Talanta 83 (2011) 1634–1642

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Determination and importance of temperature dependence of retention coefficient (RPHPLC) in QSAR model of nitrazepams' partition coefficient in bile acid micelles

Mihalj Poša*, Ana Pilipović, Mladena Lalić, Jovan Popović

Department of Pharmacy, Faculty of Medicine, University of Novi Sad, Hajduk Veljkova 3, 21000 Novi Sad, Serbia

ARTICLE INFO

Article history:
Received 13 July 2010
Received in revised form 28 October 2010
Accepted 22 November 2010
Available online 30 November 2010

Keywords:
QSAR
Retention coefficient
PCA
MLR
Bile acids

ABSTRACT

Linear dependence between temperature (t) and retention coefficient (k, reversed phase HPLC) of bile acids is obtained. Parameters (a, intercept and b, slope) of the linear function $k = f(t)$ highly correlate with bile acids' structures. Investigated bile acids form linear congeneric groups on a principal component (calculated from $k = f(t)$) score plot that are in accordance with conformations of the hydroxyl and oxo groups in a bile acid steroid skeleton.

Partition coefficient (K_p) of nitrazepam in bile acids' micelles is investigated. Nitrazepam molecules incorporated in micelles show modified bioavailability (depo effect, higher permeability, etc.). Using multiple linear regression method QSAR models of nitrazepams' partition coefficient, K_p are derived on the temperatures of 25 °C and 37 °C. For deriving linear regression models on both temperatures experimentally obtained lipophilicity parameters are included (PC1 from data $k = f(t)$) and *in silico* descriptors of the shape of a molecule while on the higher temperature molecular polarisation is introduced. This indicates the fact that the incorporation mechanism of nitrazepam in BA micelles changes on the higher temperatures. QSAR models are derived using partial least squares method as well. Experimental parameters $k = f(t)$ are shown to be significant predictive variables. Both QSAR models are validated using cross validation and internal validation method. PLS models have slightly higher predictive capability than MLR models.

1. Introduction

The aim of Quantitative Structure Activity (Property) Relationship (QSA (P) R) research is to find functional dependence between molecule structure and its pharmaco-biochemical activities or physico-chemical properties. Important feature of derived mathematical model is its ability to predict activity (property) of molecules not directly included in experiment (molecules not yet synthesised or those with limited *in vivo* and *in vitro* experiments due to economic or ethical reasons). Mathematical models give additional information about activity of a molecule and through correlations with molecular descriptors explain receptors binding places (enzyme, ionic channel, etc.) [1–8].

Bile acids (BAs) are surface active molecules with steroid skeleton [9–11]. Besides their well known physiological role in lipid metabolism regulation, BAs are used as promoters in transport of some drugs through the cell membrane or other physiological barriers (blood–brain barrier, etc.) [12–15]. Above certain concentration – critical micellar concentration BAs form aggregates i.e. micelles that have the possibility to accept hydrophobic molecule – guest (drug) changing thus its bioavailability [16–18]. Interaction between BA micelle and its hydrophobic guest can be described with partition coefficient (K_p) [19]. If a BA is more hydrophobic it has a higher capacity to accept hydrophobic drug so its effect on BA bioavailability is higher.

Everything mentioned above indicates importance of bile acids' hydrophobicity (lipophilicity) in describing interactions between their micellar solutions and nitrazepam. Thus, it is expected that QSAR model for partition coefficient contains descriptor for lipophilicity of bile acids. Bile acids' lipophilicity is usually expressed as a logarithm of partition coefficient between 1-octanol and water (log P). Traditional shake flask method for deriving log P is shown to be less precise and reproductive than different chromatographic parameters [20]. Because of that, goal of the first part of our work was to find a temperature dependence (t) of retention coefficient (k) obtained in reversed phase chromatography (RPHPLC) in order to gain experimental, predictive BA variables that describe the change of molecules' hydrophobicity [21], i.e. to derive a novel chromatographic parameter for describing bile acids' lipophilicity. According to that, hydrophobicity parameters should

* Corresponding author. Tel.: +38 121422760; fax: +38 121422760.
E-mail address: mihaljp@uns.ac.rs (M. Poša).

0039-9140/$ – see front matter
doi:10.1016/j.talanta.2010.11.050

be extracted from chromatographic data ($k=f(t)$) that describe the best structural i.e. conformational characteristics of bile acids. The second part of the work deals with deriving QSAR model between nitrazepams' (probe molecule) partition coefficient (K_p) and BA structure using multiple linear regression (MLP) and partial least square (PLS) methods. Experimentally obtained lipophilicity parameters of temperature dependence of BA retention coefficient (k) and *in silico* molecular descriptors (topological and electronic) are used as predictive variables in deriving QSAR model. Parameters of lipophilicity are included as predictive variables because a lot of molecular descriptors calculated from the molecular graph have the same value if they belong to the same congeneric group (for instance the cholic acid and its keto derivatives have the same value of Wieners' index which is 2045). In this paper a novel molecular descriptor (ND) is introduced which describes bile acids' steroid skeleton i.e. spatial orientation of substituents and their mutual distance in the BA steroid skeleton. Descriptor ND has characteristics of both 2D and 3D topological descriptors. That is the reason why bile acids from the same congeneric group have different ND descriptor values [22]. The relationship between chromatographic descriptors (descriptors gained from $k=f(t)$ is observed as well and *in silico* descriptors where ND is included). BAs with one, two and three hydroxyl group and their glyco-, tauro and oxo derivatives are investigated (Fig. 1). Particular attention is paid to BA oxo derivatives that have growing pharmacological use due to lower membranolytic activity [12,23].

2. Materials and methods

2.1. Chemicals and solutions

Bile acids (1–14), 98% purity purchased from Sigma, New Zealand were used as starting compounds for the synthesis of its oxo derivatives (15–25). The syntheses of bile acids oxo derivatives and their transformation to sodium salts were carried out according to previously described procedures [9,16]. Methanol, HPLC grade was obtained from Carlo Erba Reagenti, Italy, KH_2PO_4 and Na_2PO_4 from Lachner, Czech Republic. Nitrazepam 99, 98% purity was purchased from Sigma, New Zealand and NaCl pro analysis from Merck, Germany.

2.2. Reverse phase HPLC method

The HPLC system Agilent 1100 Series, equipped with degasser, binary pump, automatic injector and DAD detector with software system for data processing AgilentChemStation was used and the analyses were performed on a reversed-phase C-18 column: Eclipse Plus C18 (250 mm × 3 mm, 5 μm, 250 Å) column (Zorbax SD). The mobile phase was 0.01 M phosphate buffer:methanol = 70:130 (v/v) maintained at pH 7 and the injection volume was 10 μL. Solutions of bile acids and their derivatives in mobile phase were prepared in concentration of 1 mg/ml. All separations were performed isocratically at a flow rate of 1 ml/min and a column temperature changing from 20 to 45 °C. The detection was performed at 210 nm [24].

The HPLC capacity factor (k) was calculated from the eluted peak retention time (t):

$$k = \frac{t_x - t_0}{t_0}$$

where t_x and t_0 are the retention times of the bile acids and the unretained solvent front respectively. Linear dependence between (k) and temperature (t) is calculated for the each BA:

$$k = a + bt \quad (1)$$

2.3. Spectrophotometric determination of nitrazepams' partition coefficient in bile salt micelles

Experiments were carried out according to De Castro et al. [19], using Agilent 8453 spectrophotometer equipped with the Peltier thermostated cell holder (25 and 37 °C). Critical micellar concentrations of bile salts (cholic acid and its keto derivatives) used for calculation of nitrazepams' partition coefficient were taken from [10,9].

2.4. Data treatment

For PCA, MLR and PLS analyses Statistica 7 was used [25].

3. Molecular descriptors

The molecular descriptors calculated with SciQSAR option of the molecular modeling computer program ALCHEMY 2000 [24] were the following: the first-order (^1X) and the third-order (^3X) connectivity index, the zero-order ($^0X^v$) and first-order ($^1X^v$) valence connectivity index, the third-order shape index for molecule ($^3K_\alpha$), the Wiener (W) index, volume (V), molar mass (M), dipole moment (DM), molecular polarizability, specific molar polarizability (SP), the largest positive charge over the atoms in molecule, in electrons (Q_+), the largest negative charge over the atoms in a molecule, in electrons (Q_-), the sum of absolute values of the charges on each atom of the molecule, in electrons (SQ), the sum of absolute values of the charges on the nitrogen and oxygen in the molecule, in electrons (SQ_{NO}) and the partition coefficient (log P). Molecular ovality (Oval) and Connolly excluded volume (CSEV) are calculated with ChemBio3D Drew 10 software [26]. In all cases the structures of the compounds were pre-optimized with the Molecular Mechanics Force Field (MM+) procedure included in Hyperchem version 7.5 [27], and the resulting geometries were further submitted to the semi empirical method PM3 (Parametric Method-3) using the Fletcher–Reeves algorithm and a gradient norm limit of 0.009 kcal/Å.

The next formula, introduced in this paper, was used to calculate the novel descriptor (ND):

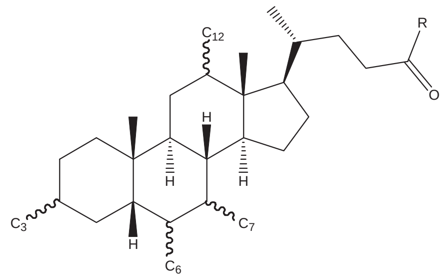
$$\mathrm{ND} = \frac{\frac{1}{n}\sum \angle_{\mathrm{O},a\mathrm{M}}}{\sum d_{\mathrm{O,O}} + \sum d_{\mathrm{O,ph}}}$$

where n represents the number of carbon atoms with hydroxyl and oxo groups in BAs steroid skeleton; $\angle_{\mathrm{O},a\mathrm{M}}$ represents the angle between β axial (a) methyl group and hydroxyl or oxo group in the proper Newmanns' projection formulas ($\angle_{\mathrm{O},a\mathrm{M}}$: $\alpha(a)$ OH = 180°; α(equatorial, e) OH or oxo = 120°; $\beta(e)$ OH or oxo = 60°), $d_{\mathrm{O,O}}$ represents distance between carbon atoms with hydroxyl or oxo groups from steroid skeleton (number of single connection is taken as a unit), while $d_{\mathrm{O,ph}}$ represents distance between carbon atoms with hydroxyl or oxo substituents and polar head of the side chain (as a unit the number of single connection as the shortest way in BA molecule graph is taken).

4. Results and discussions

4.1. Lipophilicity parameters: temperature dependence of retention coefficient

Linear equation (Eq. (1)) that connects BA retention coefficient (k) with temperature (t) fits very well with experimental data (Table 1). Linear model explains 96–99% of the whole variance (determination coefficient (R^2), Table 2). There is a good correlation (Eq. (2)) between the slope (b) and the intercept (a) in Eq. (1)



	Trivial name (abbreviation)	Substituent				
		C_3	C_6	C_7	C_{12}	R
1	lithocholic acid (**L**)	α-OH				OH
2	deoxycholic acid (**D**)	α-OH			α-OH	OH
3	chenodeoxycholic acid (**CD**)	α-OH		α-OH		OH
4	cholic acid (**C**)	α-OH		α-OH	α-OH	OH
5	ursodeoxycholic acid (**UD**)	α-OH		β-OH		OH
6	hyodeoxycholic acid (**HD**)	α-OH	α–OH			OH
7	hyocholic acid (**HC**)	α-OH	α-OH	α-OH		OH
8	Glycine conjugate of chenodeoxycholic acid (**G-CD**)	α-OH		α-OH		$NHCH_2COOH$
9	Taurine conjugate of deoxycholic acid (**T-D**)	α-OH			α-OH	$NH(CH_2)_2SO_3H$
10	Glycine conjugate of cholic acid (**G-C**)	α-OH		α-OH	α-OH	$NHCH_2COOH$
11	Glycine conjugate of deoxycholic acid (**G-D**)	α-OH			α-OH	$NHCH_2COOH$
12	Taurine conjugate of lithocholic acid (**T-L**)	α-OH				$NH(CH_2)_2SO_3H$
13	Taurine conjugate of chenodeoxycholic acid (**T-CD**)	α-OH		α-OH		$NH(CH_2)_2SO_3H$
14	Glycine conjugate of lithocholic acid (**G-L**)	α-OH				$NHCH_2COOH$
15	3α,7α-dihydroxy-12-oxocholanic acid (**12-OxC**)	α-OH		α-OH	=O	OH
16	3α,12α-dihydroxy-7-oxocholanic acid (**7-OxC**)	α-OH		=O	α-OH	OH
17	7,12-dioxolithocholic acid (**7,12-dOxC**)	α-OH		=O	=O	OH
18	12α-hydroxy-3,7-dioxocholanic acid (**3,7-dOxC**)	=O		=O	α-OH	OH
19	7α -hydroxy-3,12-dioxocholanic acid (**3,12-dOxC**)	=O		α-OH	=O	OH
20	3,7,12-trioxocholanic acid (dehydrocholic a.) (**3,7,12-tOxC**)	=O		=O	=O	OH
21	12-oxolithocholic acid (**12-OxD**)	α-OH			=O	OH
22	3,12-dioxocholanic acid (**3,12-dOxD**)	=O			=O	OH
23	7-oxolithocholic acid (**7-OxCD**)	α-OH		=O		OH
24	3,7-dioxocholanic acid (**3,7-dOxCD**)	=O		=O		OH
25	6-oxolithocholic acid (**6-OxHD**)	α-OH	=O			OH

Fig. 1. Investigated bile acids.

Table 1
Dependence of retention coefficient (k) of temperature (t) (°C).

Bile acids		20	25	30	35	40	45
1	L	29.32	25.25	20.73	16.86	14.18	11.53
2	D	15.09	13.28	10.87	9.34	7.71	6.43
3	CD	12.83	10.87	8.88	7.45	6.33	5.27
4	C	6.02	5.21	4.34	3.69	3.19	2.74
5	UD	3.24	2.74	2.36	2.05	1.78	1.50
6	HD	4.22	3.49	2.96	2.83	2.51	2.03
7	HC	3.92	3.40	2.83	2.41	2.07	1.77
8	G-CD	11.41	10.23	8.63	7.23	5.91	5.00
9	T-D	13.67	11.67	9.54	8.05	6.84	5.79
10	G-C	5.58	4.83	4.04	3.46	2.98	2.56
11	G-D	14.54	12.15	10.11	8.53	7.24	6.13
12	T-L	27.43	22.37	18.43	15.38	12.91	10.8
13	T-CD	11.53	9.52	7.97	6.73	5.72	4.71
14	G-C	28.69	23.23	19.17	16.05	13.45	10.9
15	12-OxC	2.41	2.08	1.82	1.77	1.54	1.33
16	7-OxC	2.18	1.89	1.66	1.62	1.43	1.23
17	7,12-dOxC	0.50	0.47	0.44	0.42	0.39	0.36
18	3,7-dOxC	0.56	0.51	0.46	0.45	0.41	0.35
19	3,12-dOxC	0.50	0.47	0.43	0.42	0.39	0.34
20	3,7,12-tOxC	0.40	0.35	0.33	0.32	0.28	0.25
21	12-OxD	5.91	5.03	4.35	4.23	3.65	3.13
22	3,12-dOxD	3.22	2.81	2.47	2.42	2.13	1.85
23	7-OxCD	4.72	4.02	3.51	3.40	2.93	2.53
24	3,7-dOxCD	2.73	2.38	2.12	2.06	1.83	1.59
25	6-OxHD	4.75	3.95	3.38	3.23	2.73	2.32

in a group of the examined BAs (Fig. 2) in Eq. (2).

$$a = 0.7638 + 58.8522b, \quad R = 0.9996 \tag{2}$$

This correlation indicates the possibility of forming homologic series in the examined BA molecule group [28]. In order to check if parameters a and b are good enough to describe bile acid's structural characteristics, grouping of bile acids in the a–b plain is performed. On a plot of the a–b data it can be seen that (Fig. 2) BAs form three groups. Lithocholic acid and its conjugates (BAs with one hydroxyl group in the steroid skeleton) belong to the first (I) group. In the second (II) group there are deoxycholic and chenodeoxycholic acid and their conjugates (BAs with two hydroxyl groups in the steroid skeleton), while the third group (III) contains BA oxo derivatives and next hydroxyl BA derivatives: ursodeoxycholic (UD, 5), hyocholic (HC, 7), and hyodeoxycholic (HD, 6) acid. For UD, HC and HD molecules it is characteristic that besides equatorial (e) C3 hydroxyl group (that has the same spatial orientation in each examined BA) they possess additional hydroxyl group with equatorial (e) orientation (HC and HD: C6 α(e) OH; UD: C7 β(e) OH). Oxo groups' oxygen atom has equatorial orientation as well (Fig. 3). According to that, common thing for bile acid of the third group is a presence of minimum two oxygen atoms (from the hydroxyl or oxo group) on the steroid skeleton whose orientations are switched for 60° (reference position is α, axial orientation) in a proper Newmanns' projection formulas (Fig. 3), i.e. they are switched toward angular methyl groups. The cholic (C, 4) and glycocholic acid (G-C, 10) belong to the third group (Fig. 2) as well. Those two molecules structurally (according to the orientation of the hydroxyl group) i.e. conformationally do not show common elements with other molecules from the group III. However, it is known that the bile

Table 2
The intercept (a) and slope (b) in linear regression $k = f(t)$, first principal component scores calculated from Table 1.

Bile acids		a	b	R	R^2	PC1	PC2
1	L	43.0501	0.7200	−0.9939	0.9879	5.37637	0.023150
2	D	21.8893	0.3518	−0.9953	0.9906	1.79015	−0.051081
3	CD	18.4260	0.3021	−0.9903	0.9807	1.05575	0.020646
4	C	8.4873	0.1319	−0.9913	0.9828	−0.66245	0.016292
5	UD	4.4884	0.0680	−0.9921	0.9843	−1.40668	0.031446
6	HD	5.6203	0.0804	−0.9766	0.9538	−1.11512	−0.034642
7	HC	5.5477	0.0864	−0.9918	0.9836	−1.23205	0.041543
8	G-CD	16.6862	0.2651	−0.9973	0.9946	0.85970	−0.024730
9	T-D	19.5503	0.3165	−0.9901	0.9802	1.30724	−0.001736
10	G-C	7.8569	0.1213	−0.9920	0.9841	−0.77443	0.014749
11	G-D	20.6286	0.3336	−0.9893	0.9786	1.51098	−0.012582
12	T-L	39.1628	0.6544	−0.9863	0.9728	4.65836	0.029036
13	T-CD	16.3808	0.2671	−0.9897	0.9796	0.70668	0.007651
14	G-C	41.1016	0.6926	−0.9874	0.9750	4.95053	0.026326
15	12-OxC	3.1340	0.0402	−0.9851	0.9704	−1.58139	−0.004996
16	7-OxC	2.8202	0.0353	−0.9837	0.9676	−1.64238	−0.002356
17	7,12-dOxC	0.7141	0.0078	−0.9884	0.9769	−2.12906	0.056420
18	3,7-dOxC	0.6100	0.0054	0.9987	0.9974	−2.11715	0.056756
19	3,12-dOxC	0.6237	0.0060	0.9875	0.9752	−2.12981	0.056122
20	3,7,12-tOxC	0.5042	0.0056	−0.9868	0.9738	−2.17086	0.066570
21	12-OxD	7.7659	0.1040	−0.9824	0.9652	−0.57871	−0.116428
22	3,12-dOxD	4.1490	0.0511	−0.9839	0.9681	−1.32315	−0.048688
23	7-OxCD	6.1746	0.0817	−0.9831	0.9665	−0.91824	−0.078586
24	3,7-dOxCD	3.4944	0.0423	0.9851	0.9704	−1.46618	−0.031508
25	6-OxHD	6.3614	0.0913	−0.9832	0.9667	−0.96809	−0.039373

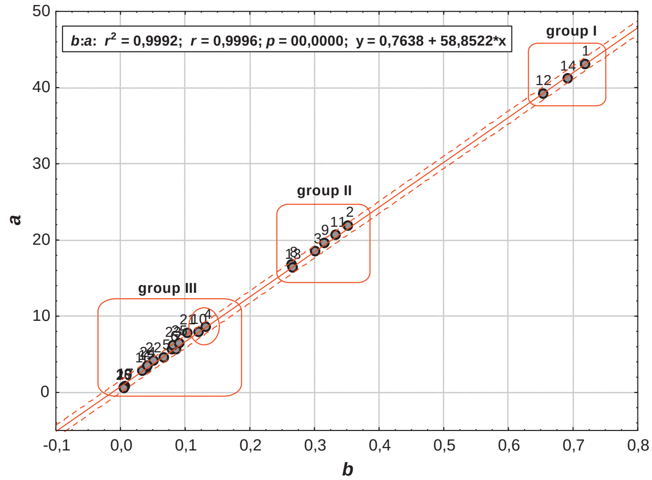


Fig. 2. Grouping of bile acids on the scatter plot of *a*–*b*.

acids' lipophilicity can be expressed using principal component analysis (PCA) method on retention coefficients obtained by means of reverse phase thin layer chromatography (RP TLC) [20]. For deriving parameters of lipophilicity PCA method was implemented on the data from Table 1 ($k = f(t)$). First two principal components (PC1 and PC2) are found to be significant i.e. their sum (PC1 + PC2) explains 100% of the variance (Table 1). On the scatter plot of the scores PC1–PC2 three congeneric groups are formed (Fig. 4) with identical elements (molecules) as on the plot *a*–*b*. However, if the linear regression is used between values of PC1 and PC2 in group III then the Cooks' distance (Fig. 4) in the linear model is highest for molecules C; 4 and G-C; 10. Total variance square on the scatter plot PC1–PC2 for linear congeneric group III is $s_0^2 = 0.0285$ (group-model is determined in relation to the straight regression line, Fig. 4), residual variance square for the molecules C and G-C are $s_i^2 = 0.1223$ and $s_i^2 = 0.0909$. Since $s_i^2 > s_0^2$ [29] it can be concluded that bile acids C and G-C are outliers and do not belong to the group III. Thus, when PC1 and PC2 scores are used for grouping bile acids, molecule stereochemistry (steroid core) is more clearly expressed. Bile acid structures are shown to be better explained using first two principal component values (PC1 and PC2) than by regression coefficients (Eq. (1)). This is because, depending on bile acid itself, Eq. (1) explains 96–99% of the variance in a temperature dependence of retention coefficient while PC1 + PC2 explain 100% of the variance in Table 1.

Relationship between experimentally obtained descriptors (*a*, *b*, PC1 and PC2, related to temperature dependence on retention coefficient) and *in silico* descriptors are also calculated using PCA. Fig. 5 presents loadings of the principal components PC1 and PC2 calculated from the data matrix. Matrix columns are *in silico* molecular descriptors, new molecular descriptor ND, and experimentally obtained parameters of lipophilicity: regression coefficients *a*, *b*; PC1 and PC2 scores calculated from the temperature dependence of retention coefficient (Table 1). It can be concluded (Fig. 5) that *a*, *b* and PC1 are strongly correlated, and that most of *in silico* molecular descriptors are correlated to each other and orthogonal with *a*, *b*, PC1 and ND. Among *in silico* descriptors, ND shows the highest correlation with experimental parameters *a*, *b*, PC1 which indicates the importance of the spatial orientation of hydroxyl and oxo groups of BAs; steroid skeleton in a temperature dependence on retention coefficient.

4.2. QSAR model for partition coefficient of nitrazepam

Partition coefficients (K_p) of nitrazepam in BA micelles are measured on temperatures of 25 °C and 37 °C (Table 3). On both temperatures increase in the number of OH group in the steroid skeleton i.e. substitution of oxo with hydroxyl group leads to the decrease of the K_p. It is the consequence of the decrease in hydrophobicity of the BAs' steroid skeleton especially when oxo groups whose oxygen atom is directed toward β (hydrophobic) side of the steroid skeleton are introduced (Fig. 3). Lower hydrophobicity of the BAs steroid skeleton causes the lower hydrophobicity of the internal micellar cage as well as the lower potential for accepting hydrophobic guest (nitrazepam).

For deriving QSAR model besides *in silico* molecular descriptors, parameters of lipophilicity are used (regression coefficient of the linear dependence of partition coefficient of temperature: *a*, *b* and principal components PC1 and PC2 (Table 2) calculated from the data matrix $k = f(t)$ (Table 1). By implementing multiple linear regressions on the calibration set (17 BA molecules (Table 3)) and eliminating predictive variables by forward stepwise method next equations are obtained for the K_p:

$$\text{KP} \quad (25 \quad {}^\circ\text{C}) = 1778.73 + 348.42 \quad \text{PC1} - 566.36 \quad {}^3X + 18.45 \quad V \tag{3}$$

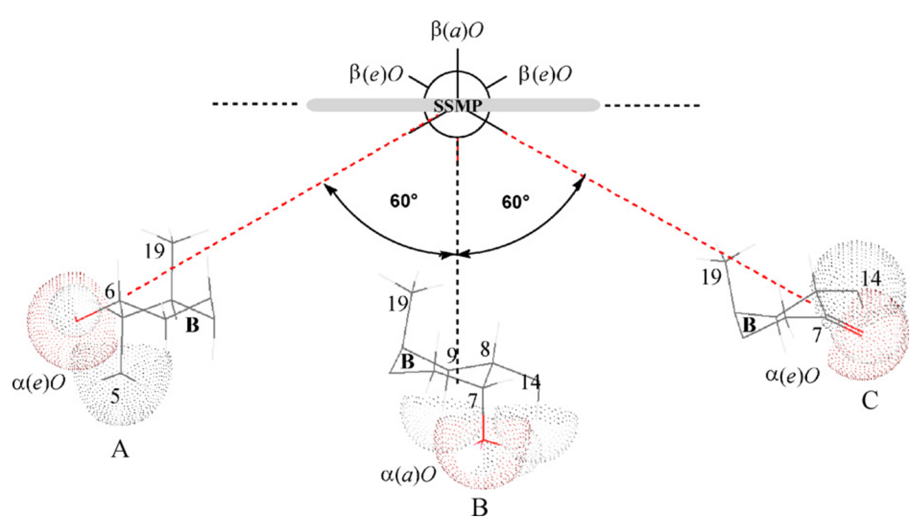


Fig. 3. Spatial orientation of axial OH group (A) cholic acid, equatorial OH group (B) hyodeoxycholic acid and equatorial oxo group (C) 7-oxolithocholic in conformational formulas of the B ring of the steroid core and in Newmanns' projection formula (SSMP = steroid skeleton mean plain).

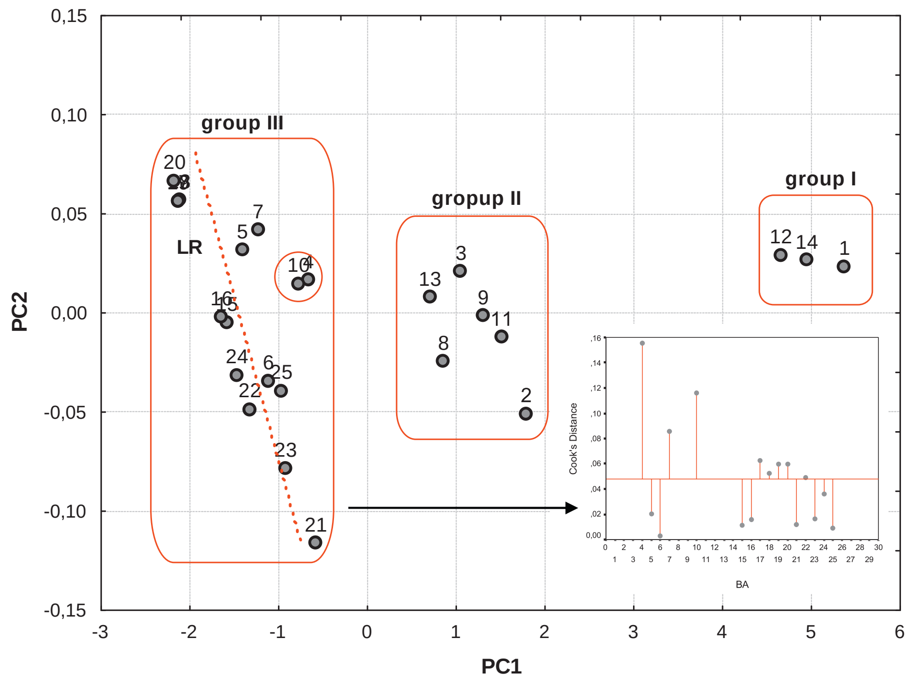


Fig. 4. The scatter plot of scores on the plane described by PC1 and PC2, grouping of bile acids (PC1 (99.99715%)–PC2 (0.05287%)), principal components are calculated from data matrix $t = f(k)$.

$n = 17$; $R = 0.994$; $F = 350$; $s = 92.08$

$$\text{KP}\ (37\ ^\circ\text{C}) = -3142.80 + 274.53\ \ \text{PC1} - 27.36\ \ M + 375.89\ \ \text{MP} \tag{4}$$

$n = 17$; $R = 0.994$; $F = 389$; $s = 86.62$

Derived linear models (Eqs. (3) and (4)) suggest the importance of BA steroid skeleton hydrophobicity on the inclusion of nitrazepam in micelles. In both equations the first principal component, calculated from the BA's retention coefficients, which is highly correlated to the hydrophilic–hydrophobic balance of BA molecules takes part. On 25 °C in Eq. (3) connectivity of the third-order (^3X) and volume of the molecule (V) take part. Those are molecular descriptors that explain the shape of a molecule. BA's geometry is very important in micelle formation especially its internal cage whose hydrophobicity, size and shape determine level of guest molecule intake (nitrazepam). With temperature rise (Eq. (4)), besides molecular descriptors of the size of the molecule i.e. mass molecule (M), molecular polarizability (MP) is introduced in linear regression equation. It is known that, on higher temperatures micelle formation (formation of the mixed micelles) is not caused by entropic reasons as on the room temperature but by enthalpy force [16]. According to that, if the molecular polarizability (MP) is higher, induced dipoles are formed more easily i.e. changes of energy during formation of the secondary chemical bonds inside the micelles are higher.

If the internal validation is implemented on the calibration set (cross validation leave-one-out, LOO) the regression coefficient of the internal validation are $q^2_{CV} = 0.893$ (Eq. (3)) and $q^2_{CV} = 0.837$ (Eq. (4)) [30]. Derived linear models are significant according to the K_p value of nitrazepam predictive capability (models are significant in a sense of predictivity if $q^2_{CV} > 0.3$ [3]). In order to implement independent (external) validation [30] certain molecules are left out from each congeneric BAs group (Fig. 4) and test set is formed. From group II CD(3) and G-C(10), from group III HD (6), 12-OxD (21) and 3,12-dOxD(22) are left out. In both models (Eqs. (3) and (4)) for each molecule from the test group standard error of prediction from cross-validation, SEP_{CV} [29] (Table 4) is lower than standard deviation of experimental data (Table 3).

If experimentally obtained parameters are not used as predictive variables the next regression equations are derived for K_p:

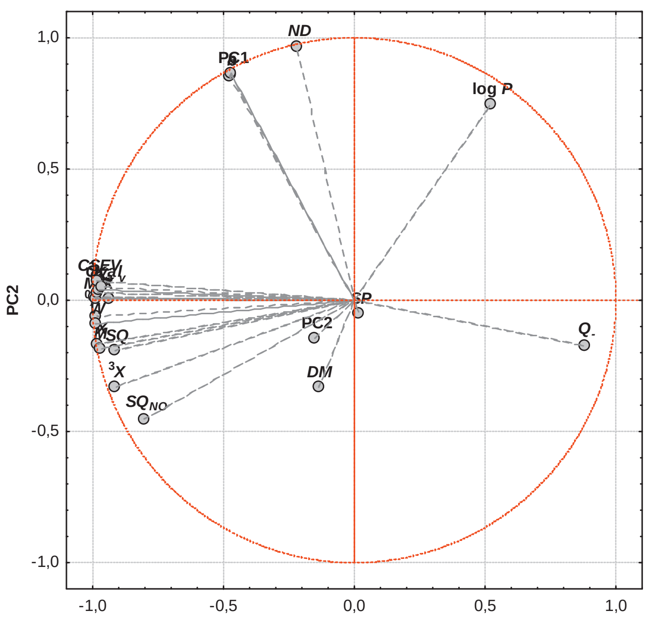


Fig. 5. Principal component's loadings obtained from the predictive variables: *in silico* descriptors and experimental parameters: *a*, *b*, PC1 (calculated from data matrix $t = f(k)$) and PC2 (calculated from data matrix $t = f(k)$); (PC1 (63.679%) and PC2 (18.875%)).

$$\text{KP}\ (25\ ^\circ\text{C}) = -6992.58 + 342.32\ \ \text{ND} - 33.87\ \ V + 15.48\ \ M \tag{5}$$

Table 3
Partition coefficient (K_p) of nitrazepam values.

Bile acids		Set	Experimental		Estimated by MLR		Estimated by PLS	
No	Aberration		t = 25 °C	t = 25 °C	t = 36 °C	t = 25 °C	t = 25 °C	t = 36 °C
1	L	Pred			4589	4222	4522	4410
2	D	Cal	3007 ± 64	2923 ± 47	3154	3042	3067	2991
3	CD	Test	2935 ± 42	2879 ± 35	3005	2841	2936	2872
4	C	Cal	2260 ± 43	2232 ± 49	2232	2169	2165	2132
5	UD	Cal	2082 ± 31	2048 ± 39	2154	2166	2085	2052
6	HD	Test	2197 ± 35	2171 ± 32	2259	2246	2367	2331
7	HC	Cal	1998 ± 40	1970 ± 38	1996	2013	1966	1938
8	G-CD	Cal	3271 ± 52	3205 ± 46	3174	3143	3311	3250
9	T-D	Cal	3475 ± 57	3412 ± 40	3507	3418	3442	3365
10	G-C	Test	2351 ± 37	2289 ± 53	2412	2498	2442	2398
11	G-D	Cal	3401 ± 45	3322 ± 44	3284	3322	3366	3291
12	T-L	Pred			4381	4184	4428	4323
13	T-CD	Cal	3328 ± 59	3251 ± 57	3405	3254	3357	3290
14	G-L	Pred			4662	4462	4573	4462
15	12-OxC	Cal	1705 ± 28	1689 ± 38	1782	1766	1755	1728
16	7-OxC	Cal	1695 ± 26	1680 ± 32	1763	1749	1759	1738
17	7,12-dOxC	Cal	1452 ± 24	1431 ± 22	1494	1464	1507	1485
18	3,7-dOxC	Cal	1443 ± 27	1424 ± 19	1466	1467	1487	1471
19	3,12-dOxC	Cal	1456 ± 33	1421 ± 23	1471	1464	1436	1416
20	3,7,12-tOxC	Cal	1437 ± 24	1430 ± 29	1352	1301	1360	1351
21	12-OxD	Test	2079 ± 45	2054 ± 39	2212	2070	2234	2186
22	3,12-dOxD	Test	1902 ± 38	1945 ± 33	1855	1894	1965	1944
23	7-OxCD	Cal	2357 ± 41	2316 ± 37	2195	2157	2239	2202
24	3,7-dOxCD	Cal	1893 ± 37	1870 ± 25	1907	1854	1951	1935
25	6-OxHD	Cal	2251 ± 40	2232 ± 32	2174	2143	2249	2220

Cal = calibration set (cross validation), test set (internal validation), pred = predicted values.

Table 4
Values of standard error of prediction from cross-validation SEP_{CV}.

Bile acids		MLR		PLC	
		t = 25 °C	t = 36 °C	t = 25 °C	t = 36 °C
3	CD	16.97	9.21	0.24	0.64
6	HD	11.88	17.46	41.23	38.79
10	G-C	35.65	50.68	35.65	26.41
21	12-OxD	41.3	8.87	37.58	32.00
22	3,12-dOxD	11.39	12.37	15.26	0.24

$n = 17$; $R = 0.970$; $F = 139$; $s = 144.71$

$$\mathrm{KP}\ (37\ {}^\circ\mathrm{C}) = -6100.54 + 316.95\ \mathrm{ND} + 34.16\ V - 16.22\ M \quad (6)$$

$n = 17$; $R = 0.984$; $F = 145$; $s = 145.00$

The importance of the descriptor ND can be seen according to linear models above. This demonstrates the importance of the certain parts of the BA molecule as are distance between OH and oxo groups and their distance from planar head of a C17 side chain i.e. their penetration to β side of the steroid skeleton (toward angular groups).

Thus, ND determines hydrophobicity change of the BA molecule which can be seen from the correlation i.e. its position in relation to experimentally obtained variables (PC1, a, b) on the scatter plot of scores (Fig. 5). ND descriptor can be considered as a local descriptor [3] because it does not include information about distance and connectivity of other atoms of the BA molecule which is not the case for Wiener index or connectivity descriptors. Internal validation regression coefficients for models (Eqs. (5) and (6)): $q^2_{CV} = 0.427$ (Eq. (5)), $q^2_{CV} = 0.439$ (Eq. (6)) are lower than for models in which experimentally obtained data are included as predictive variables. Nevertheless, models (Eqs. (5) and (6)) are still significant in the sense of predictive capability. Higher predictivity of regression equations (Eqs. (3) and (4)) suggests importance of the temperature dependence of BAs retention coefficients (through predictive variable PC1) i.e. of dependence $k = f(t)$ (Table 1) that is related to hydrophobicity of the examined molecules.

PLC cross validation method is done and predictive residual sums of squares ($PRESS_{CV}$) are calculated after addition of each factor. $PRESS_{CV}$ [29] does not change after three factors added to

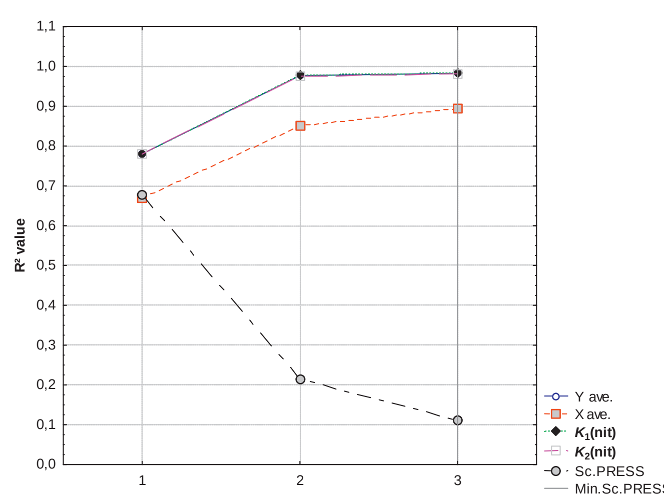


Fig. 6. Determination of the number of significant factors in PLC model.

Table 5
Scaled regression coefficient values obtained using PLC method.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
	a	b	PC1	PC2	$^1\chi$	$^0\chi^v$	$^1\chi^v$	log P	V	W	$^3\chi$	M	SP	DM	MP	Q_+	$^3K_\alpha$	SQ	SQ_{NO}	Q_-	CSEV	Oval	ND
K_p 25 °C	0.16	0.16	0.16	−0.11	0.02	0.03	0.02	0.03	0.05	0.03	−0.04	0.00	−0.03	−0.00	0.04	0.05	0.06	0.00	−0.06	−0.03	0.05	0.07	0.14
K_p 37 °C	0.16	0.16	0.16	−0.11	0.02	0.03	0.02	0.03	0.05	0.03	−0.04	0.00	−0.03	−0.00	0.04	0.05	0.06	0.00	−0.06	−0.03	0.05	0.07	0.14

Table 6
Statistical parameters to evaluate linear relation between observed and predicted partition coefficient (K_p) of nitrazepam considering both multiple regression methods.

Statistical parameter	MLR		PLC	
	Eq. (3) t = 25 °C	Eq. (4) t = 36 °C	Eq. (5) t = 25 °C	Eq. (6) t = 36 °C
R	0.99278	0.9914	0.99435	0.99422
F	1370	1153	1754	1715
p	0.0000	0.0000	0.0000	0.0000
s	83.26	87.66	73.71	72.09
a_0 (intercept)	3.59	5.44	−22.01	−12.80
a_1 (slope)	0.9928	0.99258	1.00	0.99778

PLC models (Fig. 6). For deriving model for K_p of nitrazepam using PLC, BA set is divided in to calibration and test set in the same way as for MLR method (Table 3). On the basis of calibration set that consists of 17 BA molecules in model of nitrazepams' K_p (standardized equation, Table 5) the most important predictive variables (on both temperatures) are: $a = b$ = PC1 > ND > PC2, i.e. experimentally obtained parameters and ND descriptor i.e. descriptors whose vectors on the scatter plot scores form the lowest angle (a, b, PC1 and ND; Fig. 5). PC2 is shown to be a significant variable of PLC model as well.

This suggests that principal components with small eigen values contain important information about structural differences between different molecules. Internal validation regression coefficients of the PLC model ($q^2_{CV} = 0.903$) are higher than for MLR model. Standard errors of prediction from cross-validation, SEP_{CV} (Table 4) as for MLR model have the smaller values than experimental values K_p standard deviations. Statistical parameters of the linear regression between predicted (MLR and PLC) and experimentally obtained K_p values of nitrazepam for calibration and test set of BA molecules are shown in Table 6. Statistical parameters are slightly better for PLC than for MLR models.

Experimental values of K_p of nitrazepam are not determined for the BAs of the first group (lithocholic acid and its derivatives) since these BAs do not form micelles. However, on basis of regression models MLR and PLC, predicted values of K_p match (Table 3). Predicted K_p values for the first group of BAs have the highest values among all examined BAs. Those BAs are the most hydrophobic. Practical usage of predicted K_p values is that those BAs can be most efficiently used for microcapsulation of nitrazepam with addition of some cosurfactant (for example. sodium dodecyl sulfate) [3], and in such a binary system micelles with very hydrophobic cage would form.

5. Conclusions

Experiment used for obtaining $k = f(t)$ is fast and simple. Thus for the short period of time experimental parameters for a large number of compounds can be acquired, which is desirable in QSAR research. Temperature (t) dependence of retention coefficient (k) is linear. Parameters of linear dependence (the a, intercept, and the b, slope) strongly correlate to a BA structure. In QSAR models $t = f(k)$ is the best represented with principal component scores calculated from the temperature dependence k matrix. In this work linear model for partition coefficient of nitrazepam in BA micelles on temperatures of 25 °C and 37 °C using MLR method are derived. On both temperatures regression models are derived from experimentally predictive variable PC1 ($t = f(k)$) and *in silico* descriptors of the shape of a molecule while on temperature 37 °C molecular polarization is included. This clearly indicates the fact that the incorporation mechanism of nitrazepam in BA micelles changes on higher temperatures. Derived regression models for K_p of nitrazepam using

PLC method indicate the importance of experimentally predictive variables. PLC models for K_p of nitrazepam have slightly higher predictive capability than derived MLR models. Correlation between experimental parameters PC1, *a* i *b* and *in silico* descriptor ND points to the importance of OH and oxo groups' spatial orientation on BAs' steroid skeleton in a temperature dependence of retention coefficient as well as for describing nitrazepam partition coefficient.

Acknowledgement

The Ministry of Science and Technological Development of the Republic of Serbia (Project No. 172021) supported this study.

References

[1] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S.P. De Yong, J. Levi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics, Elsevier, Amsterdam, 1997.
[2] F.D. King, Medicinal Chemistry Principles and Practice, R.S.C., Cambridge, 2002.
[3] W. Camile, The Practice of Medicinal Chemistry, Academic Press, Oxford, 2003.
[4] R. Kaliszan, Quantitative Structure–Chromatographic Retention Relationships, Wiley, New York, 1987.
[5] G. Ioele, M. De Luca, F. Oliverio, G. Ragno, Talanta 79 (2009) 1418–1424.
[6] C. Sarbu, C. Onisor, M. Poša, S. Kevrešan, K. Kuhajda, Talanta 75 (2008) 651–657.
[7] M.P. Freitas, J.A. Martins, Talanta 67 (2005) 182–186.
[8] K. Valko, J. Chromatogr. A 1037 (2004) 299–310.
[9] M. Poša, K. Kuhajda, Steroids 75 (2010) 424–431.
[10] A. Roda, A.F. Hofmann, K.J. Mysels, J. Biol. Chem. 258 (1983) 6362–6370.
[11] M. Calabresi, P. Andreozzi, C. La Mesa, Molecules 12 (2007) 1731–1754.
[12] M. Mikov, J.P. Fawcett (Eds.), Bile Acids, Medishet Publisher, Geneva, 2007.
[13] C. Bowe, L. Mokhtarzadeh, P. Venkatesen, S. Babu, H. Axelrod, M.J. Sofia, R. Kakarla, T.Y. Chan, J.S. Kim, H.J. Lee, G.L. Amidon, S.Y. Choe, S. Walker, D. Kahne, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 12218–122223.
[14] G.S. Gordon, A.C. Moses, R.D. Silver, J.R. Flier, M.C. Carey, Proc. Natl. Acad. Sci. U.S.A. 82 (1985) 7419–7423.
[15] M. Poša, V. Guzsvány, J. Csanádi, S. Kevrešan, K. Kuhajda, Eur. J. Pharm. Sci. 34 (2008) 281–292.
[16] M. Poša, S. Kevrešan, M. Mikov, V. Ćirin-Novta, K. Kuhajda, Colloid Surf. Biointerface 64 (2008) 151–161.
[17] M. Poša, V. Guzsvány, J. Csanádi, Colloid Surf. Biointerface 74 (2009) 84–90.
[18] M. Atanacković, M. Poša, H. Heinle, Lj. Gojković-Bukarica, J. Cvejić, Colloid Surf. Biointerface 72 (2009) 148–154.
[19] B. De Castro, P. Gameiro, C. Guimarães, J.L.F.C. Lima, S. Reis, J. Pharm. Biomed. Anal. 24 (2001) 595–602.
[20] C. Sarbu, K. Kuhajda, S. Kevrešan, J. Chromatogr. A 917 (2001) 361–366.
[21] M. Poša, A. Pilipović, M. Lalić, J. Popović, Acta Chim. Slov. 57 (4) (2010) 828–835.
[22] M. Poša, Steroids, doi:10.1016/j.steroids.2010.09.003.
[23] M. Poša, Z. Farkaš, Collect. Czech. Chem. Commun. 75 (8) (2010) 767–784.
[24] A. Roda, A. Minutello, M.A. Angellotti, A. Finit, J. Lipid Res. 31 (1990) 1433–1443.
[25] http://www.statsoft.com.
[26] http://www.cambridgesoft.com.
[27] http://www.organic-chemistry.org/prog/peo/osiris property explorer.
[28] T. Cserhati, J. Biochem. Biophys. Methods 27 (1993) 133.
[29] M. Otto, Chemometrics, Wiley/VCH, Weinheim, 2007.
[30] P. Gramatica, QSAR Comb. Sci. 26 (2007) 694–701.