Journal of the Serbian Chemical Society
JSCS-info@shd.org.rs • www.shd.org.rs/JSCS
J. Serb. Chem. Soc. 77 (10) 1443–1456 (2012) UDC 547.587.51:544.722.123+544.022:615.9
JSCS–4365 *Original scientific paper*

Quantitative structure–toxicity relationship study of some natural and synthetic coumarins using retention parameters

EL HADI M. A. RABTTI[1], MAJA M. NATIĆ[1#], DUŠANKA M. MILOJKOVIĆ-OPSENICA[1#], JELENA Đ. TRIFKOVIĆ[1#], TOMISLAV TOSTI[1#], IVAN M. VUČKOVIĆ[1], VLATKA VAJS[2] and ŽIVOSLAV Lj. TEŠIĆ[1#*]

[1]*Faculty of Chemistry, University of Belgrade, P. O. Box 51, 11158 Belgrade, Serbia and* [2]*Institute of Chemistry, Technology and Metallurgy, University of Belgrade, Njegoševa 12, 11000 Belgrade, Serbia*

(Received 16 July, revised 6 September 2012)

Abstract: Four lipophilicity descriptors (R_M^0, *b*, C_0 and *PC1*) for twelve coumarin derivatives were determined by reversed-phase thin-layer chromatography in order to analyze the descriptor which best describes the lipophilicity of the investigated coumarins. Moreover, possible chemical toxicity of coumarins, expressed as the probability of a compound to cause organ-specific health effects, was calculated using ACD/Tox Suite program. The quantitative relationships between toxicity and molecular descriptors, including experimentally determined lipophilicity descriptors obtained in current study were investigated using partial least square regression. The best models were obtained for kidney and liver health effects. Quantitative structure–toxicity relationship models revealed the importance of electric polarization descriptors, size descriptors and lipophilicity descriptors. The obtained models were used for the selection of the structural features of the compounds that are significantly affecting their absorption, distribution, metabolism, excretion and toxicity.

Keywords: lipophilicity parameters; thin-layer chromatography; toxicity; partial least squares regression.

INTRODUCTION

Lipophilicity of a compound is an important physico-chemical parameter. It determines biological processes as it is related to absorption, bioavailability, hydrophobic drug-receptor interactions, metabolism and toxicity. The lipophilic nature of a drug might be represented by the logarithm of the octanol–water partition coefficient, log *P*, introduced into medicinal chemistry by Hansch and Fujita.[1] Instead of the traditional shake-flask method, partition chromatographic data can be used for quantitative comparisons of relative lipophilicities. For this

* Corresponding author. E-mail: ztesic@chem.bg.ac.rs
# Serbian Chemical Society member.
doi: 10.2298/JSC120716091R

purpose, the most suitable are the intercepts of the linear relationships between the logarithm of retention constants R_M and the volume fraction of the organic modifier in a binary mobile phase obtained in reversed-phase thin-layer chromatographic (RP TLC) experiments. The relation is given by Eq. (1):

$$R_M = R_M^0 + b\varphi \quad (1)$$

where φ stands for the concentration of the organic component in the mobile phase and b is the slope, which indicates the rate at which the solubility of the solute in the mobile phase increases with changes in its composition. Parameter b is related to the specific hydrophobic surface area of the solutes in contact with the non-polar stationary phase.[2] Two theories relate the slope with the specific hydrophobic surface area. The first one correlates the slope to the number of mobile phase molecules in the solvation sphere of the solute, which are released after formation of the stationary phase–solute complex.[3] This depends on the non-polar (hydrophobic) area of a molecule in the case of reversed-phase chromatography. Another approach is based on the explanation that the surface tension of the mobile phase changes with its composition, thereby altering the energy of vacancy formation required for the accommodation of solute molecules.[2]

Besides the lipophilicity parameter R_M^0, the parameter C_0, defined as the ratio of the intercept and slope values, is frequently used in this type of investigations:[4]

$$C_0 = -R_M^0/b \quad (2)$$

C_0 could be understood as the concentration of an organic modifier in the mobile phase for which the distribution of the solute between the two phases is equal, *i.e.*, $R_M = 0$, $R_F = 0.5$. It could also be interpreted as the hydrophobicity per unit of specific hydrophobic surface area.

Several studies show that the retention is much better correlated with lipophilicity parameters if principal component analysis (PCA) is employed.[5,6] Principal components (PCs) combine all chromatographic data in one single feature, possessing in this way properties of interpolated quantities, while R_M^0, b and C_0 are extrapolated. PCA is a multivariate statistical method that is usually used to reduce the dimensionality (number of variables) of a large number of interrelated variables, while retaining as much of the information (variation) as possible. The first principal component (PC1, *i.e.*, a linear combination of the R_M values obtained under different chromatographic conditions) is chosen in the direction of the largest variance in the dataset, followed by the second one that encloses the rest of the variability and so on.[7,8]

All the above-mentioned chromatographic descriptors are equally present in the literature and are commonly used for assessing the lipophilicity of unknown solutes.

Different computational and mathematical models can be a very useful approach in prediction of biological activity of novel compounds. Quantitative structure–activity relationships (QSARs) are mathematical models that are used to correlate molecular descriptors with biological activity of a given group of compounds. Similar to QSARs, quantitative structure–retention relationships (QSRRs) relate molecular descriptors to chromatographic retention. Among the many examples of the measures that may be predicted from QSARs, modeling health effects could be considered as one of the most diverse.[9] A wide range of software tools are available for predicting physico-chemical properties and biological effects. Many of these packages are commonly used in the assessment of chemical toxicity. ToxBoxes (now called ACD/Tox Suite), marketed by ACD/ /Labs and Pharma Algorithms, provides predictions of various toxicity endpoints, including human Ether-à-go-go Related Gene (hERG) channel inhibition, genotoxicity, cytochrome P450 (CYP3A4) inhibition, Estrogen Receptor (ER) binding affinity, irritation, rodent acute lethal toxicity (LD_{50}), aquatic toxicity, and organ-specific health effects (http://www.acdlabs.com/products/admet/tox/). The predictions are associated with confidence intervals and probabilities, thereby providing a numerical expression of prediction reliability. The software incorporates the ability to identify and visualize specific structural toxicophores, giving insight into which parts of the molecule are responsible for the toxic effect.[10] Continuing research on the screening of plant extracts,[11–14] in this paper, attention is focused on naturally occurring coumarins, compounds of diverse pharmacological properties.[15] The majority of coumarins have been isolated from green plants. The genus *Seseli*, part of Apiaceae family, is a well-known source of linear or angular pyranocoumarins, an interesting subclass of coumarins possessing antiproliferative,[16] antiviral[17] and antibacterial activities.[18] Numerous species of the genus have been used in folk medicine since ancient times.

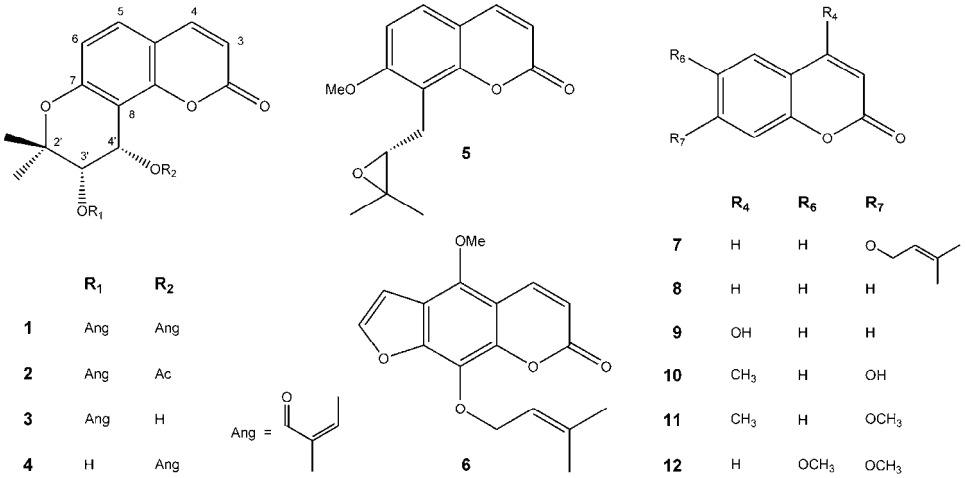
In addition to previous research on chromatographic behavior of the mentioned coumarins,[14] the first goal of this study was to determine the descriptors that best describe their lipophilicity based on thin-layer chromatographic data. The research was focused on the calculation of the probability of a compound causing organ-specific health effects; on the identification of structural features that contribute to diverse health effects; and on establishing a relationship between toxicity data and molecular descriptors, using partial least square regression, in order to determine crucial factors governing activity, *i.e.*, to reveal mechanisms of action and to propose structural features that would contribute to improved ADME-Tox profiles of the compounds.

EXPERIMENTAL

Reagents

The structures of the twelve studied coumarins are presented in Fig. 1. Coumarins **1**–**5** were isolated from *Seseli montanum* subsp. *tommasinii*.[19] Coumarin **6** was isolated from the

roots of *Seseli annuum*[20] and coumarin **7** was obtained from *Achillea tanacetifolia*.[21] Compounds **8**–**12** were purchased from Sigma–Aldrich (Steinheim, Germany). Their purity was proven by HPLC or NMR spectroscopy.



	R_1	R_2
1	Ang	Ang
2	Ang	Ac
3	Ang	H
4	H	Ang

	R_4	R_6	R_7
7	H	H	O
8	H	H	H
9	OH	H	H
10	CH_3	H	OH
11	CH_3	H	OCH_3
12	H	OCH_3	OCH_3

Fig. 1. Chemical structures of the investigated coumarins: anomalin (**1**), isopteryxin (**2**), isolaserpitin (**3**), laserpitin (**4**), meranzin (5), phellopterin (**6**), 7-*O*-prenylumbelliferone (**7**), coumarin (**8**), 4-hydroxycoumarin (**9**), 7-hydroxy-4-methylcoumarin (**10**), 7-methoxy-4-methylcoumarin (**11**) and 6,7-dimethoxycoumarin (**12**).

Thin-layer chromatography

The analytical procedure for the TLC was described in detail previously.[19]

Software

Software-predicted lipophilicity of the compounds was calculated with the available programs (http://www.vcclab.org/lab/alogps/). The ALOGPS 2.0 program package for prediction of lipophilicity and aqueous solubility of compounds was developed using the efficient partition algorithm and an associative neural network (ASNN) approach. The database used in the current program included 13,360 compounds with experimental values for lipophilicity (log *P*) covering a diverse range.[22]

The prediction of toxicity was realized using ADME/Tox WEB Software (http://pharma-algorithms.com/webboxes/). The predictions of genotoxicity by ToxBoxes are based on the probability of a query compounds to be genotoxic in the Ames test. The training data used in the software contained the results of Ames genotoxicity assays for several strains of *S. typhimurium*, with or without metabolic activation. A neural network model was built using structural fragments as descriptors. The molecules were decomposed into atomic- and chain-based fragments. Fragments containing 2 to 5 atoms present in at least 10 training set molecules were used to develop the model. The model makes a prediction if the chemical structure is more than 75 % covered by fragments in the training set.[10]

Optimized geometrical representations of the coumarins were obtained by Hyperchem Professional software (version 7.0, Hybercube). Molecular Modeling Program Plus (MMP Plus) software was employed for the calculation of the physicochemical properties (http://www.norgwyn.com/com/mmpplus.html). PCA and Partial least square regression (PLS)

were performed using PLS_Toolbox statistical package (Eigenvectors Inc. v. 5.7) from MATLAB, v. 7.4.0.287 (R2007a) (MathWorks INC, Natick, MA, USA). PCA was performed as an exploratory data analysis using a singular value decomposition algorithm (SVD) and a 0.95 confidence level for Q and T^2 Hotelling limits for outliers. The PLS method was employed using the SIMPLS algorithm without forcing orthogonal conditions to the model in order to condense Y-block variance into first latent variables. Model calibration was performed using the random samples cross-validation method. The calibration model was characterized by the root mean square errors of calibration (*RMSEC*) and root mean square errors of cross-validation (*RMSECV*). The explained variances are defined as the sum of squares due to regression divided by the sums of the squares about the mean: R^2Y (cum), the square of the multiple correlation coefficients for the calibration objects, and Q^2Y (cum), the square of the multiple correlation coefficients for the cross-validation segments.[7,8] The data were mean-centered and scaled to unit variance before statistical analyses. Autoscaling of the data was chosen as a pretreatment method in order to prevent highly abundant components from dominating components present in much smaller quantities. The values of the probabilities of health effects were the dependent variables in the quantitative structure–toxicity relationship (QSTR) equations, and they were regressed against the molecular structural descriptors (*i.e.*, independent variables).

RESULTS AND DISCUSSION

Lipophilicity of the compounds

Retention behavior of the investigated coumarins was described in detail in a previous work.[19] Reversed-phase thin-layer chromatography was performed on an octadecyl-modified silica stationary phase with three different binary solvent systems composed of water and organic modifier (methanol, tetrahydrofuran or acetonitrile). Data for linear correlation between R_M and the volume fraction of organic modifier in the mobile phase along with slopes, the correlation coefficients and standard errors of estimation were previously reported.[19] The calculated R_M^0 values were different for the individual compounds due to different substituents. The aforementioned paper also described the determination of octanol–water partition coefficients, log P_{OW}, as a measure of the lipophilicity of tested compounds. The log P_{OW} values were experimentally obtained using eight standard solutes with known log P_{OW} values, which were analyzed under the chosen chromatographic condition (methanol–water, 75/25 %, v/v), the same as for the target substances.[19] The determined lipophilicity of the investigated compounds was in accordance with their chromatographic behavior. The experimentally established R_M^0 and log P_{OW} values, obtained with methanol as the organic modifier, were correlated against log P values calculated using different software packages. It was concluded that the RPTLC retention constants, R_M^0, and the log P_{OW} values of the investigated compounds reflect their lipophilicity.

In addition to mentioned observations, correlations between R_M^0 (intercept) and b (slope) of the linear relationship between R_M and the volume percent of organic modifier in the aqueous mobile phase was performed in order to evaluate the possibility of the use of the slopes as lipophilicity parameters. These linear

relations are presented as equations given together with correlation coefficients (r), standard deviation (s), and Fisher test (F) calculated for the 95 % level of significance. Highly significant linear relationships between the retention constants R_M^0 and b were obtained (the calculated Student's t-values were greater than the critical one):

$$b_{(\mathrm{ACN})} = -1.067(\pm 0.215) - 0.758(\pm 0.012)\, R_M^0$$

$$r = 0.7890, \quad s = 1.06, \quad F = 42.106, \quad t = 4.06\, t_{cr} = 2.26$$

$$b_{(\mathrm{MeOH})} = -0.807(\pm 0.063) - 0.905(\pm 0.021)\, R_M^0$$

$$r = 0.9940, \quad s = 0.08, \quad F = 1772.254, \quad t = 28.74\, t_{cr} = 2.26$$

$$b_{(\mathrm{THF})} = -0.900(\pm 0.273) - 1.123(\pm 0.015)\, R_M^0$$

$$r = 0.8950, \quad s = 0.55, \quad F = 94.652, \quad t = 6.34\, t_{cr} = 2.26$$

The obtained statistically significant relationships indicate that the slopes could be considered as an alternative to the intercepts lipophilicity parameters. The slopes were further compared with the previously calculated log P values[19] and the statistical parameters of these dependences are given in Table I.

TABLE I. Relationships between the slopes and the log P values determined using different computational programs

b *vs.* log P	Organic modifier	Equation	r	s
b – *A*log *Ps*	Acetonitrile	b = –0.618(0.265) – 0.613(0.090)*A*log *Ps*	0.806	0.972
	Methanol	b = –0.341(0.275) – 1.033(0.093)*A*log *Ps*	0.918	1.046
	Tetrahydrofuran	b = 1.696(0.234) – 0.641(0.079)*A*log *Ps*	0.854	0.759
b – *AC*log *P*	Acetonitrile	b = –0.5745(0.252) – 0.647(0.088)*AC*log *P*	0.829	0.859
	Methanol	b = –0.276(0.245) – 1.087(0.085)*AC*log *P*	0.936	0.810
	Tetrahydrofuran	b = –1.597(0.158) – 0.696(0.055)*AC*log *P*	0.935	0.338
b – *A*log *P*	Acetonitrile	b = –0.539(0.267) – 0.634(0.090)*A*log *P*	0.816	0.923
	Methanol	b = –0.217(0.279) – 1.064(0.094)*A*log *P*	0.921	1.004
	Tetrahydrofuran	b = –1.566(0.187) – 0.679(0.063)*A*log *P*	0.914	0.450
b – *M*log *P*	Acetonitrile	b = 0.390(0.775) – 1.168(0.328)*M*log *P*	0.515	2.435
	Methanol	b = 1.690(0.979) – 2.111(0.414)*M*log *P*	0.694	3.886
	Tetrahydrofuran	b = –0.306(0.611) – 1.365(0.259)*M*log *P*	0.710	1.512
b – log P_{KOWWIN}	Acetonitrile	b = –1.045(0.231) – 0.496(0.081)log P_{KOWWIN}	0.767	1.167
	Methanol	b = –1.039(0.249) – 0.844(0.088)log P_{KOWWIN}	0.893	1.354
	Tetrahydrofuran	b = –2.086(0.160) – 0.541(0.056)log P_{KOWWIN}	0.892	0.562
b – *X*log *P2*	Acetonitrile	b = –0.492(0.382) – 0.739(0.115)*X*log *P2*	0.783	1.088
	Methanol	b = –0.139(0.351) – 1.240(0.134)*X*log *P2*	0.884	1.470
	Tetrahydrofuran	b = –1.546(0.258) – 0.779(0.098)*X*log *P2*	0.847	0.796
b – *X*log *P3*	Acetonitrile	b = –0.669(0.338) – 0.642(0.124) *X*log *P3*	0.701	1.502
	Methanol	b = –0.358(0.389) – 1.108(0.143)*X*log *P3*	0.843	1.980
	Tetrahydrofuran	b = –1.664(0.261) – 0.705(0.096)*X*log *P3*	0.828	0.896

The lower *r* values obtained for the analyzed correlations indicate that it is necessary to use a proper statistical test to see whether the correlation coefficients are indeed significant, bearing in mind the number of points used in the cali-bration. The *t*-test for the correlation confirmed that no linear relationship between *S* and *M*log *P* was obtained with the chromatographic system using acetonitrile as the organic modifier (t = 1.90, $t_{cr(0.05;10)}$ = 2.23). In all other cases, the correlations were statistically significant (t = 3.05 – 8.41). Although the results indicate that the slopes of RPTLC equations may be applied for lipophilicity expression of the investigated compounds, it could be noticed that intercepts, as a parameter of lipophilicity, are more reliable according to the Pearson's coefficients and standard errors of estimation. Within the observed correlations, the best results were obtained for methanol, *i.e.*, the correlation coefficients are the highest and deviations from the ideal correlation (slope ≈ 1 and intercept ≈ 0) are less pronounced than in the case of the other two organic modifier. In addition, the best correlations were achieved between the slopes and the *AC*log *P* values.

Lower quality correlations were obtained between R_M^0 and parameter C_0:

$$C_{0(\mathrm{ACN})} = 0.324(\pm 0.084) + 0.215(\pm 0.046) R_M^0$$

$$r = 0.6570, \quad s = 0.16, \quad F = 22.068, \quad t = 2.75\, t_{cr} = 2.26$$

$$C_{0(\mathrm{MeOH})} = 0.497(\pm 0.055) + 0.109(\pm 0.019) R_M^0$$

$$r = 0.7480, \quad s = 0.06, \quad F = 33.615, \quad t = 3.56\, t_{cr} = 2.26$$

$$C_{0(THF)} = 0.458(\pm 0.066) + 0.085(\pm 0.028) R_M^0$$

$$r = 0.4270, \quad s = 0.03, \quad F = 9.204, \quad t = 1.49\, t_{cr} = 2.26$$

and C_0 was not further considered as a potential parameter of lipophilicity.

PCA was performed on the set of retention data (R_M values obtained for the three chromatographic systems) in order to reveal possible similarities among the studied compounds governed by both their intrinsic structural properties and specific interactions that occurred in the different chromatographic systems, and to obtain the values of *PC1*, as a measure of lipophilicity. PCA applied on the entire set of molecular descriptors resulted in a three-component model explaining 97.64 % of the data variation (first principal component comprised 91.76 % of the variances).

The scores plot of the first two principal components (Fig. 2) indicates that there were no outliers among the analytes (all the data lie inside the Hotelling T^2 ellipse). Samples are clustered into two main separate groups, similar to PCA analysis of molecular descriptors of investigated substances, reported previously.[19] Clustering was performed according to the lipophilicity of the coumarins. PC1 distinguished samples consistent to the number of the rings present in the molecule (bicyclic and tricyclic compounds). The exception was compound **7**

with the hydrophobic side-chain substituent, 3-methylbut-2-enyloxy, which exhibits positive *PC1* score values together with the tricyclic compounds.

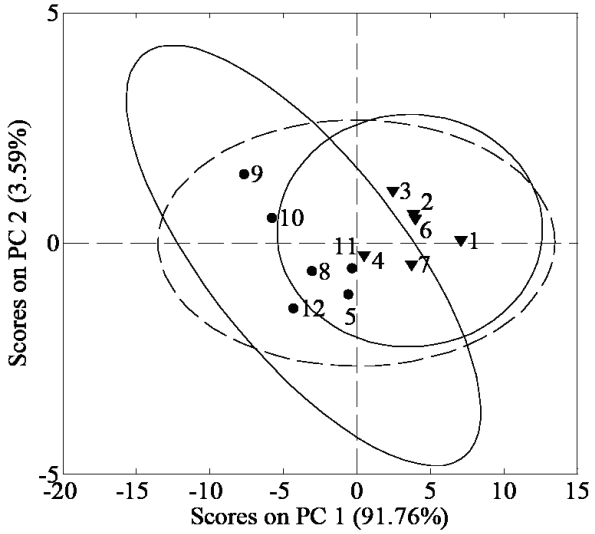


Fig. 2. Score values of the first and second PCs.

Highly significant linear relationships between the retention constants R_M^0 and *PC1* were obtained:

$$PC1 = -8.434(\pm 0.939) + 5.099(\pm 0.511) R^0_{M(ACN)}$$

$$r = 0.9000, \quad s = 20.24, \quad F = 99.715, \quad t = 6.53\, t_{cr} = 2.62$$

$$PC1 = -9.445(\pm 0.615) + 3.556(\pm 0.211) R^0_{M(MeOH)}$$

$$r = 0.9620, \quad s = 7.58, \quad F = 282.967, \quad t = 11.14\, t_{cr} = 2.62$$

$$PC1 = -16.461(\pm 1.188) + 7.181(\pm 0.502) R^0_{M(THF)}$$

$$r = 0.9490, \quad s = 10.35, \quad F = 204.468, \quad t = 9.52\, t_{cr} = 2.62$$

Taking into account the satisfactory quality of the obtained relationships, the values of *PC1* were correlated with the calculated log *P* values. The statistical parameters of these dependences are listed in Table II. Statistically significant correlations were obtained in all the investigated linear dependences, indicating that *PC1* could be used as a parameter of lipophilicity for the investigated coumarins.

A general remark related to the previous discussion could be that the variables describing directly the partitioning of the solute between the stationary and mobile phase, such as R_M^0, are more suitable for lipophilicity estimation of the investigated coumarins than the parameter proportional to the molecular hyd-

rophobic surface area (*b*) or the interpolated quantity that combines all chromatographic data (*PC1*).

TABLE II. Relationships between the *PC1* and the log *P* values determined using different computational programs

PC1 vs. log *P*	Equation	*r*	*s*
PC1 – *A*log *Ps*	*PC1* = –11.136(1.049) + 4.008(0.477)*A*log *Ps*	0.864	27.527
PC1 – *AC*log *P*	*PC1* = –11.434(1.285) + 4.235(0.448)*AC*log *P*	0.889	22.367
PC1 – *A*log *P*	*PC1* = –11.682(1.380) + 4.155(0.464)*A*log *P*	0.878	24.599
PC1 – *M*log *P*	*PC1* = –19.948(3.686) + 8.592(1.561)*M*log *P*	0.727	55.093
PC1 – log P_{KOWWIN}	*PC1* = –8.758(0.879) + 3.407(0.309)log P_{KOWWIN}	0.916	16.902
PC1 – *X*log *P2*	*PC1* =–11.755(1.808) + 4.746(0.693)*X*log *P2*	0.807	38.995
PC1 – *X*log *P3*	*PC1* =–11.455(1.485) + 4.451(0.545)*X*log *P3*	0.857	28.926

Quantitative structure–toxicity relationship

In a previous study, PLS modeling was performed in order to qualify relationships between the factors governing the lipophilicity of the studied coumarins.[19] The two proposed PLS models (the dependent variables were R_M^0 and log P_{OW}) were statistically significant and their statistical quality was comparable. From these models, it could be seen that the descriptors that describe the size and the shape of the molecule as well as their polar properties determined the lipophilic behavior of the investigated compounds.

Continuing this previous investigation, the toxicity of the analyzed substances, expressed as organ-specific health effects, was predicted and correlated with the molecular descriptors and retention parameters for all three chromatographic systems, as parameters of lipophilicity.

The numerical expressions of the prediction reliability for different health effects (blood, cardiovascular system, gastrointestinal system, kidney, liver and lungs) for the twelve investigated coumarins are given in Table III. Compounds **1**–**4** have pronounced toxic effects on blood, the cardiovascular system, the gastrointestinal system, and kidney. Compound **6** has a considerable impact on all the investigated health effects and among the observed analogs, it is the most active one. Compounds **11** and **12** have the largest influence on the gastrointestinal system.

The structural features contributing to the diverse health effects are presented in Figs. S1–S12 (Supplementary material) for all the analyzed substances. The mentioned structural features are identified on the molecules with highlighting and color mapping (red – associated with toxic action, green – unrelated to the health effects under investigation).

In order to qualify the relationships between the factors governing the toxicity of the studied compounds, PLS modeling was performed on the data of the probabilities of health effects. The number of latent variables (Num. LVs) was

selected based on minimum *RMSECV* and the minimum difference between *RMSEC* and *RMSECV*. The obtained models are summarized in Table IV.

TABLE III. The values of the probabilities of health effects for the investigated coumarins

Compd.	Probability of health effects					
	Blood	Cardiovascular system	Gastrointestinal system	Kidney	Liver	Lungs
1	0.94	0.98	0.90	0.73	0.46	0.44
2	0.82	0.97	0.87	0.54	0.60	0.41
3	0.89	0.95	0.91	0.65	0.75	0.40
4	0.86	0.94	0.90	0.67	0.44	0.40
5	0.93	0.54	0.49	0.17	0.33	0.33
6	0.90	0.94	0.72	0.86	0.73	0.70
7	0.34	0.44	0.41	0.13	0.07	0.35
8	0.27	0.29	0.23	0.09	0.09	0.16
9	0.45	0.11	0.18	0.04	0.02	0.32
10	0.23	0.04	0.28	0.07	0.03	0.19
11	0.33	0.68	0.87	0.10	0.17	0.12
12	0.60	0.77	0.78	0.12	0.19	0.13

TABLE IV. The statistical parameters of the derived PLS models

Parameter	Probability of health effects					
	Blood	Cardiovascular system	Gastrointestinal system	Kidney	Liver	Lungs
R^2Y	0.738	0.867	0.798	0.823	0.959	0.899
Q^2Y	0.508	0.446	0.276	0.638	0.634	0.124
RMSEC	0.141	0.122	0.125	0.125	0.052	0.050
RMSECV	0.191	0.260	0.254	0.182	0.158	0.169
Num. LVs	2	3	3	2	4	4

The contribution of the molecular descriptors and lipophilicity parameters that exhibit the strongest influence on toxic activity was analyzed using variable importance in projection scores (*VIP*). The variables with *VIP* scores higher than 1 were considered as the most relevant for the explanation of the dependant variable *Y*, while those significantly lower than 1 (arbitrarily a value lower than 0.5 is taken) have extremely low or almost no contribution. The descriptors included in the final models are presented in Table V in descending order of their coefficient values in regression graphs together with the notification of the sign of their contribution to the dependent variable.

The statistical parameters calculated for the models obtained after elimination of the variables that only contribute to noise (variables with low values of coefficients and low *VIP* values) confirmed that only in the case of the gastrointestinal system and kidney toxic activity was a simpler and better model obtained (gastrointestinal system – R^2Y = 0.792; Q^2Y = 0.508; *RMSEC* = 0.126; *RMSECV* = 0.199; *Num. LVs* – 3; kidney – R^2Y = 0.914; Q^2Y = 0.738; *RMSEC* = = 0.088; *RMSECV* = 0.154; *Num. LVs* – 2). Taking into account the parameters

that represent the quality of the model, it could be concluded that the PLS models for kidney and liver health effects are statistically significant.

TABLE V. Molecular descriptors included in the PLS models

Probability of health effects	Molecular descriptors
Blood	LUMO (–), Molecular width (+), Polar surface area (+), Molecular depth (+), Total energy (–), Molecular weight (+), Mass (+), Volume (+), Binding energy (–), Refractivity (+), Polarizability (+), Parachor (+), MR (+), Surface area (+)
Cardiovascular system	Hansen dispersion (–), LUMO (–), HOMO (+), $R_M{}^0{}_{(ACN)}$ (–), Total energy (–), Molecular weight (+), Mass (+), $R_M{}^0{}_{(MeOH)}$ (–), Molecular depth (–), Surface area (–), Refractivity (+), Polarizability (+), Parachor (+), MR (+), Binding energy (–), $R_M{}^0{}_{(THF)}$ (–), Volume (+)
Gastrointestinal system	HOMO (+), Hansen dispersion (–), $R_M{}^0{}_{(ACN)}$ (–), Polar surface area (+), H bond acceptor (+), Total energy (–), Molecular depth (–), Surface area (–)
Kidney	LUMO (–), Molecular width (+), Hydrophilic surface area (+), Polar surface area (+), $R_M{}^0{}_{(THF)}$ (+), Refractivity (+), Mass (+), Molecular weight (+), Volume (+), Surface area (+), MR(+), Parachor (+), $R_M{}^0{}_{(MeOH)}$ (+), Total energy (–), Binding energy (+), $R_M{}^0{}_{(ACN)}$ (+)
Liver	LUMO (–), Molecular width (+),
Lungs	LUMO (–), Molecular width (+), $R_M{}^0{}_{(THF)}$ (+)

The most relevant descriptors influencing the probabilities of health effects are electric polarization descriptors, size descriptors and lipophilicity descriptors. All the obtained models indicate the importance of the LUMO parameter with a negative contribution to toxicity. The mentioned descriptor is related to the electron affinity and is a measure of the electrophilicity of a molecule. Other electric polarization descriptors that encode information about the charge distribution in the molecule, such as polarizability and refractivity index, have a positive influence on the values of the biological activities, while the Hansen dispersion exhibits a negative influence. Descriptors related to the size of the molecule, such as molecular weight, depth, width, mass and volume, have a positive impact on all the observed health effects. The polar surface area, present in the model for blood and the gastrointestinal system, is defined as the part of the surface area of the molecule associated with oxygens, nitrogens, sulfurs and the hydrogens bonded to any of these atoms. This surface descriptor, which is related to the hydrogen-bonding ability of the compounds, has a positive impact on two mentioned health effects. On the contrary, the surface area of a substance, as the sum of all areas that cover the surface of the molecule, have different influences subject to the determined toxicity. Experimentally obtained lipophilicity parameters $R_M{}^0{}_{(ACN)}$, $R_M{}^0{}_{(MeOH)}$ and $R_M{}^0{}_{(THF)}$ are present with negative influences in the models for

the cardiovascular and gastrointestinal systems, and in models for kidney and lungs with a positive impact. Lipophilicity parameters are not included in the final models for blood and liver. This leads to the assumption that the toxicity of the investigated coumarins in these cases is probably not determined by their lipophilicity, but by specific interactions with the receptor active center. Similar results could be found elsewhere in the literature.[23,24]

CONCLUSIONS

The present work focused on identifying the most important descriptors affecting the lipophilicity of twelve coumarin derivatives. Four commonly used descriptors for assessing the lipophilicity of unknown solutes, obtained from the thin-layer chromatographic data, were compared. As a general remark, it could be stated that a variable describing directly the solute partitioning between the stationary and mobile phase, such as $R_M{}^0$, are more suitable for lipophilicity estimation of the investigated coumarins than the parameter proportional to the molecular hydrophobic surface area (*b*), or the interpolated quantity that combined all chromatographic data (*PC1*).

The toxicity of the coumarins was used for establishing QSTRs including calculated and experimentally determined molecular descriptors, and also partial least square regression. Taking into account the parameters that represent the quality of the QSTR model, it could be concluded that the best models were obtained for kidney and liver health effects. Descriptors included in the final equations were electric polarization descriptors, size descriptors and lipophilicity descriptors. The obtained models were used for the selection of the structural features of the compounds that significantly affect their absorption, distribution, metabolism, excretion and toxicity.

SUPPLEMENTARY MATERIAL

Structural features of compounds **1**–**12** contributing to diverse health effects are available electronically from http://www.shd.org.rs/JSCS/, or from the corresponding author on request.

Acknowledgements. This work was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, Grant No. 172017.

ИЗВОД

АНАЛИЗА ЗАВИСНОСТИ СТРУКТУРЕ И ТОКСИЧНОСТИ НЕКИХ ПРИРОДНИХ И СИНТЕТИЧКИХ КУМАРИНА КОРИШЋЕЊЕМ РЕТЕНЦИОНИХ ПАРАМЕТАРА

EL HADI M. A. RABTTI[1], МАЈА М. НАТИЋ[1], ДУШАНКА М. МИЛОЈКОВИЋ-ОПСЕНИЦА[1], ЈЕЛЕНА Ђ. ТРИФКОВИЋ[1], ТОМИСЛАВ ТОСТИ[1], ИВАН М. ВУЧКОВИЋ[1], ВЛАТКА ВАЈС[2] и ЖИВОСЛАВ Љ. ТЕШИЋ[1]

[1]Хемијски факултет Универзитета у Београду, П. фах 51, 11158 Београд и [2]Институт за хемију, технологију и металургију Универзитета у Београду, Његошева 12, 11000 Београд

Применом реверзно-фазне танкослојне хроматографије на дванаест деривата кумарина одређена су четири параметара липофилности ($R_M{}^0$, *b*, C_0 и *PC1*). Корелацијом добијених резултата са израчунатим log *P* вредностима утврђен је дескриптор који на

најбољи начин описује липофилност испитиваних кумарина. Поред тога израчуната је могућа хемијска токсичност кумарина, изражена као вероватноћа утицаја поменутих једињења на специфичне органе (крв, кардиоваскуларни систем, гастроинтестинални систем, бубреге, јетру и плућа) а која је израчуната применом ACD/Tox Suite програма. Добијене вредности токсичности корелисане су са молекулским дескрипторима и експериментално одређеним параметрима липофилности, применом методе парцијалне регресије најмањих квадрата (*partial least square regression*). Узимајући у обзир параметре који описују квалитет модела зависности структуре и токсичности, утврђено је да су најбољи модели добијени за утицај кумарина на бубреге и јетру. Сви добијени модели указују на значај електрично поларизационих дескриптора, као и дескриптора који описују величину и липофилност једињења, а употребљени су за утврђивање структурних карактеристика које значајно утичу на њихову апсорпцију, дистрибуцију, метаболизам, излучивање и токсичност.

(Примљено 16. јула, ревидирано 6. септембра 2012)

REFERENCES

1. C. Hansch, T. Fujita, *J. Am. Chem. Soc.* **86** (1964) 1616
2. C. Horváth, W. Melander, I. Molnar, *J. Chromatogr.* **125** (1976) 129
3. F. Murakami, *J. Chromatogr.* **178** (1979) 393
4. M. L. Bieganowska, A. Doraczynska-Szopa, A. Petruczynik, *J. Planar Chromatogr. - Mod. TLC* **8** (1995) 122
5. C. Sarbu, T. Sorina, *J. Chromatogr., A* **822 (**1998) 263
6. C. Sarbu, K. Kuhajda, S. Kevresan, *J. Chromatogr., A* **917** (2001) 361
7. R. G. Brereton, *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*, Wiley, Chichester, UK, 2003, p. 297–323
8. K. Varmuza, P. Filzmaser, *Introduction to Multivariate Statistical Analysis in Chemometrics*, CRC Press, Taylor & Francis Group, Boca Raton, FL, 2008
9. T. W. Schultz, M. T. D. Cronin, T. I. Netzeva, *J. Mol. Struct. THEOCHEM* **622** (2003) 23
10. A. Worth, S. Lapenna, E. Lo Piparo, A. Mostrag-Szlichtyng, R. Serafimova, *JRC Scientific and Technical Reports*, EUR 24705 EN, Publications Office of the European Union, European Union, Luxembourg, 2011
11. M. Sajewicz, D. Staszek, M. Natic, M. Waksmundzka-Hajnos, T. Kowalska *J. Chromatogr. Sci.* **49** (2011) 560
12. M. Sajewicz, L. Wojtal, M. Natic, D. Staszek, M. Waksmundzka-Hajnos, T. Kowalska, *J. Liq. Chromatogr. Related Technol.* **34** (2011) 848
13. M. Sajewicz, D. Staszek, M. Natic, L. Wojtal, M. Waksmundzka-Hajnos, T. Kowalska, *J. Liq. Chromatogr. Related Technol.* **34** (2011) 864
14. A. B. Abubaker, M. Natic, T. B. Tosti, D. M. Milojkovic-Opsenica, I. Z. Djordjevic, V. V. Tesevic, M. B. Jadranin, S. M. Milosavljevic, M. J. Lazic, S. S. Radulovic, Z. Lj. Tesic, *Biomed. Chromatogr.* **23** (2009) 250
15. I. Rostova, *Curr. Med. Chem. - Anti-Cancer Agents* **5** (2005) 29
16. P. Magiatis, E. Melliou, A. L. Skaltsounis, S. Mitaku, S. Léonce, P. Renard, A. Pierré, G. Atassi, *J. Nat. Prod.* **61** (1998) 982
17. L. Xie, Y. Takeuchi, L. M. Cosetino, K. H. Lee, *J. Med. Chem.* **42** (1999) 2662
18. E. Melliou, P. Magiatis, S. Mitaku, A. L. Skaltsounis, E. Chinou, I. Chinou, *J. Nat. Prod.* **68** (2005) 78

19. E. H. M. A. Rabtti, M. M. Natić, D. M. Milojković-Opsenica, J. Đ. Trifković, I. M. Vučković, V. E. Vajs, Ž. Lj. Tešić, *J. Braz. Chem. Soc.* **23 (**2012) 522
20. I. Vučković, V. Vajs, M. Stanković, V. Tešević, S. Milosavljević, *Chem. Biodiv.* **7** (2010) 698
21. S. Trifunović, *PhD. Thesis*, Comparative investigation of the chemical composition of selected *Achilea* species, Faculty of Chemistry, University of Belgrade, Belgrade, Serbia 2006.
22. I. Tetko, V. Yu. Tanchuk, *J. Chem. Inf. Comput. Sci*. **42** (2002) 1136
23. S. Šegan, F. Andrić, A. Radoičić, D. Opsenica, B. Šolaja, M. Zlatović, D. Milojković–Opsenica, *J. Sep. Sci.* **34** (2011) 2659
24. S. Šegan, J. Trifković, T. Verbić, D. Opsenica, M. Zlatović, J. Burnett, B. Šolaja, D. Milojković-Opsenica, *J. Pharm. Biomed. Anal.*, 2012, In press, DOI: 10.1016/ /j.jpba.2012.08.025.