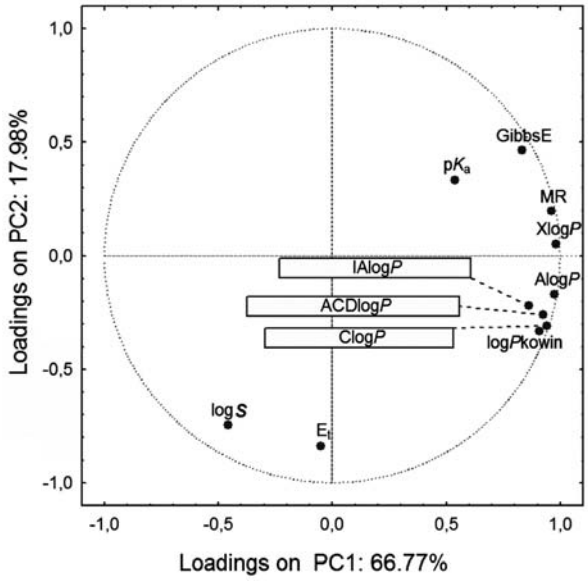


Figure 1. Factor loadings of partitions coefficients and molecular descriptors for the first two PCs

The highest positive impact to the PC1 is recorded by partition coefficients, Gibbs energy distribution, and molar refractivity. The negative influence on the score values of the PC1 has log *S*. Score plot visualize similarities/differences between investigated compounds (Figure 2).

Because of missing values for pK_a and log *S*, the compound III.4 was excluded from the PCA and HCA analysis.

The score plot reveals that compounds from fourth series are well separated from the others. Compounds of series IV, becouse of their small molecular volume significantly differ from other investigated seriies. These are monosubstituted molecules, therefore the least hydrophobic, which results in the lowest values of log *P*. Unfortu-

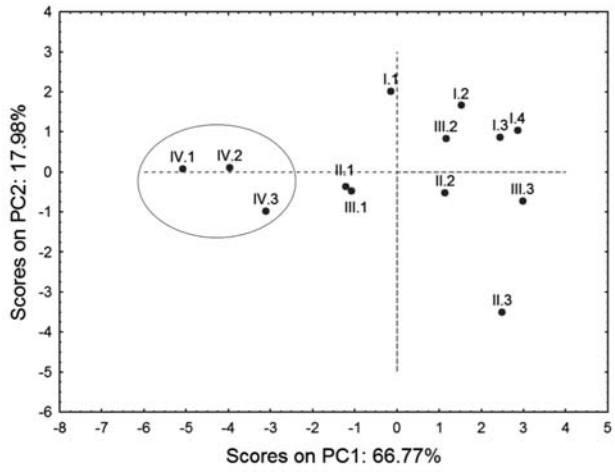


Figure 2. Score plot based on the partition coefficients and molecular descriptors of the compounds investigated

nately, scores plot does not reveal separation of compounds from I, II, and III. group. In order to determine differences between the I, II and III. group, and to confirm the diversity of the IV. group of analyzed molecules, HCA has been applied. Clustering is based on Ward's linkage method[35] and Euclidean distance. As result of HCA, dendogram is shown in the Figure 3. Obtained dendogram shows clustering of investigated compounds in three well-separated clusters: cluster A contains compounds from the I. series, cluster B groups compounds from the II. and III. series, and individual cluster C contains compounds from the IV. series. It is obviously that compounds in cluster B have similar properties: (compounds of series II and III are disubstituted molecules (high molecular volume) and have not halogen atom in structure (series I). It is well known that the presence of halogen atoms certainly affect the retention behavior. Hence, it may be concluded that by using HCA, compounds are more precisely separated than by PCA.

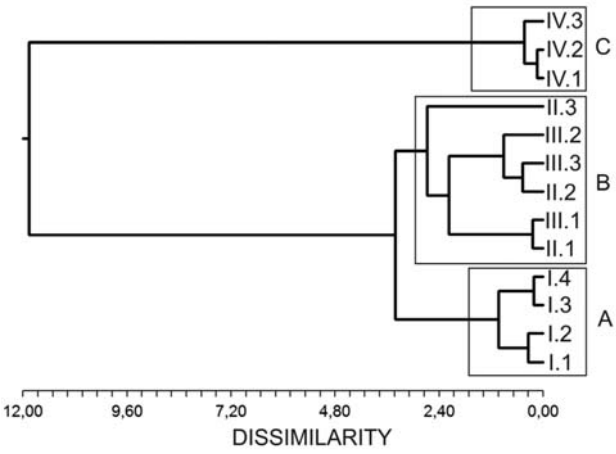


Figure 3. Dendogram of 13 compounds in the space of 11 molecular descriptors

3. 1. 2. Introduction to MLR Based on Retention Data and Molecular Descriptors

The reduction of the number of descriptors was performed before the model construction. The descriptors obtained by stepwise regression routine served as the input data for MLR analysis.

Pearson's correlation matrix has been performed on all descriptors by using NCSS Statistical Software[28] (Table 3). The correlation matrix presented that some parameters selected in the study is highly correlated with the other. However, it is very important not to derive the models containing descriptors which are highly correlated (highly correlated variables lead to unstable MLR models).

Mathematical models were formed by a stepwise addition of terms. A deletion process was then employed where each variable in the model was held out in turn and using the remaining parameters models were generated. Each descriptor was chosen as an input for the software package of NCSS and then the stepwise addition method implemented in the software was used to choose the descriptors contributing to the retention of s-triazines derivatives.

The correlation coefficients for monoparametric models were presented in Table 4. It can be concluded that the partition coefficient (log *P*) tends to correlate with retention constant exclusively. Results presented in Table 4 have confirmed expected correlation between R_M^0 and log *P*.[36, 37]

The specifications for the best-selected MLR models are shown in Table 5. MLR method only can be used when a relatively small number of molecular descriptors are used (at least five to six times smaller than the total number of compounds). In this case (for fourteen compounds), only two descriptors can be used to develop a good QSRR model in order to avoid a high chance of spurious correlations. In this approach, only the biparametric QSRR models were derived. The statistical quality of the generated models is determined by statistical measures: the square of the correlation coefficient (r^2), the standard error of estimation (*s*), and *F*-test (Fisher's value) for statistical significance.[38–40] The square of the correlation coefficient (or coefficient of multiple determination) is a relative measure of fit by the regression equation. Correspondingly, it represents the part of variation in the observed data that is explained by the regression. The correlation coefficient values closer to 1.0 represent the better fit of the regression. Standard deviation is measured by the error mean square, which expresses the variation of the residuals or the variation about the regression line. Thus, standard deviation is an absolute measure of quality of fit and should have a low value for the regression to be significant. The *F*-test reflects the ratio of the variance explained by the model and the variance due to the error in regression. High values of the *F*-test indicate that the model is statistically significant.

However, it is well known that there are three important steps in any QSRR study: development of models, validation of models and utility of developed models. Validation is a crucial aspect of any QSRR analysis.[41] The statistical quality of the resulting models, as depicted in Table 5, is determined by r^2, *s*, and *F*.[42–44] It is noteworthy that all these equations were derived using entire data set of compounds ($n = 14$) and no outliers were identified. The *F* - value presented in Table 5 is found statistically significant at 99% level since all the calculated *F* values are higher as compared to tabulated values.

Table 3. Correlation between different molecular descriptors

r	*A*log *P*	*IA*log *P*	*C*log *P*	log P_{Kowin}	*X*log *P*	*ACD*log *P*	pK_a	log *S*	*MR*	*E*t	*GibbsE*
*A*log *P*	1.000	0.893	0.966	0.918	0.962	0.969	0.402	–0.313	0.913	0.059	0.726
*IA*log *P*		1.000	0.879	0.769	0.838	0.843	0.190	–0.187	0.804	0.001	0.603
*C*log *P*			1.000	0.966	0.898	0.930	0.408	–0.176	0.852	0.199	0.655
Kowwin				1.000	0.847	0.906	0.550	–0.220	0.777	0.340	0.595
*X*log *P*					1.000	0.931	0.451	–0.445	0.968	–0.158	0.843
*ACD*log *P*						1.000	0.337	–0.255	0.829	0.133	0.601
pK_a							1.000	–0.658	0.483	0.042	0.555
log *S*								1.000	–0.506	0.454	–0.635
MR									1.000	–0.284	0.934
*E*t										1.000	–0.458
GibbsE											1.000

Table 4. Correlation coefficients (r) for the relationships between R_M^0 and molecular descriptors

Modifier		*A*log *P*	*IA*log *P*	*C*log *P*	Log P_{Kowin}	*X*log *P*	*ACD*log *P*	pK_a	log *S*	*MR*	*E*t	*GibbsE*
Acetonitrile		0.924	0.800	0.906	0.924	0.903	0.900	0.708	–0.468	0.868	0.311	0.758
Acetone		0.845	0.715	0.840	0.838	0.794	0.760	0.731	–0.399	0.809	0.270	0.719
Tetrahydrofuran	R_M^0	0.864	0.750	0.830	0.773	0.799	0.741	0.587	–0.297	0.848	0.168	0.755
Methanol		0.931	0.855	0.937	0.924	0.902	0.898	0.568	–0.295	0.875	0.325	0.755
2-Propanol		0.915	0.762	0.871	0.832	0.848	0.846	0.571	–0.325	0.838	0.236	0.680

Table 5. Statistical parameters for multiple dependence between R_M^0 and calculated descriptors

Modifier	$R_M^0 = a_1 + b_1 \text{Alog } P + c_1 \text{p}K\text{a}$						
R_M^0	a_1	b_1	c_1	r^2	F	s	eq.
acetone	–1.447	0.746	0.288	0.876	35.913	0.359	8
acetonitrile	–0.831	0.528	0.153	**0.956**	113.345	0.129	9
tetrahydrofuran	–0.979	0.686	0.134	0.823	23.294	0.337	10
2-propanol	–1.250	0.604	0.084	0.937	76.415	0.154	11
methanol	–2.986	1.052	0.163	0.874	35.156	0.403	12
	$R_M^0 = a_2 + b_2 \text{Clog } P + c_2 \text{p}K\text{a}$						
R_M^0	a_2	b_2	c_2	r^2	F	s	eq.
acetone	–0.020	0.467	0.282	0.895	42.596	0.333	13
acetonitrile	0.256	0.314	0.155	0.931	68.604	0.165	14
tetrahydrofuran	0.415	0.412	0.134	0.805	20.686	0.354	15
2-propanol	–0.027	0.363	0.084	0.918	56.249	0.178	16
methanol	–0.943	0.652	0.156	0.891	40.832	0.377	17
	$R_M^0 = a_3 + b_3 \text{ACDlog } P + c_3 \text{p}K\text{a}$						
R_M^0	a_3	b_3	c_3	r^2	F	s	eq.
acetone	–0.181	0.583	0.329	0.821	22.877	0.436	18
acetonitrile	–0.114	0.461	0.172	**0.974**	195.377	0.099	19
tetrahydrofuran	0.311	0.504	0.177	0.695	11.461	0.442	20
2-propanol	–0.379	0.513	0.108	0.927	65.296	0.166	21
methanol	–1.386	0.871	0.209	0.839	26.279	0.457	22

Table 6. Cross-validation parameters

Modifier	**eq.**	**PRESS**	**SSY**	**PRESS/SSY**	S_{PRESS}	r^2_{CV}	r^2_{adj}
	8	1.941	10.584	0.183	0.372	0.817	0.853
acetone	13	1.857	10.584	0.175	0.364	0.825	0.874
	18	3.270	10.584	0.309	0.483	0.691	0.785
	9	0.306	3.992	0.077	0.146	**0.923**	**0.949**
acetonitrile	14	0.724	3.993	0.181	0.227	0.819	0.918
	19	0.148	3.993	0.037	0.103	**0.963**	**0.970**
	10	1.894	6.428	0.295	0.368	0.705	0.788
tetrahydrofuran	15	2.056	6.428	0.320	0.383	0.680	0.766
	20	3.691	6.427	0.574	0.513	0.426	0.635
	11	0.403	3.889	0.104	0.170	0.896	0.926
2-propanol	16	0.815	3.889	0.209	0.241	0.791	0.902
	21	0.470	3.889	0.121	0.183	0.879	0.915
	12	2.448	13.059	0.187	0.418	0.813	0.851
methanol	17	2.632	13.060	0.202	0.433	0.798	0.869
	22	3.529	13.059	0.270	0.502	0.730	0.808

The developed models were validated by the calculation of following statistical parameters: PRESS, SSY, S_{PRESS}, r^2_{CV}, and r^2_{adj} (Table 6).

Cross-validation parameters of the best models (9 and 19) are obtained with acetonitrile as mobile phase showed the best validation. From both above presented models, it can be concluded that the strong influence of the lipophilicity, log *P*, is important for the retention behavior and this parameter is usually related to pharmacological activity.[45]

The only way to estimate the true predictive power of a model is to test its ability to predict accurately the retention behavior of compounds. In order to verify the predictive power of the developed model, predicted (R_M^0 values) of s-triazine derivatives investigated were calculated by using models 9 and 19 and compared with the experimental values (Table 7). The data presented in Table 7 show that the observed and the estimated activities are very close to each other. The residual activity (difference between experimentally observed R_M^0 values and QSRR calculated R_M^0 values is less than equal to 0.702.

Table 7. Retentive screening summary

Comp.	observed	I Model acetonitrile		II Model acetonitrile	
		predict	residual	predict	residual
I.1	2.110	2.403	0.293	2.247	0.137
I.2	2.786	2.797	0.011	2.746	–0.04
I.3	2.887	2.817	–0.07	2.746	–0.141
1.4	2.896	2.916	0.02	2.812	–0.084
II.1	2.388	2.316	–0.072	2.461	0.073
II.2	2.753	2.815	0.062	2.880	0.127
II.3	3.000	2.933	–0.067	2.937	–0.063
III.1	2.356	2.397	0.041	2.596	0.24
III.2	3.284	3.146	–0.138	3.037	–0.247
III.3	3.101	3.092	–0.009	3.247	0.146
III.4	3.556	2.854	–0.702	3.088	–0.468
IV.1	1.497	1.454	–0.043	1.501	0.004
IV.2	1.590	1.664	0.074	1.684	0.094
IV.3	1.956	1.765	–0.191	1.732	–0.224

Further, the plot that shows predicted R_M^0 values against the observed R_M^0 values also proves the usefulness of the derived models (Figure 4).

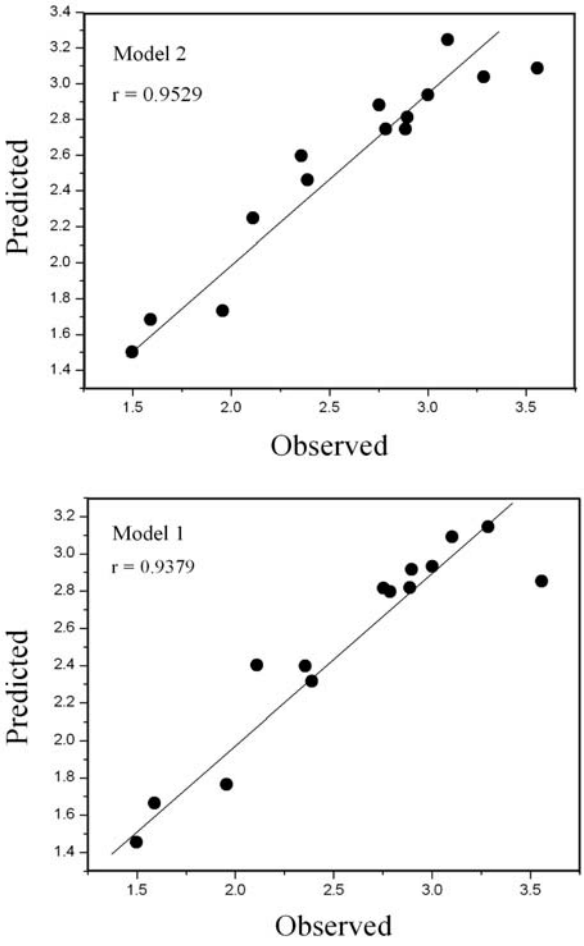


Figure 4. Plots of predicted versus experimentally observed retention parameters

In order to investigate the existence of a systemic error in developing the QSRR models, the residuals of predicted R_M^0 values were plotted against the observed R_M^0 values (Figure 5). The propagation of the residuals on both sides of the zero axis indicates that no systemic error in the development of regression models exists, as suggested by Jalali-Heravi and Kyani.[46]

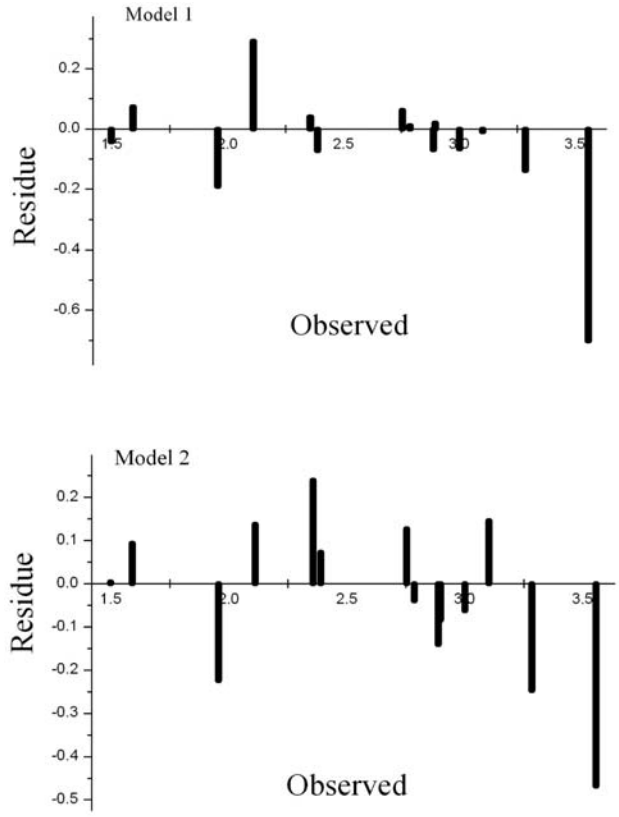


Figure 5. Plots of residual values against the experimentally observed values of retention parameters

The results of this investigation indicate that the retention constants of the tested compounds are governed by the partition coefficient, log *P*, and dissociation constant, pK_a. Lipophilicity as a pharmacokinetic descriptor has an important effect on retention behavior and this parameter is usually related to retention parameter, R_M^0.[47] To conclude, results of this study indicate that Alog *P* and ACDlog *P* are equaly suitable for prediction of retention behavior of investigated series of s-triazines.

By observing correlation of R_M^0 with different molecular descriptors, it could be concluded which chromatographic system is the best for prediction of the retention behavior of s-triazines investigated. The data presented in Tables 5 and 6 show that the thin layer of octadecil silica gel with acetonitrile-water is the best RP TLC system for prediction.

3. 2. Correlations of Retention Parameters (R_M^0) and Lipophilic Substituent Constant (π)

Lipophilic substituent constants (π), also known as hydrophobic substituent constants, represent the contribution that a certain functionality of the molecule makes to the partition coefficient. The parameter was defined by Hansch and co-workers by the equation:[48]

$$\pi = \log P_X - \log P_H \qquad 23)$$

where P_H and P_X are the partition coefficients of the parent compound and its monosubstituted derivative, respectively. The π value varies depending on the solvent system used in the determination of the partition coefficients. Most of the π values are determined using the n-octanol/water system. A positive π value indicates that a substituent has a higher lipophilicity than hydrogen, and that potentiates the distribution of the compound to the n-octanol phase. Thus, higher concentrations of such derivative are expected in the lipid material of biological systems. Negative π values show that the substituent has a lower lipophilicity than hydrogen, increasing the concentration of the compound in the aqueous media of biological systems.

When estimating the lipophilicity of compounds, lipophylic substituent constants can be used as an alternative to the partition coefficient, when studying a series of analogues in which only the substituents are different. This is based on the assumption that lipophilic effects of the unchanged parts of molecules are similar for each of the analogue. Consequently, the π values stress the effects of the substituents to the lipophilicity of the molecule. Furthermore, biological activity – π relationships that have high regression constants and low standard deviations demonstrate that the substituents are important in determining the lipophilic character of the drug. Dependence of π on R_M^0 values for compounds of series I is given in the Figure 6.

For all examined modifiers observed dependences between R_M^0 and π could be described with polynomial functions (24–28) of second order with good correlation.

$$R_M^0(\text{acetone}) = 3.187 - 0.496\pi + 2.065\pi^2 \quad r = 0.998,\ s = 0.046,\ n = 4, \qquad (24)$$

$$R_M^0(\text{acetonitrile}) = 2.109 + 1.792\pi - 1.012\pi^2 \quad r = 0.999,\ s = 0.019,\ n = 4 \qquad (25)$$

$$R_M^0(\text{methanol}) = 2.449 - 0.094\pi + 2.845\pi^2 \quad r = 0.991,\ s = 0.199,\ n = 4 \qquad (26)$$

$$R_M^0(\text{2-propanol}) = 1.901 + 1.267\pi - 0.319\pi^2 \quad r = 0.999,\ s = 0.028,\ n = 4 \qquad (27)$$

$$R_M^0(\text{tetrahydrofuran}) = 3.437 + 0.932\pi - 1.181\pi^2 \quad r = 0.967,\ s = 0.043,\ n = 4 \qquad (28)$$

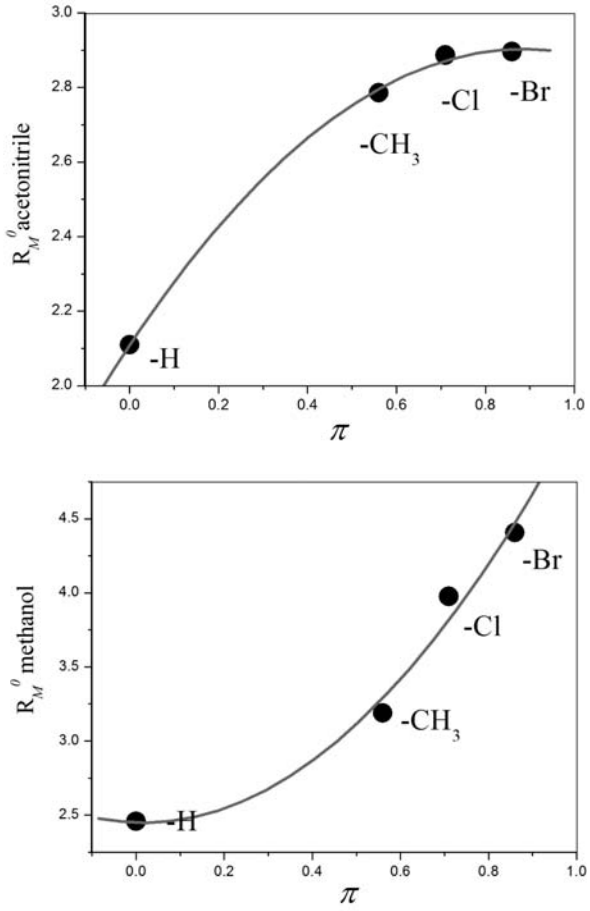


Figure 6. Correlation between R_M^0 and π for series I compounds.

4. Conclusions

Molecular modeling and QSRR analysis were performed to find the quantitative effects of the molecular structure of the compounds on their retention behavior. For all investigated derivatives, calculated retention parameters could be very good correlated with some of the molecule physicochemical properties, such as *in silico* calculated bioactivity descriptors for lipophilicity and solubility. Accurate mathematical models (MLR) were developed for predicting the retention behavior of some s-triazine derivatives. The validity of the model has been established by the determination of suitable statistical parameters. The established models were used to predict the retention of the s-triazine investigated and close agreement between experimental and predicted values was obtained. The low residual activity and high cross-validated r^2 values (r^2_{CV}) obtained suggest a good predictive ability of the developed QSRR models. It indicates that the retention constants of series of s-triazine derivatives can be successfully modeled using various molecular descriptors. It can be concluded that the strong influence of the lipophilicity, log P, and dissociation constant, pK_a are important for the retention behavior.

5. Acknowledgement

These results are part of the projects No. 114-451-2707/2012-01, financially supported by the Provincial Secretariat for Science and Technological Development of Vojvodina and projects No. 172012, No.172014 and No.31055, supported by the Ministry of Science and Technological Development of the Republic of Serbia, 2011-2014.

6. References

1. A. Solankee, K. Kapadia, A. Ćirić, M. Soković, I. Doytchinova, A. Geronikaki, *Eur. J. Med. Chem.* **2010**, *45*, 510–518.
2. H. D. Patel, H. K. Chikhalia, N. K. Shah, P. D. Patel, B. Kaswala, M. V. Buha, *J. Enzym. Inhib. Med. Ch*. **2010**, *25*, 121–125.
3. K. Srinivas, U. Srinivas, K. Bhanuprakash, K. Harakishore, H. Murthy, R. Jayathirtha, *Eur. J. Med. Chem.* **2006**, *41*, 1240–1246.
4. J. Vora, B. Vasava, D. Patel, C. Parmar, K. Chauhan, S. Sharma, *E-J. Chem*. **2009**, *6*, 201–206.
5. L. Fishbein, *Environ. Health Persp.* **1978**, *27*, 125–131.
6. B. Barr, P. Panuwet, V. Nguyen, S. Udunka, L. Needham, *Environ. Health Perspect*. **2007**, *115*, 1474–1478.
7. K. Ghosh, P. Ligy, *Global Nest J.* **2006**, *8*, 159–178.
8. J. Mooibroek, P. Gamez, *Inorg. Chim. Acta* **2007**, *360*, 381–404.
9. C. Ma, A. Dougherty, *Chem. Rev*. **1997**, *97*, 1303–1324.
10. J. Švarc-Gajić, General Toxicology. Novascience Publishers. New York, 2009.
11. T. Djaković-Sekulić, A. Smoliński, *Drug Dev. Ind. Pharm.* **2010**, *36*, 954–961.
12. N. Perišić-Janjić, R. Kaliszan, P. Wiczling, N. Milošević, G. Uscumlic, N. Banjac, *Mol. Pharmaceut.* **2011**, *8*, 555–563.
13. T. Djaković-Sekulić, A. Smoliński, N. Perišić-Janjić, M. Janicka, *J. Chemometrics* **2008**, *22*, 195–202.
14. A. Nasal, D. Siluk, R. Kaliszan, *Curr. Med. Chem*. **2003**, *10*, 381–426.
15. S. Šegan, F. Andrić, A. Radoičić, D. Opsenica, B. Šolaja, M. Zlatković, D. Milojković-Opsenica, *J. Sep. Sci*. **2011**, *34*, 2659–2667.
16. J. Thurston, J. Dudley, D. Kaiser, I. Hechenbleikner, F. Schaefer, D. Holm-Hansen, *J. Am. Chem. Soc*. **1951**, *73*, 2981–2983.
17. D. Kaiser, J. Thurston, J. Dudley, F. Schaefer, I. Hechenbleikner, D. Holm-Hansen, *J. Am. Chem. Soc*. **1951**, *73*, 2984–2986.
18. Ž. Tadić, G. Bončić-Caričić, M. Muškatirović, *J. Chem. Soc. Perkin.* **1975**, *1*, 1275–1259.
19. Ž. Tadić, K. Ries, *J. Agr. Food Chem*. **1971**, *19*, 46–49.
20. N. Perišić-Janjić, T. Đaković-Sekulić, L. Jevrić, B. Jovanović, J. *Planar Chromatogr*., **2005**, *18***,** 212–216.
21. E. Bate-Smith, R. Westall, *Biochim. Biophys. Acta* **1950**, *4*, 427–427.
22. HyperChem 7.5., Hypercube Inc., 419 Phillip St., Waterloo, Ontario, Canada, N2L 3X2, http://www.hyper.com.
23. CS. Chem. Office, Version 7.0., Cambridge Soft Corporation, 100 Cambridge Park Drive, Cambridge, MA 02140-2317, U.S.A (2001).
24. J. Chou, P. Jurs, *J. Chem. Inf. Comput. Sci*. **1979**, *19*, 172–178.
25. I. Tetko, V. Tachuk, *Virtual Computational Chemistry Laboratory*, VCC-lab **2002**, http://146.107.217.178/lab/alogps/start.html
26. Interactive Analysis log *P* Predictor, http://www.logP.com/
27. ACD/log *P*, Scientific Computing, http://www.acdlabs.com/freelogP/.
28. NCSS Statistical Software, http://www.ncss.com.
29. Chem Silico Product Secure WWW Site, http://www.chemsilico.com.
30. Molinspiration Cheminformatics, http://www.molinspiration.com.
31. N. Perišić-Janjić L. Jevrić, G. Bončić-Caričić, B. Jovanović, *J. Planar Chromatogr.* **2001**, *14*, 277–282.
32. L. Jevrić, B. Jovanović, S. Velimirović, A. Tepić, G. Koprivica, N. Mišljenović, *Hem. Ind.* **2011**, *65*, 533–540.
33. J. N. Miller, J. C. Miller, Statistics and Chemometrics for Analytical Chemistry, 6th edition, Pearson Education Limited, Harlow, UK, **2010**.
34. T. L. Đaković Sekulić, A. Smoliński, *Drug Dev. Ind. Pharm.* **2010**, *36* (8), 954–961.
35. J. H. Ward, *J. Am. Stat. Assoc.* **1963**, *58*, 236–244.
36. N. Perišić-Janjić, S. Podunavac-Kuzmanović, *J. Planar Chromatogr*. **2008,** *21*, 135–141.
37. N. Perišić-Janjić, S. Podunavac-Kuzmanović; S. Balaž, Đ. Vlaović, *J. Planar Chromatogr*. **2000,** *13,* 123–129.
38. W. Snedecor, G. Cochran, Statistical Methods (**1967**) Oxford and IBH, New Delhi.
39. S. Chaltterjee, S. Hadi, B. Price, Regression Analysis by Examples (**2000**) Wiley VCH, New York.
40. V. Diudea, QSPR/QSAR Studies for Molecular Descriptors (2000) Nova Science, Huntingdon, New York.
41. G. Topliss, P. Edwards, *J. Med. Chem.* **1979**, *22*, 1238–1244.
42. W. Snedecor, G. Cochran, Statistical Methods; Oxford and IBH, New Delhi, p. 381, **1967**.
43. S. Chaltterjee, S. Hadi, B. Price, Regression Analysis by Examples; Wiley VCH, New York, **2000**.
44. V. Diudea, QSPR/QSAR Studies for Molecular Descriptors; Nova Science, Huntingdon, New York, **2000**.
45. E J. Lien, Side Effects and Drug Design, Marcel Dekker, New York, **1987**.
46. M. Jalali-Heravi, A. Kyani, J. Chem. Inf. Comput. Sci. **2004**, *44*, 1328–1335.
47. L. Jevrić, S. Velimirović, G. Koprivica, N. Mišljenović, T. Kuljanin, A. Tepić, *Romanian Biotehnological letters.* **2012,** *17*, 6882–6892.
48. H. Könemann, R. Zelle, F. Busser, W.E. Hammers, *J of Chromatogr*. **1979,** *178,* 559–565.

Povzetek

Z QSRR (»Quantitative structure-retention relationship«) metodo smo raziskovali lastnosti, pomembne za lipofilnost štirih serij sintetiziranih derivatov s-triazina. Pri proučevanju kromatografskih lastnosti se je izkazalo, da velja linearna korelacija med vrednostmi R_M in volumskim deležem modifikatorja mobilne faze. Določili smo zvezo med retenzijskimi konstantami, R_M^0, in teoretično izračunanimi deskriptorji bioaktivnosti za lipofilnost in topnost. Najboljši QSRR modeli so bili nadalje validirani z eksperimentalnimi tehnikami in izračunom statističnih parametrov za določanje teoretičnih modelov. Dobljene vrednosti R_M^0 proučevanih derivatov s-triazina so primerne tudi za opis lipofilnosti in določanje farmakokinetičnih lastnosti.