ACADEMIA ROMÂNĂ
Revue Roumaine de Chimie
http://web.icf.ro/rrch/

Rev. Roum. Chim.,
2011, *56*(3), 217-229

MODELING OF CHROMATOGRAPHIC LIPOPHILICITY INDICES OF BILE ACIDS AND THEIR DERIVATIVES

Cristina ONIŞOR,[a] Mihalj POŠA[b] and Costel SÂRBU[a]*

[a]Babeş-Bolyai University, Faculty of Chemistry and Chemical Engineering, Arany János 11, 400028 Cluj-Napoca, Roumania
[b]University of Novi Sad, Department of Pharmacy, Faculty of Medicine, Hajduk Vejkova 3, 21000 Novi Sad, Serbia

Received August 4, 2010

Lipophilic indices for 27 structurally diverse bile acids and their derivatives with distinct functional groups were determined by reversed-phase high-performance liquid chromatography (RP-HPLC) on C18 (LiChroCART, LiChrosphere RP-18e), C8 (Zorbax, Eclipse XDB-C8) and CN (SAULENTECHNIK KNAUER LiChrosphere 100CN) columns. The highest log k_w values were obtained on C8 and C18 columns compared to CN column. Highly significant correlations were obtained between different experimental indices of lipophilicity and computed log P values and C8 column seems to be more suited for estimating the lipophilicity of the investigated compounds. In addition, the results obtained in this study applying PCA may be used in interpreting the molecular mechanism of interactions between eluents and columns with different polarities and to explain the chromatographic behavior of compounds. The contribution of 2D and 3D descriptors which are related to atomic mass and volumes, together with reactivity parameters such as polarizability and electronegativity seem to control the chromatographic mechanism (lipophilicity) on all columns.

INTRODUCTION

The complex chemistry and physiology of bile acids have fascinated a wide range of experimentalists for centuries.[1-3] The pharmacokinetic behavior of drugs and the access to their target sites are strongly dependent on passive diffusion and on interactions with biological membranes. During the last decades the n-octanol–water partition coefficient (log P or log D) was recognized as the standard lipophilicity parameter to simulate drug partitioning in membranes and has been widely used in the field of quantitative structure-activity relationships (QSAR) and drug design.[4-6]

Quantitative-structure property-relationships (QSPR) models provide insight into aspects of the molecular structure that influence the target properties. This statistical technique is often used to replace expensive biological tests or experiments of a given physicochemical property with calculated descriptors, and can also be used to predict responses of interest for new compounds.[7]

Since the importance of lipophilicity for drug design and QSAR/QSPR is recognized, the methods for log P calculation were extensively developed and improved during the past, thus, anybody can freely access via the Internet some the available tools.

Reversed-phase chromatographic indices, log k_w as alternative measure, are usually calibrated towards the octanol–water system and conditions are chosen so that the best parallelism with log P is achieved.[8, 9]

The definition of precise metabolic and physico-chemical roles for different molecular species of bile acids have been hampered by the difficulty involved in separating and quantifying the configurationally similar bile acid species. The development of sophisticated HPLC technology has provided new possibilities for rapid and precise bile acid separation and measurement.[10, 11]

The aim of this study was to determine the lipophilicity indices of 27 bile acids and their derivatives and to investigate the molecular mechanism of retention in order to find an objective manner of quantitative comparison of

* Corresponding author: Tel. +0040-264-593877, Fax: +0040-264-590818, E-mail: csarbu@chem.ubbcluj.ro

retention properties of different chemically bonded phases used in reversed-phase high performance liquid chromatography. In addition, oxo- and diacetoxy- derivatives of bile acids were included in the examination mainly because their chromatographic behavior was not yet investigated and their pharmacological use highly increases.

EXPERIMENTAL

1. Solvents and chemicals

Bile acids and their derivatives (Table 1) were delivered by the Department of Pharmacy, Faculty of Medicine, University of Novi Sad (Serbia). The compounds (1-15) are Sigma, New Zealand, 98%. Cholic (4), deoxycholic (2), chenodeoxycholic (3) and hyodeoxycholic acids (7) were used as starting compounds for the synthesis of the oxo derivatives (16-27). The synthesis of oxo derivatives was carried out according to known procedures.[12]

2. Instrumentation

The chromatography was performed on an Agilent 1100 Series LC system consisting of a vacuum degassing unit, a binary high pressure pump, a standard automatic sample injector, a column thermostat and a diode array detector (DAD). The system was connected to an 1100 MSD mass spectrometer. The chromatographic behavior of the compounds was studied on C18 (LiChroCART, LiChrosphere RP-18e, 125x4 mm, 5 μm), C8 (Zorbax, Eclipse XDB-C8, 150x4.6 mm, 5 μm) and CN (SAULENTECHNIK KNAUER LiChrosphere 100CN, 250x4 mm, 5 μm) columns.

Table 1

Abbreviations of the bile acids and their conjugates

Compounds	Abbreviations	Position and orientation of hydroxyls, diacetoxy- and oxo- groups
1. Lithocholic acid	LC	3α
2. Deoxycholic acid	DC	3α,12α
3. Chenodeoxycholic acid	CDC	3α,7α
4. Cholic acid	C	3α,7α,12α
5. Ursodeoxycholic acid	UDC	3α,7β
6. Hyocholic acid	HC	3α,6α,7α
7. Hyodeoxycholic acid	HDC	3α,6α
8. Glycochenodeoxycholic acid sodium salt	GCDC	Glyco conjugate of CDC
9. Taurodeoxycholic acid sodium salt	TDC	Tauro conjugate of DC
10. Glycocholic acid sodium salt	GC	Glyco conjugate of C
11. Glycodeoxycholic acid sodium salt	GDC	Glyco conjugate of DC
12. Taurolithocholic acid sodium salt	TLC	Tauro conjugate of LC
13. Taurochenodeoxycholic acid sodium salt	TCDC	Tauro conjugate of CDC
14. Glycolithocholic acid sodium salt	GLC	Glyco conjugate of LC
15. Taurocholic acid sodium salt	TC	Tauro conjugate of C
16. 3α,7α-Dihydroxy-12-oxo-5β-cholanic acid	12-oxo C	3α,7α,12-oxo
17. 3α,12α-Dihydroxy-7-oxo-5β-cholanic acid	7-oxo C	3α,12α,7-oxo
18. 3α-Hydroxy-7,12-dioxo-5β-cholanic acid	7,12-dioxo C	3α,7,12-dioxo
19. 12α-Hydroxy-3,7-dioxo-5β-cholanic acid	3,7-dioxo C	12α,3,7-dioxo
20. 3,7,12-Trioxo-5β-cholanic acid	3,7,12-trioxo C	3,7,12-trioxo
21. 3α-Hydroxy-12-keto-5β-cholanoic acid	3-OH,12-oxo C	3α,12-oxo
22. 3,12-Diketo-5β-cholanoic acid	3,12-dioxo C	3,12-dioxo
23.Methyl ester of 3α-Acetoxy-12-keto-5β-cholanoic acid	Methyl ester of 3-acetoxy,12-oxo C	3α-OAc,12-oxo
24. 3α-Hydroxy-7-keto-5β-cholanoic acid	3-OH,7-oxo C	3α,7-oxo
25. Methyl ester of 3α,7α-Diacetoxy-12-keto-5β-cholanoic acid	Methyl ester of 3α,7α-diacetoxy,12-oxo C	3α,7α-OAc,12-oxo
26. Methyl ester of 3α,12α-Diacetoxy-7-keto-5β-cholanoic acid	Methyl ester of 3α,12α-diacetoxy,7-oxo C	3α,12α-OAc,7-oxo
27. Methyl ester of 3,6-Diketo-5β-cholanoic acid	Methyl ester of 3,6-dioxo C	3,6-dioxo

3. Determination of retention indices

The mobile phases used were methanol-water mixtures in different portion volumes varying between 30-35% (v/v) (1.25% per step) for a part of BA and 38-40% (v/v) (0.5% per step) for the rest of BA on C18 column; 80-84% (v/v) (1% per step) on C8 column, respectively 61-65% (v/v) (1% per step) on CN column. The injection volume was 10 μL, and the mobile phase flow was 1 mL/min. The temperature was kept constantly at 25°C. The detection was realized by mass

spectrometer with negative fragmentation at 80 eV, by monitorizing the selected ion (SIM Mode), an exception being in case of CN column for compounds 23, 25, 26, si 27 for which positive fragmentation was selected at 60 eV.

4. RP-HPLC

Reverse-phase high-performance liquid chromatography (RP-HPLC) provides a variety of indices that can be used as lipophilicity estimators. The most popular lipophilicity indices measured by RP-HPLC are derived by the retention time, t_r, according to the following formula:

$$\log k = \log k_w - S\varphi \,, \tag{1}$$

where

$$\log k = \log(\frac{t_r - t_o}{t_o}) \,, \tag{2}$$

and t_o being the retention time of an non-retained solute, k_w refers to the isocratic k value for pure water as mobile phase, and is usually extrapolated value, S is related to the solvent strength of pure organic modifier as mobile phase and is specific to this solvent on the stationary phase.[13, 14]

A relatively new derived chromatographic index, φ_0, is increasingly proposed as an effective parameter to measure the lipophilicity of compounds. The φ_0 value represents the percentage (in volume) of organic modifier which is necessary to achieve its equal distribution between the two phases of the chromatographic medium and it can be calculated according to:[15, 16]

$$\varphi_0 = \log k_w / S \,, \tag{3}$$

In addition, we used also the lipophilicity scale obtained by applying Principal Component Analysis (PCA) directly to the matrix retention data (k and log k) resulted for all compounds and combinations of methanol-water and so the eigenvalues of the covariance matrix were obtained. The scores corresponding to the first principal component appeared to be one of the best solutions for the lipophilicity scale resulted from retention data.[17]

5. Molecular descriptors

The log P values were calculated by Chem3D Ultra 10[18] ($LogP_{CD,}$ $PartCoeff_{CD}$), four are given by the Dragon 5.4[19] (MLOGP-Moriguchi method, MLOGP2-Squared Moriguchi method, ALOGP-Ghose-Crippen method, ALOGP2-Squared Ghose-Crippen method), and two of them are calculated by Alchemy software ($LogP_{SciQSAR}$, $LogPc_{SciLogP}$)[20]. The ALOGPS 2.1-vcclab internet module allows the calculation of another nine log P values (ALOGPs, AClogP, AB/LogP, COSMOFraq, miLogP, ALOGP, MLOGP, KOWWIN, XLOGP2, XLOGP3, AverageLogP);[21] Chemsilico module available on the Internet offers 6 other parameters (CSLogP,CSLogWSo,CSWSo,CSPB,CSHIA,CSBBB) which give more information about the penetration of the celular membrane by BA[22] Other Log P values were given by different other programs.[23-26]

Bile acids and their derivatives were characterized by 1267 theoretical descriptors calculated using Dragon 5.4 software.[27, 28] The Dragon descriptors employed in this study can be arranged in the following groups: *descriptors 2D*: 2D autocorrelations; *descriptors 3D*: RDF, 3D-MORSE, GETAWAY, WHIM, geometrical properties and Randić molecular profiles; *others descriptors*: functional groups, atom-centered fragments, molecular properties, charge descriptors, and constitutional properties. In all cases the structures of the compounds were firstly preoptimized with the Molecular Mechanics Force Field (MM+) procedure included in Hyperchem version 7.5,[29] and the resulting geometries were further refined by means of the semi empirical method PM3 (Parametric Method-3) using the Fletcher-Reeves algorithm and a gradient norm limit of 0.009 kcal/Å.

The model significance obtained in this work, with the exclusion of redundant and noisy information, was analyzed by MobyDigs v.1.0 software[30] that calculated the regression models by using genetic algorithms (GA) to perform variable selection. The MobyDigs software provides also many statistical indices useful for evaluating the performance of the developed regression models.[31]

RESULTS AND DISCUSSION

The log k values of bile acids (BA) decrease linearly in all cases of RP-HPLC as the methanol concentration increases. High correlation coefficients (r) using Eq. 1 were obtained, in the majority of cases being over 0.98, an exception being on CN column (compound 21). The highest log k values were obtained on C8 and C18 columns compared to CN column, and it can be observed that the number of hydroxyl, keto- and diacetoxy groups, as well as their position and orientation determine the chromatographic behavior of BA and their derivatives.

The chromatographic behavior of the investigated compounds is in a very good agreement with their polarity (Table 2) as can be easily observed from the profiles of retention indices presented in Figure 1 a-e. By carefully examining the patterns the similarity and differences between the bonded phases investigated can be clearly observed.

The more lipophilic hydrocarbon surface lies on the convex (β) side of the steroid nucleus and is devoid of hydrophilic substituent. In contrast, the hydrocarbon surface on the concave (α) side is less hydrophobic due to its smaller total surface area and by the presence of one, two, or three hydroxyl (less commonly sulfate, glucuronidate, keto and diacetoxy) functions. Furthermore, the aliphatic side chain and conjugating amino and aceto functions terminate in a strong ionic polar group which contributes a strong hydrophilic moiety to the less hydrophobic side.

Studies within the bile salts demonstrate that HPLC mobility, which correlates with lipophilicity, was markedly influenced by both position and orientation, in addition to number of hydroxyl and oxo-functions, in that mobility decreased in the order C>CDC>DC>LC. The cholic acid molecule has the largest planar polarity since its β side of the molecule is separated from the hydrophilic α side. For its mono- and diketo-derivatives planar polarity decreases because the β side of the molecule becomes more hydrophilic (less lipophilic) due to shift of the oxo group toward steroid skeleton mean plane. Moreover, partial inversion of polarity occurs for DC acid because β side of the molecule becomes more polar because of displacement of the oxygen atoms at C3, C7 and C12 oxo group, and α side becomes less polar (more lipophilic) due to appearance of the lipophilic island on the α side of the steroid skeleton.

The eigenvalues obtained by applying PCA show that the first principal component accounts for 99.65% (*k*) and 97.10% (log *k*) of the total variance in the case of C18, 97.37% (*k*) and 94.29% (log *k*) for C8, and 99.72% (*k*) and 99.20% (log *k*) for CN column, respectively.

The correlation between different lipophilicity indices is presented in Table 3. The φ_0 values correlate better with log k_w on C18 and CN columns (-0.62, and -0.60). The scores corresponding to the first principal component corresponding to *k* and log *k* values are better correlated with φ_0, and the highest correlation coefficients were obtained for C8 column.

In each case a high correlation between the two regression parameters, the intercept log k_w and the slope S, was observed (-1.00 on C8, -0.92 on C18 and -0.98 on CN), and the high linear correlation was considered a possibility of finding congeneric classes within large groups of compounds.

A comparative study has also been developed between the lipophilicity indices of bile acids obtained by HPLC and calculated partition coefficients using different theoretical methods.

Among the log P values, the most similar to the experimental partition coefficients were those obtained on C8 and C18 columns (higher than 0.7) (Table 4). The highest compatibility of experimental log k_w values was found with:

(i) CSLogP, $LogPc_{SciLogP}$, miLogP, AB/LogP and MolLogP on C8 (values between 0.80-0.90); (ii) miLogP, $LogPc_{SciLogP}$, $LogP_{CD}$, VirtualLogP, XLOGP2, AvLogP and cLogP on C18 (values between 0.80-0.85) and as for CN column worse corelations were obtained.

Among calculated values of partition coefficients, CSLogP and miLogP correlate better with log k_w on C8 and C18 columns.

The lipophilicity index, φ_0, appeares to be the best solution for the lipophilicity scale resulted from retention data, in all cases the values being > 0.85. Comparison of these calculation procedures reveal that the most appropriate log P values to the bile acids chromatographic indices are the ones which combine additive atomic contributions, atom-type electrotopological-state (E-state), neural network modeling indices and group contributions.

All statements above are well supported by lipophilicity charts obtained by scatterplots of scores corresponding to log k values onto the planes described by the first two principal components. It is interesting to observe that the series of compounds investigated form practically five different congeneric classes (Figure 2 (a-c) in a good agreement with their chemical structure: diacetoxy- (23, 25-27), oxo-derivatives (16-22, 24), primary and secondary bile acids (1-7), and finally the glyco- (8, 10, 11, 14) and tauro- (9, 12 13, 15) conjugates.

The position of each compound within the graphs is also in a good agreement with the position and orientation of hydroxyls and the presence of polar groups –COOH, $–SO_3^-$, -C=O and $-OCOCH_3$, respectively.

Predictive models with three of the most contributing descriptors were ascertained through the statistical parameters present in Table 5 in order to establish the correlation of lipophilicity indices of the studied compounds with their structural and physicochemical properties. The statistical significance of each model is evaluated by the determination coefficient R^2, leave-one-out crossvalidation coefficient Q^2, predictive error sum of squares PRESS, standard error s and Fisher test F. The high Q^2 and R^2 values are considered as a proof of goodness of fit and high predictive ability and robustness of the obtained models. PRESS can be used to assess the predictive performance of each model; the best regression will have a comparatively small predictive sum of squares.

Table 2

Chromatographic retention data of studied bile acids

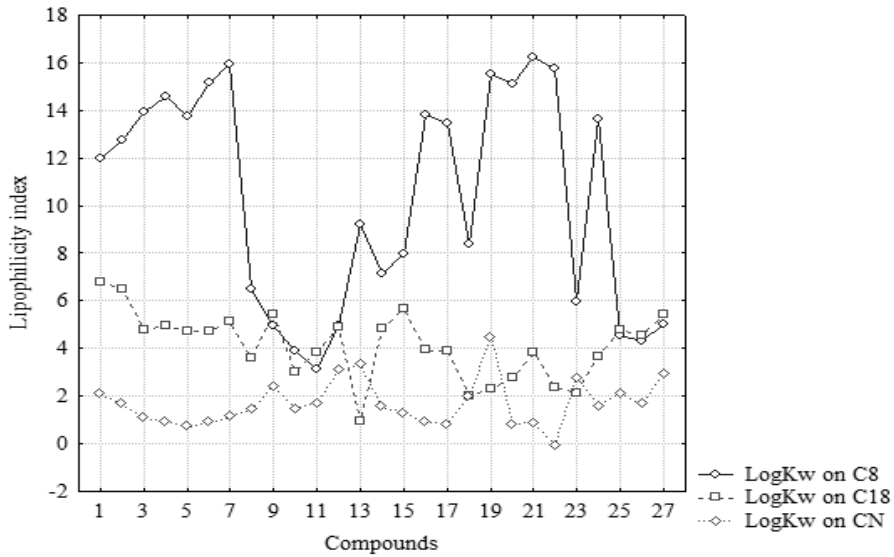
Cpd.	Column														
	C8					C18					CN				
	log k_w	S	φ_o	$PC1_k$	$PC1_{\log k}$	log k_w	S	φ_o	$PC1_k$	$PC1_{\log k}$	log k_w	S	φ_o	$PC1_k$	$PC1_{\log k}$
1	11.988	-0.139	-86.494	-8.449	-1.661	6.821	-0.076	-89.398	-7.941	-1.004	2.093	-0.033	-62.838	-0.012	-0.129
2	12.745	-0.151	-84.235	-3.179	-1.022	6.477	-0.075	-86.132	-0.032	-0.425	1.666	-0.028	-59.092	0.473	0.106
3	13.934	-0.166	-83.886	-3.079	-0.969	4.763	-0.054	-88.031	1.650	-0.263	1.074	-0.019	-55.953	0.567	0.155
4	14.563	-0.178	-81.904	-0.124	-0.236	4.938	-0.060	-82.567	-20.578	-1.360	0.943	-0.018	-51.266	0.857	0.343
5	13.777	-0.170	-81.280	0.571	0.012	4.734	-0.059	-79.828	-10.399	-0.984	0.718	-0.015	-49.164	0.808	0.306
6	15.151	-0.186	-81.368	0.424	-0.004	4.728	-0.059	-80.132	-10.735	-1.008	0.900	-0.018	-50.274	0.902	0.375
7	15.925	-0.195	-81.667	0.112	-0.109	5.158	-0.064	-80.721	-18.113	-1.266	1.151	-0.021	-53.766	0.789	0.296
8	6.491	-0.083	-78.393	1.521	0.392	3.635	-0.052	-70.037	6.235	0.407	1.437	-0.025	-56.794	0.645	0.205
9	4.957	-0.065	-76.263	1.742	0.555	5.459	-0.081	-67.481	6.322	0.550	2.401	-0.032	-75.503	-3.391	-1.025
10	3.912	-0.055	-71.512	2.113	0.992	2.991	-0.047	-64.195	8.800	1.068	1.468	-0.027	-54.757	0.872	0.358
11	3.129	-0.043	-73.276	1.777	0.545	3.838	-0.053	-71.882	4.924	0.167	1.717	-0.029	-58.395	0.562	0.158
12	4.958	-0.062	-79.462	1.131	0.073	4.876	-0.072	-67.438	6.740	0.590	3.089	-0.042	-73.902	-4.268	-1.159
13	9.207	-0.117	-79.029	1.548	0.494	0.955	-0.018	-51.897	9.907	1.540	3.322	-0.047	-71.439	-3.464	-1.027
14	7.165	-0.088	-81.514	0.521	-0.181	4.854	-0.065	-74.901	-1.477	-0.445	1.549	-0.026	-60.741	0.237	-0.019
15	7.963	-0.095	-83.908	-0.956	-0.676	5.683	-0.065	-87.160	0.479	-0.383	1.280	-0.023	-56.650	0.606	0.180
16	13.793	-0.171	-80.520	1.125	0.305	3.932	-0.051	-77.548	-0.415	-0.432	0.908	-0.018	-49.612	0.950	0.411
17	13.456	-0.168	-80.192	1.298	0.408	3.908	-0.051	-76.770	0.421	-0.355	0.813	-0.017	-48.946	0.920	0.387
18	8.366	-0.106	-78.625	1.633	0.523	2.023	-0.030	-66.546	8.639	1.729	2.013	-0.035	-57.673	0.744	0.274
19	15.547	-0.196	-79.198	1.879	0.974	2.295	-0.034	-68.113	8.602	1.697	4.472	-0.074	-60.520	0.666	0.258
20	15.126	-0.191	-79.109	1.893	0.981	2.754	-0.042	-64.946	8.691	0.998	0.808	-0.015	-53.125	0.642	0.198
21	16.255	-0.205	-79.291	1.872	0.982	3.839	-0.059	-65.287	8.339	0.952	0.886	-0.018	-49.217	0.955	0.415
22	15.767	-0.199	-79.154	1.901	1.012	2.386	-0.027	-88.691	8.716	1.809	-0.053	-0.002	35.533	0.648	0.196
23	5.949	-0.066	-90.540	-5.795	-1.544	2.138	-0.032	-65.994	0.706	-0.385	2.776	-0.042	-65.476	-0.679	-0.376
24	13.634	-0.169	-80.768	0.921	0.198	3.687	-0.040	-93.104	1.337	-0.421	1.591	-0.026	-60.946	0.224	-0.024
25	4.534	-0.054	-84.117	-0.256	-0.547	4.765	-0.051	-92.891	-3.753	-0.765	2.116	-0.033	-64.319	-0.296	-0.245
26	4.293	-0.051	-83.852	-0.099	-0.499	4.554	-0.049	-92.935	-2.611	-0.681	1.682	-0.026	-64.938	-0.289	-0.247
27	5.029	-0.057	-87.611	-2.048	-0.997	5.422	-0.057	-95.959	-14.455	-1.329	2.949	-0.045	-65.246	-0.667	-0.369

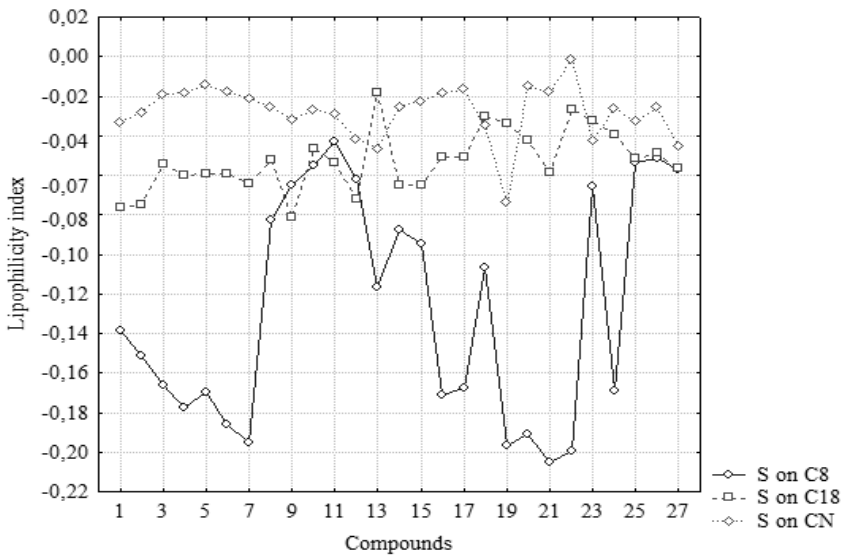
Legend: log k_w -isocratic k value for pure water, S -solvent strength of organic modifier, φ_0 -lipophilicity index,
$PC1_k$, $PC1_{\log k}$ -scores corresponding to first principal component.

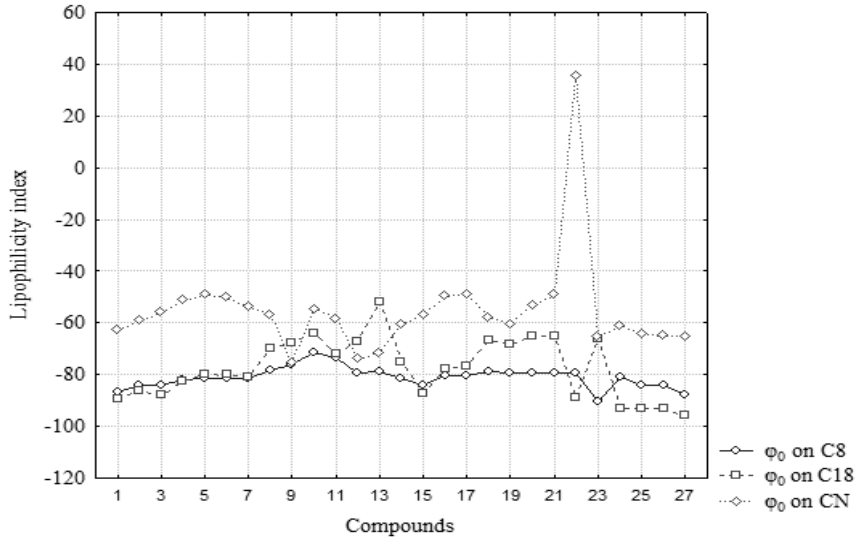
Table 3

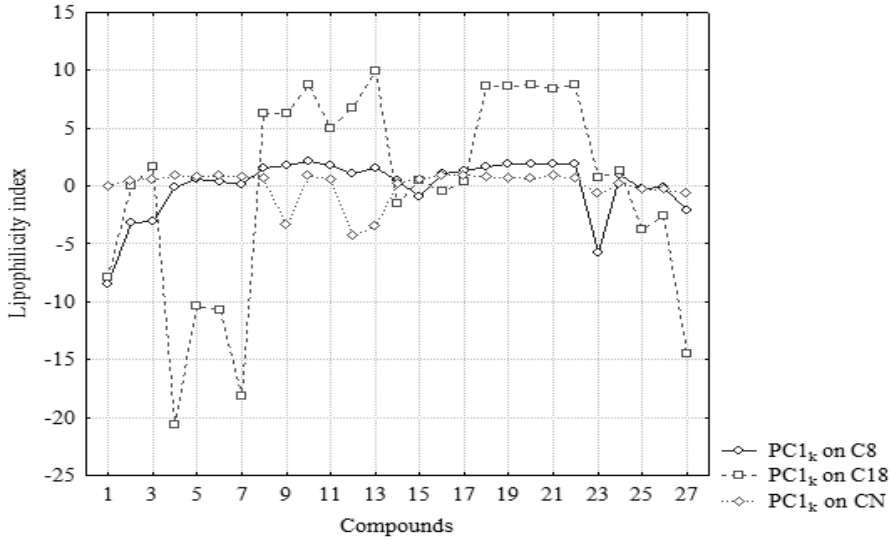
Correlation matrix of chromatographic retention indices

Column	Index	Column														
		C8					C18					CN				
		$\log k_w$	S	φ_o	$PC1_k$	$PC1_{\log k}$	$\log k_w$	S	φ_o	$PC1_k$	$PC1_{\log k}$	$\log k_w$	S	φ_o	$PC1_k$	$PC1_{\log k}$
C8	$\log k_w$	1.00	**-1.00**	-0.08	0.03	0.19	-0.04	0.09	-0.10	-0.19	-0.01	**-0.41**	0.34	**0.45**	**0.47**	**0.54**
	S		1.00	0.02	-0.10	-0.26	0.08	-0.12	0.06	0.14	-0.04	**0.42**	-0.34	**-0.47**	**-0.47**	**-0.54**
	φ_o			1.00	**0.79**	**0.86**	-0.35	0.12	**0.53**	**0.51**	**0.59**	-0.13	0.11	0.13	-0.01	0.12
	$PC1_k$				1.00	**0.91**	**-0.47**	0.30	**0.41**	**0.40**	**0.52**	-0.14	0.13	0.20	-0.01	0.12
	$PC1_{\log k}$					1.00	**-0.58**	0.37	**0.54**	**0.55**	**0.70**	-0.14	0.12	0.32	0.06	0.19
C18	$\log k_w$						1.00	**-0.92**	**-0.62**	**-0.59**	**-0.72**	-0.18	0.20	-0.23	0.03	-0.01
	S							1.00	0.30	**0.41**	**0.55**	0.13	-0.17	0.32	0.10	0.09
	φ_o								1.00	**0.58**	**0.65**	0.29	-0.28	-0.22	-0.32	-0.20
	$PC1_k$									1.00	**0.89**	0.20	-0.18	0.09	-0.24	-0.19
	$PC1_{\log k}$										1.00	0.22	-0.22	0.26	-0.19	-0.11
CN	$\log k_w$											1.00	**-0.98**	**-0.60**	**-0.59**	**-0.61**
	S												1.00	**0.57**	**0.43**	**0.46**
	φ_o													1.00	**0.42**	**0.43**
	$PC1_k$														1.00	**0.98**
	$PC1_{\log k}$															1.00


Lipophilicity index
Compounds
LogKw on C8
LogKw on C18
LogKw on CN


Lipophilicity index
Compounds
S on C8
S on C18
S on CN


Lipophilicity index
Compounds
φ0 on C8
φ0 on C18
φ0 on CN


Lipophilicity index
Compounds
PC1k on C8
PC1k on C18
PC1k on CN

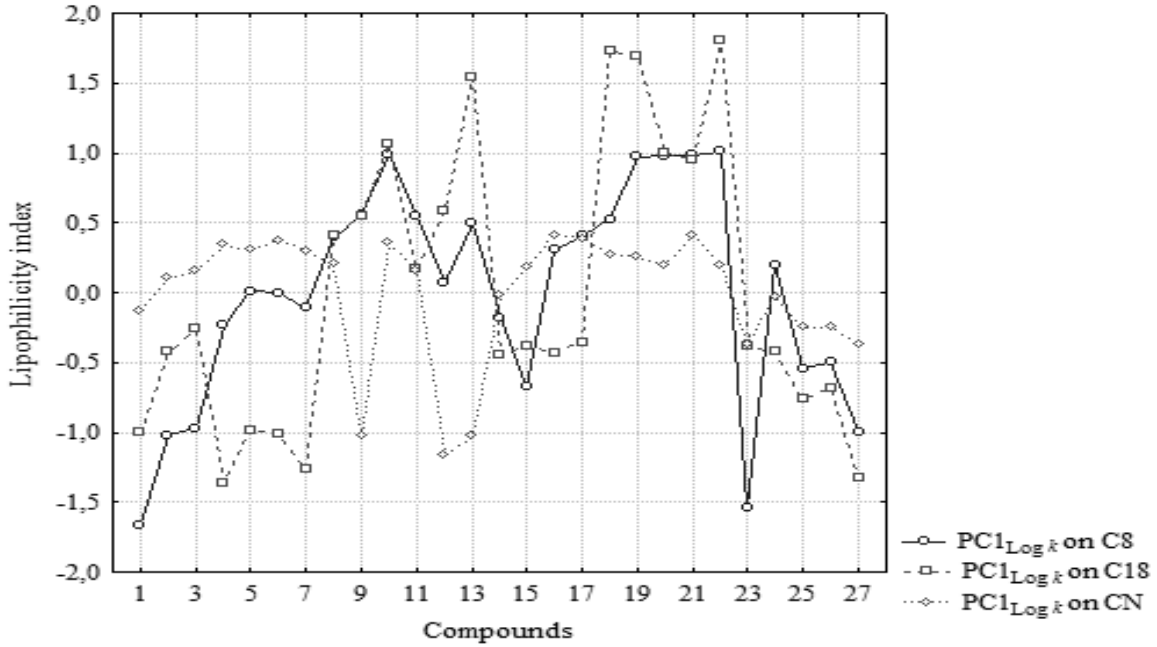


Fig. 1 – Profiles of lipophilicity indices corresponding to C8, C18 and CN column.

Table 4

Correlation coefficients of lipophilicity indices with calculated log P values

Variable	Column														
	C8					C18					CN				
	log k_w	S	φ_o	$PC1_k$	$PC1_{\log k}$	log k_w	S	φ_o	$PC1_k$	$PC1_{\log k}$	log k_w	S	φ_o	$PC1_k$	$PC1_{\log k}$
ALOGPs	0.19	-0.17	**-0.51**	**-0.46**	-0.33	0.01	0.14	-0.35	-0.16	-0.17	-0.07	0.03	0.24	0.31	0.18
AClogP	0.18	-0.15	**-0.57**	**-0.46**	**-0.41**	0.10	0.10	**-0.48**	-0.33	-0.34	-0.08	0.05	0.18	0.31	0.18
AB/logP	0.10	-0.05	**-0.72**	**-0.61**	**-0.60**	0.28	-0.13	**-0.42**	-0.35	**-0.39**	0.16	-0.12	-0.09	-0.16	-0.28
COSMOFrag	-0.29	0.34	**-0.56**	**-0.62**	**-0.71**	**0.46**	**-0.39**	-0.32	**-0.41**	**-0.56**	0.16	-0.10	-0.29	-0.28	-0.38
miLogP	0.26	-0.22	**-0.62**	**-0.53**	**-0.50**	0.22	-0.01	**-0.54**	**-0.45**	**-0.46**	-0.14	0.11	0.14	0.32	0.21
ALOGP	0.03	0.02	**-0.53**	**-0.61**	**-0.57**	0.35	-0.26	-0.36	-0.33	**-0.42**	-0.02	0.05	0.01	-0.05	-0.15
MLOGP	0.14	-0.09	**-0.63**	**-0.62**	**-0.58**	0.32	-0.16	**-0.48**	**-0.40**	**-0.45**	-0.03	0.04	0.06	0.07	-0.04
KOWWIN	0.10	-0.06	**-0.52**	**-0.52**	**-0.50**	0.26	-0.09	**-0.48**	**-0.41**	**-0.49**	-0.15	0.13	0.09	0.28	0.16
XLOGP2	-0.13	0.18	**-0.50**	**-0.62**	**-0.61**	**0.46**	**-0.39**	-0.36	-0.33	**-0.45**	0.04	0.01	-0.06	-0.15	-0.25
XLOGP3	-0.08	0.13	**-0.50**	**-0.60**	**-0.60**	**0.41**	-0.33	-0.35	-0.35	**-0.46**	0.04	0.01	-0.09	-0.13	-0.23
AverageLogP	0.06	-0.02	**-0.62**	**-0.62**	**-0.59**	0.31	-0.16	**-0.46**	**-0.39**	**-0.46**	-0.01	0.03	0.02	0.07	-0.05
$MLOGP_{Dragon}$	0.14	-0.09	**-0.63**	**-0.62**	**-0.58**	0.32	-0.16	**-0.48**	**-0.40**	**-0.45**	-0.03	0.04	0.06	0.07	-0.04
$MLOGP^2$	0.12	-0.07	**-0.66**	**-0.69**	**-0.63**	0.35	-0.18	**-0.49**	**-0.41**	**-0.47**	-0.02	0.03	0.04	0.07	-0.05
$ALOGP_{Dragon}$	0.03	0.02	**-0.53**	**-0.61**	**-0.57**	0.35	-0.26	-0.36	-0.33	**-0.42**	-0.02	0.05	0.01	-0.04	-0.15
$ALOGP^2$	0.02	0.03	**-0.57**	**-0.69**	**-0.62**	0.38	-0.28	-0.36	-0.32	**-0.42**	-0.00	0.03	-0.02	-0.03	-0.14
Virtual LogP	-0.02	0.07	**-0.69**	**-0.62**	**-0.61**	0.25	-0.07	**-0.48**	-0.37	**-0.45**	0.05	-0.04	0.00	0.05	-0.09
MolLogP	0.05	0.00	**-0.55**	**-0.63**	**-0.66**	**0.48**	-0.32	**-0.51**	**-0.60**	**-0.67**	-0.10	0.09	-0.05	0.15	0.05
cLogP	0.21	-0.18	**-0.53**	**-0.43**	-0.37	0.08	0.12	**-0.49**	-0.33	-0.34	-0.11	0.07	0.19	0.37	0.24
ClogP	0.02	0.02	**-0.52**	**-0.55**	**-0.52**	0.32	-0.20	**-0.42**	-0.31	-0.36	-0.05	0.10	0.17	-0.08	-0.19
$LogPc_{SciQSAR}$	0.19	-0.14	**-0.58**	**-0.74**	**-0.55**	0.27	-0.20	-0.21	-0.24	-0.25	0.11	-0.08	0.06	-0.08	-0.14
$LogPc_{SciLogP}$	**0.66**	**-0.65**	-0.18	0.03	0.05	0.10	-0.06	-0.21	-0.17	-0.07	-0.24	0.19	0.28	0.28	0.31
$LogP_{CD}$	0.25	-0.22	**-0.59**	**-0.49**	**-0.39**	0.10	0.04	-0.36	-0.21	-0.21	-0.01	0.01	0.19	0.07	-0.03
$PartCoeff_{CD}$	-0.07	0.11	**-0.54**	**-0.62**	**-0.61**	**0.39**	-0.29	**-0.39**	-0.35	**-0.45**	0.01	0.02	-0.05	-0.01	-0.13
CSLogP	0.12	-0.07	**-0.65**	**-0.59**	**-0.60**	0.33	-0.16	**-0.49**	**-0.58**	**-0.57**	-0.03	0.04	0.01	0.06	-0.03
CSLogWSo	0.07	-0.10	**0.50**	0.37	**0.39**	-0.02	-0.15	0.37	0.35	0.36	0.03	0.01	-0.02	-0.32	-0.20
CSWSo	-0.29	0.28	0.17	0.12	0.04	0.08	-0.16	0.16	0.23	0.15	0.14	-0.09	-0.17	-0.35	-0.31
CSPB	0.23	-0.21	-0.12	-0.25	-0.24	0.38	-0.38	-0.19	**-0.41**	-0.38	-0.12	0.17	0.04	-0.16	-0.15
CSHIA	**0.58**	**-0.58**	-0.10	-0.06	0.07	-0.05	0.10	-0.14	-0.20	-0.09	-0.28	0.20	0.28	**0.45**	**0.46**
CSBBB	0.34	-0.32	**-0.46**	-0.30	-0.22	-0.06	0.20	-0.27	-0.32	-0.20	-0.06	-0.02	0.19	**0.44**	0.38

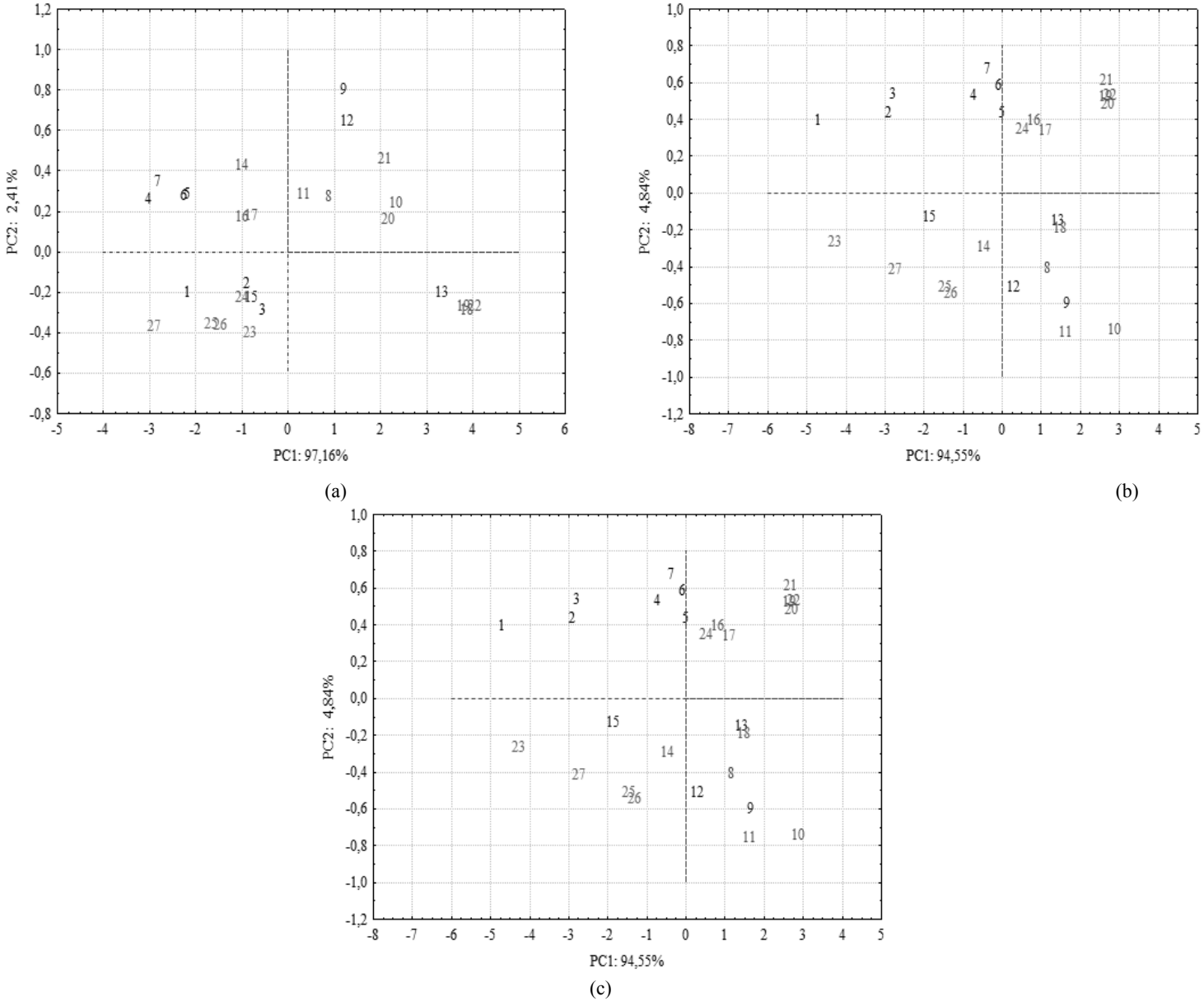


Fig. 2 – Lipophilicity charts corresponding to log k values: (a) C8 (b) C18, and (c) CN column.

The best models yield a determination coefficient between 0.75-0.94 in the case of C8 (Eq. 4-8) and between 0.74-0.82 in the case C18 (Eq. 9-12) column. C8 seems to be the most adequate column for the estimation and characterization of bile acids lipophilicity as follows:

$$\text{Log } k_w = 95.591 - 366.036\boldsymbol{PW3} - 13.409\boldsymbol{Mor32m} + 93.994\boldsymbol{HATS6e} \quad (4)$$

$$S = -1.179 + 4.514\boldsymbol{PW3} + 0.161\boldsymbol{Mor32m} - 1.159\boldsymbol{HATS6e} \quad (5)$$

$$\varphi_0 = -438.882 + 0.504D/Dr06 + 0.827RDF130e + 507.235REIG \quad (6)$$

$$PC1_k = -42.595 - 6.605ASP - 2.824MLOGP2 - 17.193BLTA96 \quad (7)$$

$$PC1_{logk} = 67.201 - 101.496X0Av + 12.012MATS4e - 0.135Tp \quad (8)$$

$$\text{Log } k_w = 10.747 - 4.934\boldsymbol{EEig15x} + \mathbf{0.229}\boldsymbol{RDF075p} + 7.040\boldsymbol{Mor20p} \quad (9)$$

$$S = 1.429 + 0.046\boldsymbol{EEig13d} - 0.123ESpm09u - 0.038\boldsymbol{Mor20u} \quad (10)$$

$$\varphi_0 = -262.220 - 1.626\boldsymbol{RDF045p} - 26.937\boldsymbol{Mor20u} - 35.679\boldsymbol{Mor05v} \quad (11)$$

$$PC1_k = -70.192 + 0.438TI1 - 32.019\boldsymbol{Mor20p} - 6.396H\text{-}047 \quad (12)$$

$$PC1_{logk} = -7.523 - 0.869\boldsymbol{Mor05m} - 3.295\boldsymbol{Mor20v} + 1.664Hypertens\text{-}80 \quad (13)$$

The regression equations obtained for CN column present also significant determination coefficients (0.73-0.91):

$$\text{Log } k_w = -151.228 - 1.038MAXDP + 154.910PCR - 3.271\boldsymbol{Infective\text{-}80} \quad (14)$$

$$S = 0.811 - 0.055\boldsymbol{EEig05d} + 0.209VEA1 + 0.046\boldsymbol{Infective\text{-}80} \quad (15)$$

$$\varphi_0 = -95.764 - 1.605RDF100u + 1.876RDF050m + 1025.198R1v+ \quad (16)$$

$$PC1_k = 35.460 - 17.857\boldsymbol{MATS4e} - 11.623\boldsymbol{EEig03d} + 38.407G1p \quad (17)$$

$$PC1_{logk} = -45.996 + 46.595Me + 12.029\boldsymbol{MATS1v} - 0.577Mor26m \quad (18)$$

All stationary phases are characterized by 3D descriptors which express the molecular size, shape, branch and charge characteristics. The topological descriptors are based on molecular graphs as a source of probability distributions to which the information theory definitions apply. GETAWAY, RDF, molecular and topological descriptors are most frequently selected by the genetic algorithm (GA) variable selection method, having the best overall performance in modeling the considered properties. Topological and RDF descriptors give holistic information on the molecular structure; most of GETAWAY describes only portions of the molecular structure, while molecular descriptors describe the structure or shape of molecules.

In case of C8 column, there are descriptors which are retained by log k_w and S lipophilicity indices: PW3 – path/walk 3-Randic shape index which has a high negative value on log k_w which states that specific polarizability accounts the favorable effects of dipole-dipole interactions between the solutes and the bulk phases; Mor32m – 3D-MoRSE-signal 32/weighted by atomic masses; HATS6e – leverage-weighted autocorrelation of lag 6/weighted by the atomic Sanderson electronegativities and application of the Sanderson electronegativities as weighting coefficients, takes into account, to some degree, charge distribution inside a molecule. Other important contributions are given by REIG descriptor upon φ_0 which has larger values for branched molecules and X0Av descriptor which encodes information about size, branching, cyclization, unsaturation and heteroatom content in the molecule.

Table 5

Quality statistical parameters of predictive models

Column	Eq.	Q^2	R^2	s	PRESS	F
C8	(4)	0.9236	0.9402	1.204	42.555	120.5
	(5)	0.9100	0.9297	0.016	0.008	101.5
	(6)	0.7636	0.8311	1.749	98.516	37.7
	(7)	0.7653	0.8158	1.160	39.411	33.9
	(8)	0.6527	0.7535	0.410	5.449	23.4
C18	(9)	0.7580	0.8230	0.632	12.557	35.6
	(10)	0.6691	0.7649	0.008	0.002	24.9
	(11)	0.7084	0.7669	5.895	999.991	25.2
	(12)	0.6621	0.7394	4.724	665.536	21.8
	(13)	0.7528	0.8170	0.454	6.4	34.2
CN	(14)	0.6354	0.7284	0.544	9.127	20.6
	(15)	0.5723	0.7292	0.008	0.002	20.6
	(16)	0.1436	0.3405	16.949	8578.822	4.0
	(17)	0.8800	0.9073	0.460	6.135	75.1
	(18)	0.8679	0.9131	0.141	0.697	80.6

As concerning C18 column, the models retained RDF descriptors weighted by atomic polarizabilities at 0.45 and 0.75 Å distance which are sensitive to the 2 and 3-dimensional molecular structure and size of the molecule; topological aspects, such as edge adjacency, dipole moment (EEig15x and EEig13d), atomic van der Waals volume and polarizability seem to characterize the way by which atoms are connected, indicating the intrinsic polarity of the molecule and also discriminating between isomers; the polarizability and volume factor decoded in Mor20p and Mor20v descriptors allows us to assert that separation of compounds depends, to a certain extent, on them being correlated with the chemical reactivity.

As for CN, the models retained molecular properties Infective-80 -Ghose-Viswanadhan-Wendoloski anti-infective-like index at 80%), EEig05d and MATS descriptors. The most contributing descriptor to the investigated φ_o model is R1v+ -R maximal autocorrelation of lag 1/weighted by atomic van der Waals volumes, characterized by a high positive coefficient which confirms that separation of compounds depend, to a certain extent, on the volume of molecules that correlated with the chemical reactivity.

On the basis of presented correlations, it may be appreciated that the lipophilicity indices determined on C8 and CN columns might be the best choices for the lipophilicity prediction of bile acids and their derivatives.

The contribution of 2D and 3D descriptors which are related to atomic mass and volumes, together with reactivity parameters such as polarizability and electronegativity seem to control the chromatographic mechanism (lipophilicity) on all columns.

CONCLUSIONS

Different indices of lipophilicity for bile acids and their derivatives were determined by reversed-phase high-performance liquid chromatography on C8, C18 and CN chromatographic columns using methanol–water as mobile phase. The highest log *k* values were obtained on C8 and C18 columns compared to CN column. Highly significant correlations were obtained between different experimental indices of lipophilicity and computed log P values, C8 column seems to be more suited for the estimation of lipophilicity. In addition, the results obtained in this study applying PCA may be used in interpreting the molecular mechanism of interactions between eluents and columns with different polarities and to explain the chromatographic behavior of compounds. The contribution of 2D and 3D descriptors which are related to atomic mass and volumes, together with reactivity parameters such as polarizability and electronegativity seem to control the chromatographic mechanism (lipophilicity) on all columns.

REFERENCES

1. D.W. Russell and K.D.R. Setchells, *Biochemistry,* **1992**, *31*, 4737.
2. J.M. Ridlon, D-J Kang and P.B. Hylemon, *J. Lipid Res.*, **2006**, *47*, 241.
3. M. Schwarz, *Drug Discov. Today*, **2004**, *1*, 205.
4. P.R. Debruyne, E.A. Bruyneel, X. Li, A. Zimber, C. Gespach and M.M. Mareel, *Mutat. Res.*, **2001**, *480–481*, 359.
5. A.F. Hofmann, *Arch Intern. Med.*, **1999**, *159*, 2647.
6. A.P. Polo, P.J. Oliviera, A.J.M. Moreno and C.M. Palmeira, *Toxicol. Sci.*, **2000**, *57*, 177.
7. S. Jönsson, L.A. Eriksson and B. van Bavel, *Anal. Chim. Acta*, **2008**, *621*, 155.
8. D. Vrakas, C. Giaginis and A. Tsantili-Kakoulidou, *J. Chromatogr. A,* **2006**, *1116*, 158.
9. X. Liu, H. Tanaka, A. Yamauchi, B. Testa and H. Chuman, *J. Chromatogr. A,* **2005**, *1091*, 51.
10. C. Giaginis and A. Tsantili-Kakoulidou, *J. Liq. Chromatogr. Relat. Technol.,* **2008**, *31*, 79.
11. M.J. Armstrong and M.C. Carey*, J. Lipid Res.,* **1982**, *23*, 70.
12. M. Mikov and J.P. Fawcett, "Bile Acids", Medishet Publisher, Geneva, 2007.
13. M. Poša, V. Guzsvány, J. Csanádi, S. Kevrešan and K. Kuhajda, *Eur. J. Pharm. Sci.,* **2008**, *34*, 281.
14. L.R. Snyder, J.W. Dolan and J.R. Gant, *J. Chromatogr.,* **1979**, *165*, 3.
15. H.A. Cooper and R.J. Hurutbise, *J. Chromatogr.,* **1986,** *360*, 313.
16. K. Valkó, P. Siégel*, J. Chromatogr.,* **1993,** *631,* 49.
17. C.M. Du, K. Valkó, C. Bevan, D. Reynolds, *Anal. Chem.,* **1998,** *70*, 4228.
18. C. Onişor, D. Kovala-Demertzi, M.A. Demertzis, C. Sârbu, M. V. Diudea, *MATCH Commun. Math. Comput. Chem.,* **2008**, *60,* 1007.
19. Chem3D Ultra 10 software, http://www.cambridgesoft.com.
20. DRAGON for Windows (software for molecular descriptor calculations), Version 5.4 – 2005. http://www.talete.mi.it.
21. Alchemy 2000 software, http://www.cambridgesoft.com.
22. ALOGPS 2.1software, http://www.vcclab.org/lab/alogps/start.html.
23. http://www.chemsilico.com.
24. OSIRIS software, http://www.organic-chemistry.org/prog/peo/osiris property explorer.
25. Drug likeness and molecular property prediction, http://www.molsoft.com.VEGA online, http://www.ddl.unimi.it.
26. Chemaxon software, http://intro.bio.umb.edu/111-112/OLLM.
27. V. Consonni, R. Todeschini, M. Pavan, *J. Chem. Inf. Model.,* **2002,** *42*, 682.
28. R. Todeschini, V. Consonni, in R. Mannhold, H. Kubinyi, H. Timmerman (Eds.) Handbook of molecular descriptors, Wiley-VCH, Weinheim, 2000.
29. HyperChem(TM) Professional 7.5 for Windows, Molecular Modeling System, Hypercube, Inc. and Autodesk, Inc.
30. R. Todeschini, Moby Digs Academic version software for variable subset selection by genetic algorithms, Rel. 1.0 for Windows, Talete, Milan, 2004. http://www.talete.mi.it.
31. L. Scotti, M.T. Scotti, H.M. Ishiki, M.J.P. Ferreira, V.P. Emerenciano, C.M. de S. Menezes, E.I. Ferreira, *Food Chem.,* **2007**, *105*, 77.