Journal homepage: www.elsevier.com/locate/jpba

Journal of Pharmaceutical and Biomedical Analysis**,** Volume **127**, 5 August **2016,** Pages **81-93.**

Special issue: Current trends in bioanalysis - In Honour of Professor Roman Kaliszan

Available: http://dx.doi.org/10.1016/j.jpba.2016.04.001

Multivariate assessment of lipophilicity scales–computational and reversed phase thin-layer chromatographic indices

Filip Andrić[1], Dávid Bajusz[2], Anita Rácz[3,4], Sandra Šegan[5], Károly Héberger[3,*]

[1] *Faculty of Chemistry, University of Belgrade, Studentski trg 12-16, 11000 Belgrade, Serbia*

[2] *Medicinal Chemistry Research Group, Research Centre for Natural Sciences, Hungarian Academy of Sciences, H-1117 Budapest XI., Magyar Tudósok krt. 2, Hungary*

[3] *Plasma Chemistry Research Group, Research Centre for Natural Sciences, Hungarian Academy of Sciences, H-1117 Budapest XI., Magyar Tudósok krt. 2, Hungary*

[4] *Department of Applied Chemistry, Faculty of Food Science, Szent István University of Budapest, H-1118 Budapest XI., Villányi út 29-43, Hungary*

[5] *Institute of Chemistry, Technology and Metallurgy, Njegoševa 12, 11000 Belgrade, Serbia*

[*]To whom the correspondence should be addressed:
Károly Héberger
Phone: +36 1 38 26 509
E-mail: heberger.karoly@ttk.mta.hu

Abstract

Needs for fast, yet reliable means of assessing the lipophilicities of diverse compounds resulted in the development of various *in silico* and chromatographic approaches that are faster, cheaper, and greener compared to the traditional shake-flask method. However, at present no accepted "standard" approach exists for their comparison and selection of the most appropriate one(s). This is of utmost importance when it comes to the development of new lipophilicity indices, or the assessment of the lipophilicity of newly synthesized compounds. In this study, 50 well-known, diverse compounds of significant pharmaceutical and environmental importance have been selected and examined. Octanol-water partition coefficients have been measured with the shake-flask method for most of them. Their retentions have been studied in typical reversed thin-layer chromatographic systems, involving the most frequently employed stationary phases (octadecyl- and cyano-modified silica), and acetonitrile and methanol as mobile phase constituents. Twelve computationally estimated log*P*-s and twenty chromatographic indices together with the shake-flask octanol-water partition coefficient have been investigated with classical chemometric approaches – such as principal component analysis (PCA), hierarchical cluster analysis (HCA), Pearson's and Spearman's correlation matrices, as well as novel non-parametric methods: sum of ranking differences (SRD) and generalized pairwise correlation method (GPCM). Novel SRD and GPCM methods have been introduced based on the Comparisons with One VAriable (lipophilicity metric) at a Time (COVAT). For the visualization of COVAT results, a heatmap format was introduced. Analysis of variance (ANOVA) was applied to reveal the dominant factors between computational log*P*s and various chromatographic measures. In consensus-based comparisons, the shake-flask method performed the best, closely followed by computational estimates, while the chromatographic estimates often overlap with *in silico* assessments, mostly with methods involving octadecyl-modified silica stationary phases. The ones that employ cyano-modified silica perform generally worse. The introduction of alternative coloring schemes for the covariance matrices and SRD/GPCM heatmaps enables the discovery of intrinsic relationships among lipophilicity scales and the selection of best/worst measures. Closest to the recommended logK_{OW} values are Clog*P* and the first principal component scores obtained on octadecyl-silica stationary phase in combination with methanol-water mobile phase, while the usage of slopes derived from Soczewinski-Matyisik equation should be avoided.

Keywords: Lipophilicity, Reversed-phase thin-layer chromatography, Benzodiazepines, Polyaromatic hydrocarbons, Phenols, Sum of ranking differences - SRD, Generalized pairwise correlation method - GPCM, Comparison with one variable at a time – COVAT, Heatmap

List of abbreviations

ANOVA – Analysis of Variance, C18 – Octadecyl silica, CEPW – Conditional Exact test with Probability Weighted (ranking), CN – Cyanopropyl-modified silica, COVAT – Comparison with One Variable at a Time, CRRN – Comparison of Ranks with Random Numbers, GPCM – Generalized Pairwise Correlation Method, HCA – Hierarchical Cluster Analysis, HILIC – Hydrophilic Interaction Liquid Chromatography, HPLC – High Performance Liquid Chromatography, IAM – Immobilized Artificial Membrane Chromatography, LSER – Linear Solvation Energy Relationships, MEKC – Micellar Electrokinetic Chromatography, MLC – Micellar Liquid Chromatography, PC – Principal Component, PCA – Principal Component Analysis, Rg – Range scaling, Rk – Rank transformation, SRD – Sum of (absolute) Ranking Differences, St – Standardized (autoscaled),TLC – Thin-layer Chromatography

1 Introduction

Throughout the last century lipophilicity evolved into an essential physicochemical parameter that is used in pharmaceutical and environmental sciences abundantly. It is related to the distribution of compounds in the environment and biota, to bioavailability and bioconcentration in the food chain, as well as to the transport in the soil-sediment-water compartments [**1**]. It is a crucial factor influencing passive transport trough biological membranes such as the blood-brain or the gastrointestinal barriers [**2,3**]. Lipophilicity has a high impact on protein binding, drug-receptor interactions, which consequentially alters the desired physiological response, as well as drug-related toxicity and adverse effects [**4,5**].

Nevertheless, since the first works of Meyer and Overton [**6,7**], lipophilicity has been tailored to suit our practical needs, while its strict definition remains ambiguous. In that sense, according to the International Union for Pure and Applied Chemistry (IUPAC), lipophilicity represents the affinity of a molecule or a moiety for a lipophilic environment [**8**]. It is still not clear what a "lipophilic environment" actually is, and how it should be modelled. Such a vague definition of the lipophilicity itself might be one of the reasons that create additional space for development of various lipophilicity measures and numerous experimental approaches for its measurement and estimation. In order to put some constraints the IUPAC gives some recommendations how lipophilicity should be or could be measured [**8**]. The traditionally adopted shake-flask method – based on the distribution between octanol and water (commonly denoted as log*P*, but more frequently replaced with logK_{OW} in contemporary literature) – is time and reagent consuming, experimentally demanding, tedious, and mostly applicable to pure compounds that have partition coefficients in the range of -3 to 4.5 log units (some modifications of the shake flask method are applicable for compounds with logK_{OW} > 4.5). In order to overcome these difficulties many chromatographic methods have been developed, and some of them have been adopted as standard methods, parts of OECD guidelines (Organization for Economic Cooperation and Development), such as Test No. 117, HPLC method [**9**]. Aside from very specific applications of chromatographic approaches that tend to mimic biosystems such as micellar liquid chromatography (MLC) [**10-15**], immobilized artificial membrane chromatography (IAM) [**16,17**], immobilized proteins *etc*. [**18**], the mainstream methods in the determination of lipophilicity are still based on typical reversed-phase chromatography including a variety of chemically bonded stationary phases [**19-22**], where octyl-, octadecyl-, and cyanopropyl-modified silica beds are the most frequently used in combination with a polar mobile phase (usually binary mixtures of miscible organic solvents and water) [**23-25**].

Both high-performance liquid chromatography (HPLC) and thin-layer chromatography (TLC) produce a high number of chromatographic lipophilicity indices. However, TLC has a significant advantage over HPLC because of its simplicity, significantly reduced costs, short analysis time, low consumption of solvents and reagents, and its ability to simultaneously handle dozens of samples.

Several lipophilicity measures stem from TLC experiments. The intercept (R_M^0) and the slope (b) of the linear dependence of the retention on the volume fraction of the organic component of the mobile phase (φ), proposed by Soczewinski and Matyisik [**26**] (Eq. 1), have been introduced among them first. The R_M value is defined according to the Eq. 2.

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

$$R_M = \log\left(\frac{1}{R_F} - 1\right) \quad (2)$$

where R_F is the retardation factor, *i.e.*, the ratio of the distance of a solute target zone and the solvent front.

The parameter b can be related to the specific hydrophobic surface area of the solute [**27**] and the surface tension of the mobile phase [**28**], while the intercept describes partitioning between pure water and the non-polar, hydrophobic stationary phase.

In addition, the concentration of the organic solvent in the mobile phase resulting in equal distribution of a solute among the stationary and mobile phase, C_0, was introduced by Bieganowska *et al*. [**29**], and is frequently used. It is defined as the intercept (R_M^0) and the slope (b):

$$C_0 = -\frac{R_M^0}{b} \quad (3)$$

Alongside the extrapolated chromatographic lipophilicity measures, the ones based on primary retention data are also used as *e.g*. the first principal component scores ($PC1/R_M$) derived from principal component analysis (PCA) of multivariate retention data [**30, 31**], and arithmetic means of R_M values, more frequently denoted as mR_M [**23-25**].

Besides the experimental methods, computational approaches for the prediction of log*P* values are extensively used. Their main advantage is that they simply do not require experimental measurements. They can be classified in two large families: substructure-based and property-based methods. Substructure-based methods decompose the molecular structure into smaller fragments (or even down to the level of single atoms). Depending on the algorithm used, each fragment is then associated with a particular log*P* contribution. The final log*P* value of the unknown compound is obtained by a summation over all fragment contributions, and using correction factors, where necessary [**32**]. Examples of fragmentation/group contribution based methods are: ClogP, AClogP, ALOGP, miLogP, KOWWIN, XLOGP2, XLOGP3 [**33-38**]. Property-based methods, on the other hand, consider the molecule as an undivided entity [**32**]. Calculation of log*P* is based on quantitative structure - property relationship (QSPR) models using physicochemical parameters such as the case with the Linear Solvation Energy Relationships (LSER) approach [**39**], or from molecular descriptors obtained from 3D representations (*e.g*. COSMOFrag) [**40**], or simple 1D topological, and electrotopological indices (MLOGP, ALOGPs) [**41,42**]. Nevertheless, both property- and substructure-based methods are accompanied by estimation errors that reach orders of magnitude for the same molecule as compared to each other. Computational methods that are used in the present work are enlisted in section 2.3.

When it comes to the selection of an appropriate approach to lipophilicity assessment there are several problems, errors, and misconceptions, especially in the case of newly synthesized compounds or novel lipophilicity indices. If there is no possibility to obtain octanol-water partitioning data, chromatographic and computational estimates are most frequently used to estimate lipophilicity. However, no systematic or widely accepted approach exists for the selection of appropriate lipophilicity measures. Many procedures use similarities among computationally estimated values and experimentally derived lipophilicity indices as a criterion to select the best one. Such similarities are most often obtained from hierarchical clustering (HCA) [**43,44**], principal component analysis [**21,23,25,45**], or simple correlations based on parametric statistics such as Pearson's correlation coefficient [**24,25,44**]. The last one is applicable only if the data is normally distributed, which is often not the case. PCA and HCA do not provide information about statistical significance of such similarities, while the use of correlation measures most often lead to selection of the most correlated pairs, neglecting the rest of the statistically significant ones.

The aim of the present work was to rank and group lipophilicity measures from the typical reversed-phase thin-layer chromatographic data, to find the most similar and dissimilar ones, to suggest suitable substitutes for the octanol-water partition coefficient as a current golden standard in lipophilicity assessment, and to give recommendations for the proper use of statistical techniques in the selection of lipophilicity scales. The present work is a continuation and extension of our previous research [**46,47**].

2 Materials and methods

2.1 *Compound set selection*

In total 50 compounds (Table 1) of low molecular mass (94.12-321.18 g mol^{-1}) of various chemical structures, molecular sizes and shapes have been selected in a way that they cover a relatively broad range of the recommended values of experimentally determined octanol-water partition coefficients ($0.62 < \log K_{OW} < 6.50$) and

their various abilities to interact with stationary and mobile phases selected according to Abraham's solvatochromic parameters:

$$0.00 < A < 0.94,\ 0.15 < B < 1.63,\ 0.79 < S < 2.49,\ 0.80 < E < 3.43,\ 0.7751 < V < 2.1924 \quad (4)$$

where *A*, *B*, *S*, *E*, and *V* are hydrogen bond donating ability, hydrogen bond accepting ability, dipolarity-polarizability, molar refractivity in excess expressed in units ($cm^3\ mol^{-1}$)/10 and McGowan's molar molecular volume (V has a unit of ($cm^3\ mol^{-1}$)/100), respectively. The full list of compounds accompanied with Abraham's solvatochromic parameters, molecular masses, p*K*a values and water solubilities is given in the supplementary material, Table S1. Special care was taken of the selection of pharmaceutically important compounds (9 benzodiazepine derivatives) and those with environmentally related issues (12 phenols, 10 polyaromatic hydrocarbons (PAH), 4 triazine herbicides, 5 aromatic amines, 6 aromatic alcohols, aldehydes and ketones, 3 aromatic acids and esters). Under experimental conditions all compounds, with the exception of 4-nitrophenol, 2,4,6-trichlorphenol, and 4-aminobenzoic acid, have been in their neutral (non-ionized) form (ionization degree < 1 %).

2.2 *Chromatographic experiments*

Two most commonly used stationary phases have been selected: octadecyl- and cyanopropyl-modified silica layers coated on aluminum sheets and glass, respectively, (Art. Nos. 5559 and 16464 respectively, Merck Darmstadt, Germany). The plates of the 10 × 10 cm size were used. Two typical organic modifiers: methanol and acetonitrile have been chosen to prepare binary mixtures with water. The mobile phase composition was tuned in a way that allows precise and reliable measurement of retention and good fitting to the Soczewinski-Matyisik linear equation (Eq. 1). The fraction of the organic component was varied in the range from 40-80 %v/v, with an increment of 5 %. All chromatographic experiments were performed in horizontal fashion using a horizontal development chamber (CAMAG, Lutenz, Switzerland). Approximately 0.3-0.5 μL of freshly prepared solutions in concentration ≈ 1 mg/ml have been applied on the surface of the plates at 5 mm distance from the edges. The chamber was saturated 15 minutes before chromatogram development. Solvent developing distance was 5 cm. The mobile phase pH range was between 5.5 and 6.5. No buffer solution was used. After development the plates have been dried in a stream of hot air and visually inspected under UV light (λ = 254 nm) allowing individual zones, corresponding to the target compounds, to be detected.

All substances and solvents used were of analytical purity grade. Benzodiazepines have been provided in small quantities from the Faculty of Pharmacy – University of Belgrade. Small amounts of PAHs have been a generous gift from the Chair of Environmental Chemistry, Faculty of Chemistry – University of Belgrade.

2.3 *Computational prediction of logP-s*

Mostly fragmental methods, either atom- or substructure-based have been employed to calculate log*P* values (with the exception of the linear solvation energy relationship (LSER) approach, AlogPs, and MLOGP, which are property-based). ALOGPs, AClogP, miLogP, ALOGP, MLOGP, XLOGP2 and XLOGP3 have been calculated through the Virtual Computational Chemistry Laboratory (VCCLAB, http://www.vcclab.org) [**55,56**], last time accessed on September 15, 2015. ALOGP and MLOGP are implemented in the Dragon software v. 6 (http://www.talete.mi.it); miLogP was developed by Molinspiration and implemented in Molinspiration property engine v2014.11 (http://www.molinspiration.com/); XLOGP3 is available through the XLOGP3 software (http://www.sioc-ccbg.ac.cn/?p=42&software=xlogp3). LSER estimated log*P* values have been calculated according to the model reported by Abraham *et al*. [**39**]. KOWWIN log*P* values have been obtained from the KOWWIN software, part of the EPI Suite package v.4.1 (U.S. EPA). ClogP was calculated using Chem Draw Ultra v. 11.0.1 (CambridgeSoft). ACDlogP and ABlogP estimates have been obtained with the freely accessible ACD I-Lab online database (https://ilab.acdlabs.com/iLab2/), last time accessed on September 15, 2015. ACDlogP was also available through the VCCLAB.

Table 1 List of compounds with their octanol-water partition coefficients (logK_{OW}). Recommended, experimentally determined values of logK_{OW} have been obtained from the EPI-Suite data base v.4.1 (EPA – U.S. Environmental Protection Agency)

No	Compound	logK_{OW}	Ref.	No	Compound	logK_{OW}	Ref.
1	Phenol	1.46*	[48]	26	Simazine	2.18*	[48]
2	4-Nitrophenol	1.91*	[48]	27	Propazine	2.93*	[48]
3	Benzyl Alcohol	1.10*	[48]	28	Ametryn	2.98*	[48]
4	1-Naphthylamine	2.25*	[48]	29	Prometryn	3.51*	[48]
5	1-Naphthol	2.85*	[48]	30	3-Nitrophenol	2.00*	[48]
6	2,4-Dichlorofenol	3.06*	[48]	31	2-Naphthol	2.70*	[48]
7	Anthracene	4.45*	[48]	32	4-Hydroxybenzaldehyde	1.35*	[48]
8	Acetophenone	1.58*	[48]	33	2-Aminophenol	0.62*	[48]
9	2,4,6-Trichlorophenol	3.69*	[48]	34	4-*t*-Butylphenol	3.31*	[48]
10	Ethyl-4 -hydroxybenzoate	2.47*	[48]	35	2,6-Dimethylphenol	2.36*	[48]
11	*p*-Anisidine	0.95*	[48]	36	4-Methoxyphenol	1.58*	[48]
12	1,2,3-benzotriazole	1.44*	[48]	37	Methyl-4-hydroxybenzoate	1.96*	[48]
13	Diphenylamine	3.50*	[48]	38	2-Nitrobenzaldehyde	1.74*	[48]
14	2,2'ipyridyl	1.50	[49]	39	3-Nitrobenzaldehyde	1.46*	[48]
15	4-Bromoaniline	2.26*	[48]	40	Phthalimide	1.15*	[48]
16	Benzophenone	3.18*	[48]	41	Oxazepam	2.24*	[48]
17	4-Aminobenzoic acid	0.83*	[48]	42	Lorazepam	2.39*	[48]
18	Pyrene	4.88*	[48]	43	Clonazepam	2.41*	[48]
19	Benzo(a)pyrene	6.13	[50]	44	Bromazepam	2.05	[53]
20	Fluorene	4.18*	[48]	45	Diazepam	2.82*	[48]
21	Acenaphthene	3.92*	[48]	46	Nitrazepam	2.25*	[48]
22	Naphthalene	3.30*	[48]	47	Chlordiazepoxide	2.44*	[48]
23	Phenanthrene	4.46*	[48]	48	Clobazam	2.12	[54]
24	Diben[a,h]anthracene	6.50*	[51]	49	Medazepam	4.41*	[48]
25	Benz[a]anthracene	5.76	[52]	50	Chrysene	5.81	[50]

*Values recommended by C. Hansch and A. Leo

2.4 *Data pretreatment and statistical analysis*

In order to put the lipophilicity indices on the same scale, several data pre-treatment methods have been investigated: a) standardization (St), also called autoscaling (mean centering and rescaling to unit standard deviation), b) range scaling between the lowest and the highest value of the shake-flask octanol water partition coefficient logK_{OW} value (0.62 and 6.75, respectively) (Rg) and c) rank transformation (Rk). All data pretreatments, descriptive statistics, PCA, HCA, and analysis of variance (ANOVA) were performed using Statistica v. 10 (Statsoft Inc. Tulsa, Oklahoma, USA).

In the case of HCA and PCA, the PLS, PCA and multivariate/Batch SPC module was used (Statistica v.10), while analysis of variance was carried out with the factorial ANOVA tool, part of the Advanced models (General linear) module (Statistica v. 10). HCA has been carried out using Ward's amalgamation rule and the Euclidian distance measure.

Two novel, non-parametric statistical methods, sum of ranking differences (SRD) and the generalized pair correlation method (GPCM) were also applied to provide a reliable comparison and ranking of the examined lipophilicity measures. These methods are entirely general and can give a fast and easy solution to comparison problems. Both methods are implemented as Microsoft Excel VBA macros and are available at http://aki.ttk.mta.hu/srd/ and http://aki.ttk.mta.hu/gpcm.

2.4.1 *Sum of ranking differences (SRD)*

SRD is a novel, fast and entirely general method for the comparison of alternative solutions to the same problem – *e.g.* different methods for the measurement/calculation of the same property (in this case, lipophilicity measures) **[57,58]**. It takes a matrix as its input, which contains the samples/molecules in its rows and variables/methods in its columns – thus, a cell in row *i* and column *j* contains the property (here, lipophilicity) value calculated/measured for the *i*th molecule with the *j*th method. SRD is based on the comparison of the rankings produced by the different methods, *i.e.* the samples are ranked (in the order of magnitude) according to each method plus a reference method (i), the differences between the rank numbers of each sample according to each method and the reference method are calculated (ii), and these ranking differences are added up for each method (iii). The reference method can be an exact "golden standard" or as in the present case the average. Using the arithmetic mean as reference instead of the recommended experimentally determined log*P*-s is justified based on two main points: *a*) the average realizes a consensus supported by the maximum likelihood principle, which yields a choice of the estimator as the value for the parameter that makes the observed data most probable (the average). [**59**]; *b*) even systematic errors cancel each other out not only the random errors, at least partially. Even if some small biases remain, we are better off using row-average than any of the individual methods. The resulting values are called SRD values and the smaller they are, the closer the method is to the reference (in terms of ranking). These SRD values are usually normalized to enable the comparison of different SRD calculations:

$$SRD_{nor} = 100SRD/SRD_{max}, \quad (5)$$

where SRD_{max} is the maximum possible SRD value.

SRD employs two validation steps: first, a Gauss-like curve is plotted based on the use of random ranks as a sort of randomization test (CRRN – Comparison of Results with Random Numbers); if a method overlaps with the Gauss-like curve, than it cannot be considered as significantly different from the random ranking. In the second step, seven-fold cross-validation is carried out (or leave-one-out cross-validation, if the number of samples is less than 14) to provide a population of SRD values, for which average, standard deviation, etc. can be calculated. An illustrative animation of the SRD calculation was published as a supplement to our recent article [**60**].

2.4.2 *Generalized pair correlation method (GPCM)*

The method is based on a 2 × 2 contingency matrix, where the frequencies of the event A, B, C and D are in the rows and columns [**61**]. These frequencies are calculated from a comparison between every selected dependent variable pairs ($X1$ and $X2$) and the reference (Y) variable (the arithmetic mean). Event A shows how many times both of the compared two variables strengthen the correlation (*i.e.*, if $Y_i > Y_j$, than $X1_i > X1_j$, and $X2_i > X2_j$). Similar to this, event D shows the amount of those cases, when both of the compared two variables weaken the correlation with Y variable (*i.e.*, if $Y_i > Y_j$, than $X1_i < X1_j$, and $X2_i < X2_j$). Events B and C are complementary: variable $X1$ strengthen and $X2$ weakens the correlation (event B) and vice versa (event C). The final decision of the comparison is based on Conditional Fisher's exact test or McNemar test [**61**]. The procedure is repeated for every possible variable pairs. A variable can win the final comparison, if it has the most "win" decisions. "No decision" results can be made if there is no significant difference between the correlations between the reference variable and the members of the pair. GPCM compares all the different variable pairs, and counts "wins", "losses" and "no decisions (ties)" between the variables (lipophilicity measures) [**61**]. The final result can be ordered in three different ways: simple ordering (which counts the number of wins), difference ordering (which calculates the differences between wins and losses) and significance ordering (the probability weighted form of difference ordering).

3 Results and discussion

3.1 *Exploratory data analysis and clustering*

Aiming to detect outliers, and explore the data structure for similarities among lipophilicity scales, PCA and HCA have been performed on the standardized dataset. Since comparison of lipophilicity measures to the arithmetic mean average (AMA) was introduced and justified in section 2.4.1, we have decided to include AMA in the PCA and HCA as well. Two PCs capture 88.83 % of the overall data variability in the data (*PC*1 84.86%, and *PC*2 3.97 %). The score plot (Figure S1, Supplementary material) reveals relatively homogeneous structure of the studied set of compounds. Only 4-aminobenzoic acid (comp. no. 17) was out of the 3 standard deviation confidence ellipse, most likely due to significant ionization under chromatographic conditions (α = 99.98 %). The rest of the solutes might be grouped into four, not entirely distinct groups: I – Comp. nos. 19, 24, 25, and 50; II – 7, 9, 13, 20, 23, and 49; III – 21, 22, 26-29, 34, 41-48; IV – 1-6, 8, 10, 12, 14, 15, 30, 32, 35-40. The first and the second group contain mostly polyaromatic hydrocarbons. All benzodiazepines, except of midazolam, and all triazine herbicides are in the third group along with a few phenolic compounds. The rest of phenols, aromatic amines, aldehydes, ketones, and esters are in the fourth group.

The majority of the lipophilicity measures responsible for such disposition of compounds have the highest loading values in the *PC*1 direction, grouped in the tight range of 0.80-0.99 units (Figure 1). Exceptions are slopes, *b*, obtained on a C18 stationary phase using methanol and acetonitrile as organic mobile phase modifiers. The majority of computational approaches (XLOGP2, miLogP, AlogPs, ACDlogP, KOWWIN, ClogP, and XLOGP3) are centralized in the extremely small range of *PC*1 *vs*. *PC*2 loading space together with the experimentally determined logK_{OW}(exp). They are further surrounded with chromatographic descriptors, mostly *PC*1/R_M, mR_M, and R_M^0 indices in the first level, and C_0 in the second one, derived under different chromatographic conditions. Water solubility, *i.e*. its negative logarithm (-log*S*) perform similarly as C_0 and it is the closest to the AMA.

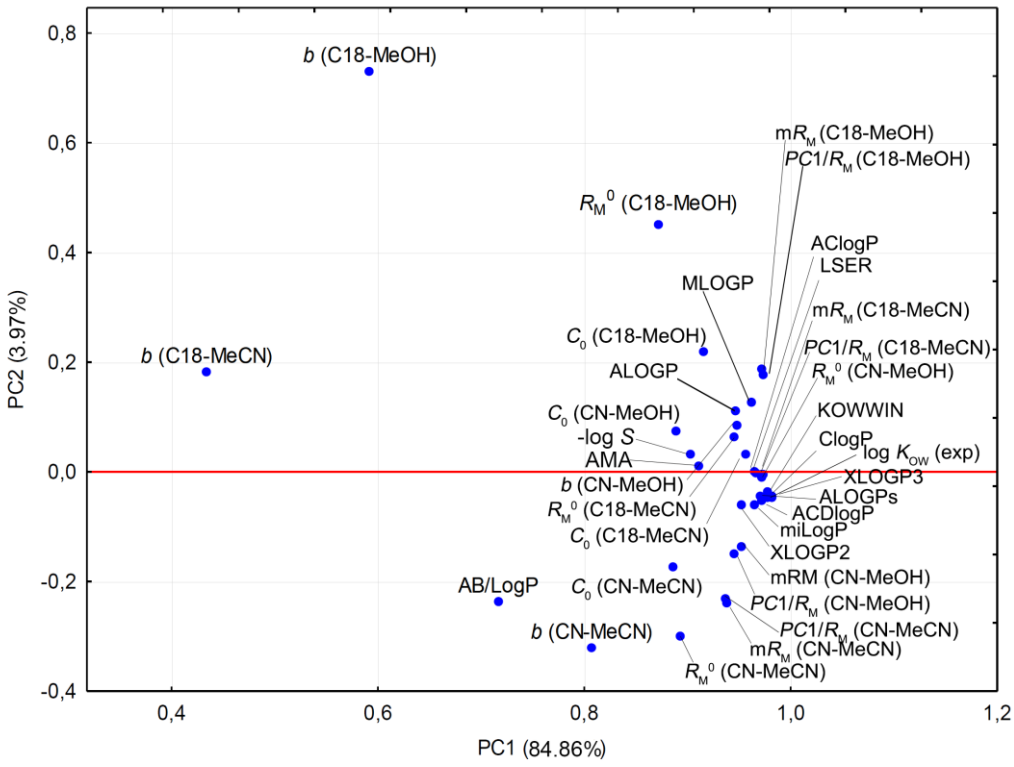

Figure 1 Principal component analysis loading plot; similar lipophilicity measures are positioned close to each other.

Fine data variability along the *PC*2 direction (3.97%) allows distinction of chromatographic indices obtained on cyano-modified silica and those derived from octadecyl-silica (Figure 1). The majority of the chromatographic indices obtained on cyano-modified silica are located in the lower part of the loading plot; the others are in the upper one. Between these two groups lies a very coherent group composed mostly of computational log*P*-s.

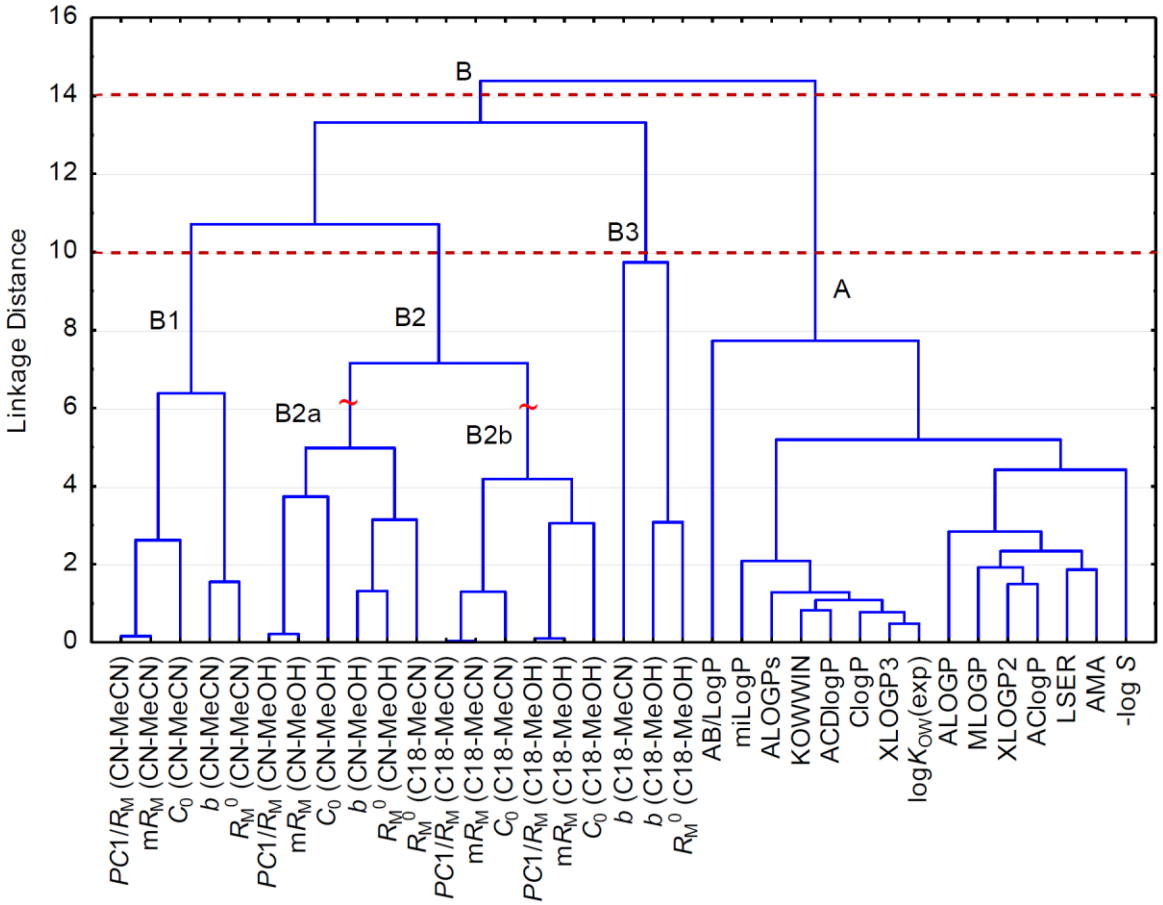

Figure 2 Hierarchical cluster analysis dendrogram showing similarities among different chromatographic lipophilicity indices (cluster B) and *in silico* predicted log*P* values (cluster A). The experimental values (logK_{OW}) and the arithmetic mean average (AMA) are also included in cluster A.

Clear distinction between computationally estimated log*P*-s and chromatographic indices is obtained by HCA (Fig 2). Cluster A, comprised of *in silico* predicted log*P*-s, log*S*, and AMA. Cluster B containing all chromatographic lipophilicity indices is separated at the level of 14 linkage distance units. However, the difference between them is only ~ 2 distance units. Further grouping of indices according to stationary and mobile phases is obvious at the level of 10 distance units. While B1 gathers only chromatographic indices obtained on cyano-modified silica, B2 includes those obtained from both stationary phases (B2a corresponds to CN-modified silica, B2b accounts for C18-modified silica). Also, the use of acetonitrile *vs*. methanol differentiates between B1 and B2a. Cluster A can be further divided into two sub-clusters. However, it cannot be explained by the subdivision of methods to property- and substructure-based ones. Also, there is the following trend in mutual similarity among the types of chromatographic descriptors on almost all chromatographic systems: the most similar to each other are mR_M and *PC*1/R_M (the shortest linkage distance), the most similar to them is C_0, while R_M^0 and *b* are gathered in separate clusters. Figure 2 clearly shows that the classical chemometric method HCA cannot establish a link between calculated and chromatographic indices. The experimental value logK_{OW} is far away and separated by calculated indices from the chromatographic ones. It is also separated from the AMA value, which is located in the first subcluster (Figure 2).

If the recommended experimental values logK_{OW}(exp) are considered as the reference, both PCA and HCA lead to the same decision about the best lipophilicity measure, *i.e.* XLOGP3. However, each of the considered lipophilicity estimation methods has systematic as well as random errors. Using the arithmetic mean as the reference instead of the recommended experimental logK_{OW} is justified based on two main points: a) the maximum likelihood principle – the average is the most probable solution and b) even systematic errors cancel each other out. According to the closeness of each method to the average point, which is included in the PCA and HCA plots the best lipophilicity estimate is obtained by log*S* in the case of PCA, and LSER in the case of HCA. However, several problems still remain. The two most important are: *i*) unknown statistical significance of obtained grouping and similarity to the reference and *ii*) loss of information due to dimensionality reduction in PCA.

3.2 *Comparison of lipophilicity measures by means of SRD and GPCM*

With the aim to overcome the aforementioned problems and answer the above questions, lipophilicity measures were compared, ranked, and grouped with non-parametric ranking methods, SRD and GPCM. Both methods also provide information regarding statistical significance of the ranking.

According to the SRD-CRRN ranking of standardized lipophilicity data (Fig. 3a), the lipophilicity estimate closest to the reference, in this case the average is Clog*P*. The parameter is closely followed by the recommended experimentally determined values of the octanol-water partition coefficient, then XLOGP3, *etc*. Actually the pseudo-continuous ranking occurs in the range of scaled SRD score values 9.28 – 14.88, including several chromatographic descriptors and the majority of *in silico* lipophilicity estimates. The farthest lipophilicity measures are the slopes *b* obtained on C18 silica. Generally, chromatographic indices obtained on CN-modified silica have lower SRD scores than those obtained on C18. All studied lipophilicity indices are able to rank the studied compounds according to their lipophilic character better than random ranking – none of the lines overlap with the random number distribution, *i.e.*, their ability to measure lipophilic character of selected group of compounds is statistically significant. Different data pretreatment methods might lead to slightly different ranking patterns. However, the milestone variables (the closest and the farthest from the reference) remain the same (Table S4a).

GPCM of standardized data provided a slightly different ranking pattern (Figure 3b) with a characteristic degeneracy of some variables (variables having the same or indistinguishably similar ranking scores). Here the average was used as a dependent variable. The lipophilicity measure closest to the average in this case was the set of recommended values, logK_{OW}(exp), closely followed by miLogP, R_M^0 (CN-MeOH), $PC1/R_M$ (C18-MeOH), *etc*., the same variables that can be found in the pseudo-continuous ranking in the case of SRD, with a slightly (not significantly) altered order of variables. The variables that are farthest from the consensus are again the slopes *b* obtained on C18-silica closely followed by *b* (CN-MeCN) and AB/logP.

If the logK_{OW}(exp) values are used as the benchmark instead of the arithmetic mean average, different ranking is obtained in the case of both SRD and GPCM, especially in terms of variable cluttering and degeneracy. However, the most important variables such as the closest ones (ClogP, XlogP3) and the farthest from the reference (*b* (C18-MeCN)) preserved their positions (Figure S2a and b, Supplementary material). Although the information about the relation of lipophilicity estimators, especially the closest and the farthest methods to this particular reference has been obtained, the information regarding the reference itself is lost. It can be only provided if the average is kept as the reference point of view.

Various methods can provide different orderings. SRD has the advantages of "multicriteria optimization", c.f. ref. [**62**]. It is clear that in this case *in silico* methods are close to the recommended logK_{OW}(exp) values, while chromatographic estimations might seem to perform worse. The reason for such, possibly "biased" behavior might be the use of the same, or at least most of the studied compounds in the training of presented *in silico* methods. External validation might provide a proof for a possible bias. However, this is not necessarily a good choice, though many authors still support it. In her recent paper Gramatica advises "to avoid the limitation of using only a single external set, we […] always verify our models on two/three different prediction sets" [**63**]. Independently

from this, our recent paper clearly shows that the ordering of merits for external validation is indistinguishable from random ranking. [**62**].

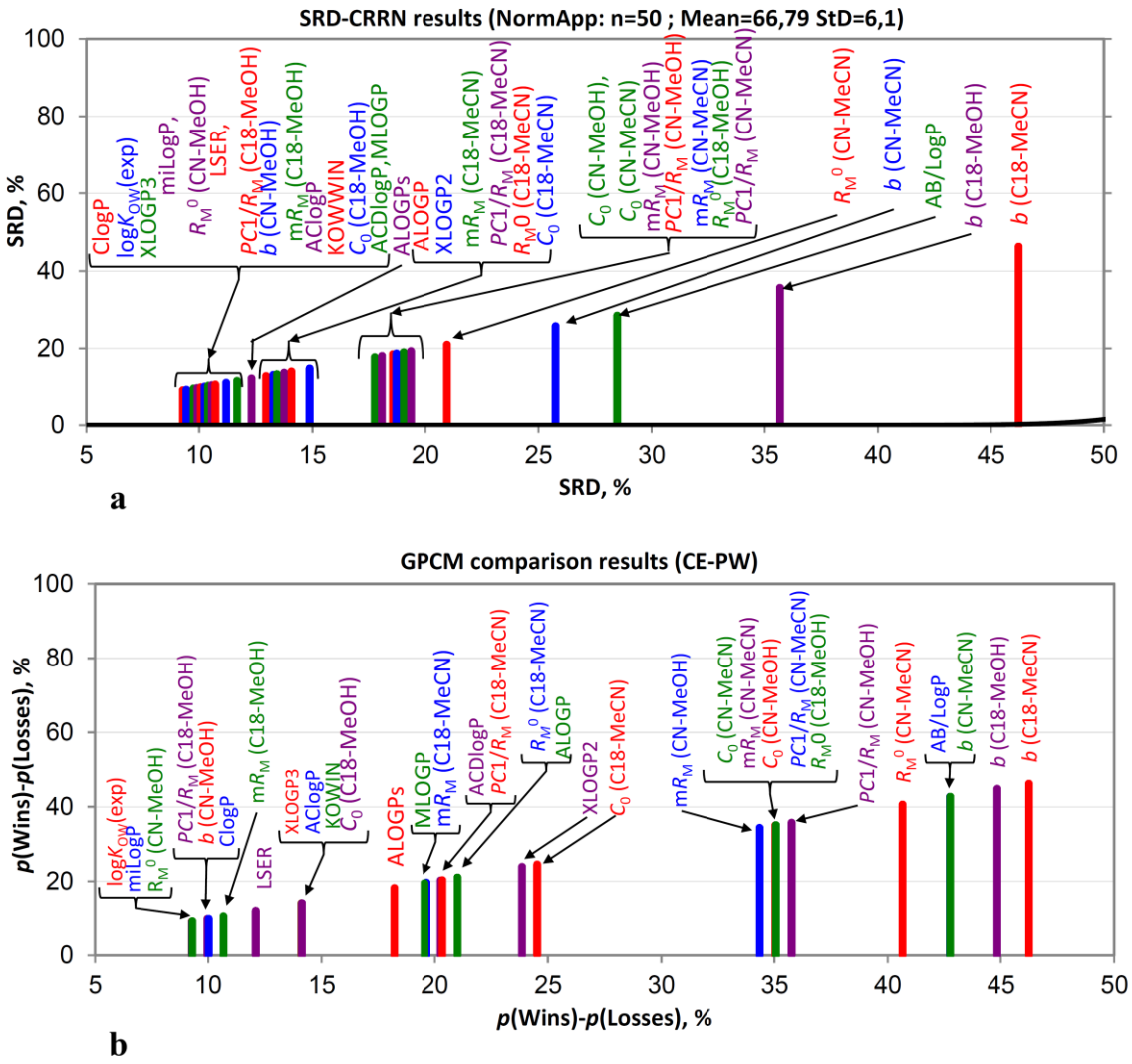

Figure 3 Comparison, ranking and grouping of chromatographic and *in silico* lipophilicity measures by SRD-CRRN (a), and GPCM-CEPW ranking (b); where CEPW stands for probability weighted ranking (PW) based on Fisher's conditional exact test (CE). Left side *y*-axes and *x*-axes are the same and denote score values in %.

Nevertheless we have carried out the SRD and the GPCM ranking of lipophilicity measures on a subset of compounds with logK_{OW} values that are likely to be correctly measured with the shake-flask method (logK_{OW} < 3 and determined with the shake-flask procedure which was verified through a meticulous tracing of the original articles, Table S1, Supplementary material). The arithmetic mean average was used as the reference.

Ranking of lipophilicity measures is slightly altered for both SRD and GPCM (Figure S3a and b, respectively in the Supplementary material), however the general trend is the same and the most important variables retained their positions compared to the ranking based on the overall set of compounds. In that sense R_M^0 (CN-MeOH) is selected as the lipophilicity measure closest to the average by SRD instead of ClogP which is the second closest (Figure S2a), while the farthest ones (*b* (C18-MeCN), AB/logP, *b* (C18-MeOH), R_M^0 (CN-MeCN)) remain in their original positions. GPCM provides ranking in a similar fashion identifying the following measures as the closest to the average: logK_{OW}(exp), miLogP, R_M^0 (CN-MeCN), and *b* (C18-MeCN) as the farthest one. Therefore, conclusions related to identified approaches using the overall set of compounds, for which logK_{OW} values originate

from different sources and possibly from different measurement techniques, are valid for the limited set of compounds for which logK_{OW} values are more likely to be measured with the shake-flask method.

In order to test whether the data pretreatment methods and ranking methods employed lead to significantly different results, analysis of variance (ANOVA) was performed on the GPCM and SRD score values after sevenfold cross-validation. ANOVA was also used to test for the possible difference among chromatographic lipophilicity indices, the use of different stationary and mobile phases, and *in silico* prediction methods.

Uncertainty has been introduced to SRD and GPCM values by a jackknife-like validation procedure (cross-validation) as follows: seven minors of the original data matrices were obtained by removing 1/7 of samples. Every truncated data set was then subjected to SRD-CRNN and GPCM-CEPW ranking procedures, providing seven score values for each of the lipophilicity measures, in total 1386 scores (33 variables (lipophilicity indices) × 3 data pretreatment methods × 2 ranking approaches (SRD and GPCM) × 7 repetitions). GPCM scores were range scaled to fit the size and order of SRD-s. Obtained scores were used as an input for ANOVA. The following factors and a full interaction model without quadratic terms were considered:

$$Score = b_0 + b_1F_1 + b_2F_2 + b_3F_3 + b_{12}F_1F_2 + b_{13}F_1F_3 + b_{23}F_2F_3 + b_{123}F_1F_2F_3 \quad (6)$$

The types of data pretreatment are incorporated in the three level factor, F_1: standardization (St), range scaling (Rg), and ranking (Rk); F_2 represents the type of lipophilicity scale ranking (two levels): SRD and GPCM; F_3 takes into account the type of lipophilicity measure at six levels: logK_{OW} (exp) – shake-flask method, Cmp – computationally estimated log*P*-s, and four types of chromatographic lipophilicity indices - C18-MeOH, C18-MeCN, CN-MeOH, CN-MeCN, referring to the use of octadecyl- and cyano-modified silica as stationary phases and methanol and acetonitrile as mobile phase components respectively.

Statistical parameters of ANOVA are summarized in Table 2. The data pretreatment methods do not differ at the predefined significance level $p = 0.05$. However, the type of ranking method, the lipophilicity measure (factors F_2, F_3) and their cross-coupling term are statistically significant. The other interaction terms are not significant (Table 2, last column).

Table 2 Univariate test for significance of factor effects for 1386 score values obtained with SRD and GPCM ranking procedures. Factors: F_1 – methods of data pretreatment: standardization (St), range scaling (Rg), rank transformation (Rk); F_2 – ranking methods: SRD and GPCM; F_3 - type of lipophilicity measures: Recommended experimental logK_{OW} values (logK_{OW}(exp)), *in silico* estimated log*P*-s (Cmp), Chromatographic indices obtained on octadecyl- and cyano-modified silica using methanol and acetonitrile as mobile phase components (C18-MeOH, C18-MeCN, CN-MeOH, CN-MeCN, respectively). Significant factors are indicated in bold.

Factor	Sum of squares	Degrees of freedom	Mean squares	*F*	*p*
Intercept	**313548.9**	**1**	**313548.9**	**3942.905**	**0.000000**
F_1	11.2	2	5.6	0.070	0.932187
$\boldsymbol{F_2}$	**11252.2**	**1**	**11252.2**	**141.498**	**0.000000**
$\boldsymbol{F_3}$	**32348.2**	**5**	**6469.6**	**81.356**	**0.000000**
$F_1 \times F_2$	4.2	2	2.1	0.026	0.974172
$F_1 \times F_3$	35.2	10	3.5	0.044	0.999996
$\boldsymbol{F_2 \times F_3}$	**6676.8**	**5**	**1335.4**	**16.792**	**0.000000**
$F_1 \times F_2 \times F_3$	10.0	10	1.0	0.013	1.000000

Factor effects are illustrated in a way that is easier to perceive, in Figure 4. Considering all types of lipophilicity parameters, the GPCM procedure resulted in generally higher scores compared to the SRD, except in the case of the shake-flask method, in which GPCM and SRD scored the same (Figure 4). Considering that the smaller the scores the better, *i.e.*, the closer to the average are the lipophilicity measures, it is easy to find that the lipophilicities obtained from the shake-flask method are the best ones. Computational methods closely follow the recommended experimental logK_{OW} values while chromatographic lipophilicity indices are close to the

computational log*P*-s. Both GPCM and SRD confirm that the use of different organic components in the mobile phase does not have any influence on the ordering of lipophilicity scales in the case of octadecyl-silica. Unlike octadecyl-silica, cyano-modified silica gel makes a significant difference with regard to the use of methanol *vs.* acetonitrile. Besides that, GPCM does not differentiate lipophilicity parameters obtained on cyano-modified silica using methanol as a mobile phase modifier from the rest of the lipophilicity scales measured on octadecyl-silica. Only chromatographic indices obtained on CN-silica using acetonitrile are significantly different. Similar conclusion might be obtained from the HCA dendrogram (Clusters B1 and B2a, Figure 2). No difference among data pretreatment methods (standardization, range scaling, and ranking) can be seen (Fig. 4). The reason why the use of acetonitrile *vs*. methanol alters lipophilicity assessment on CN-silica is most likely due to the strong dipolar properties of cyano groups of both stationary and mobile phase components. Since both have the same ability for dipolar and polarizable interactions with a solute, but expressed in opposite directions, the overall interaction impact on retention might be significantly diminished. This is not expected to occur in the case of C18-silica. The same pattern can be observed for GPCM and SRD scores: The pattern is increasing, from the recommended logK_{OW} values *via* computational measures and further on to C18-MeOH and C18-MeCN, then, an exception can be observed: a decrease at CN-MeOH, then an increase again at CN-MeCN.

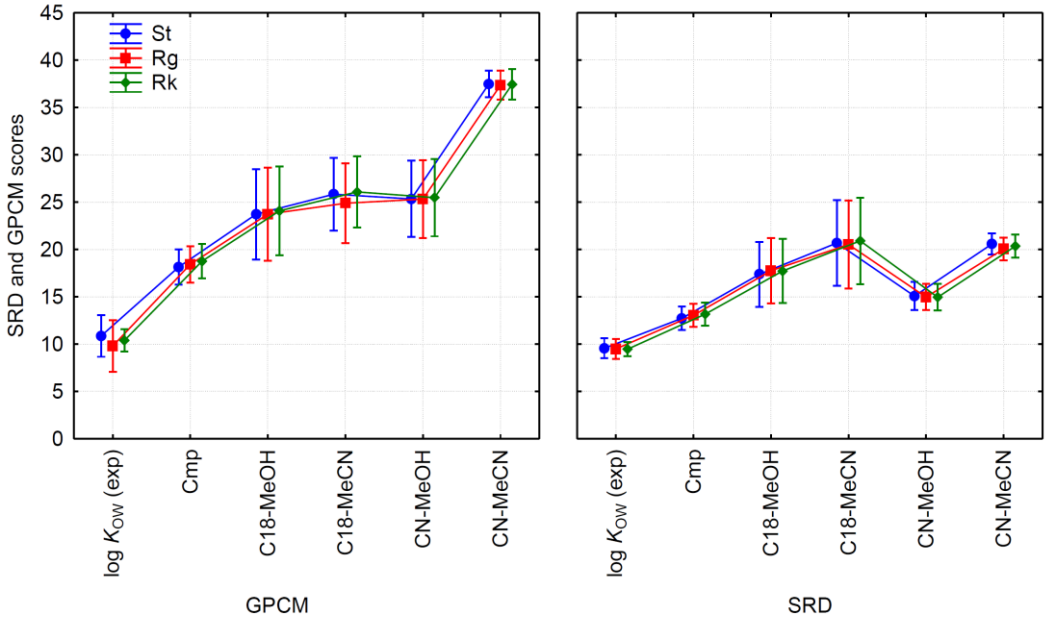

Figure 4 Effect of factors by analysis of variance for sevenfold cross-validated SRD and GPCM score values; the average was used for reference in ranking. Score values were plotted on the *y*-axis. Vertical bars denote 0.95 confidence intervals.

ANOVA of SRD and GPCM scores provides information about the statistical significance of differences among lipophilicity measures, which is an important issue, not adressed by PCA, or HCA. Considering GPCM scores, and based on 95% confidence intervals (denoted as up and down whisker-like lines at each data point, Figure 4), no statistically significant differences can be detected among *in silico* determined log*P*-s, and chromatographic indices obtained on C18- and CN-modified silica, except of those obtained on CN-silica in combination with acetonitrile as a modifier. In the case of SRD scores, no differences can be observed among C18-based and CN-silica-based chromatographic indices. However, statistical difference among computationally calculated log*P*-s and chromatographic indices is a borderline case. The recommended logK_{OW} values (logK_{OW}(exp)) are the closest to the reference (consensus) and clearly statistically distinct according to both comparison methods, SRD and GPCM.

3.3 *Pattern recognition between lipophilicity measures by non-parametric correlations based on SRD and GPCM*

Sometimes the selection of the benchmark (golden standard) is not unambiguous. In that case it is of particular interest to employ a methodology that provides information about how different variables relate to each other in an easily perceivable way. For this type of problems, correlation matrices are most often used. Pearson's correlation coefficient has been already extensively used for the assessment of novel lipophilicity indices [**24,25,44]**. In order to extend the capabilities of SRD to this type of problems, we have implemented a MS Excel VBA macro for the generation of "SRD heatmaps", where we use all of the variables in turn as the reference to produce a matrix of SRD values. We have termed this approach Comparison with One VAriable at a Time (or COVAT, as we refer to it in the rest of the article). The final results are presented in a heatmap format, with three coloring schemes: *relative*, *absolute* and *Gaussian*. With *relative* coloring, the range of (normalized) SRD values occurring in the heatmap are divided into ten sub-ranges of the same size (i.e. $SRD_{max}/10$, as $SRD_{min} = 0$ per definition for the diagonal elements) and a color is assigned to each of these sub-ranges. *Absolute* coloring facilitates the comparability of different heatmap SRD calculations, as the ten sub-ranges are fixed in this case (0-10, 10-20…90-100%). SRD values overlapping with the Gaussian distribution of random ranking can be highlighted with the *Gaussian* coloring scheme. A color reference is provided with each output table created with the macro. To enable a better perception of the underlying structure of the SRD matrix, the rows and columns of the heatmap are reordered in the ascending order of the row-wise average SRD values (which is at the same time, the ascending order of the column-wise average SRDs as the matrix is ideally symmetric). As a consequence, clusters of similar methods/models/etc. (here, lipophilicity measures) can be detected along (both sides of) the diagonal. While the resulting SRD matrices are ideally symmetric, the presence of tied values in the input matrix can introduce a small extent of asymmetry. However, if the occurrence of tied values is not too frequent, this usually does not impair the rearrangement of the matrix or the perception of the underlying data structure. The VBA script to produce SRD-COVAT heatmaps is available for download on our website: http://aki.ttk.mta.hu/srd/.

A similar approach was taken for the production of GPCM-COVAT matrices. However, a significant difference is that GPCM-COVAT matrices will be asymmetric by definition, as in the case of GPCM, probability weighted scores differ whether X_i or Y_i is used as a benchmark (therefore the complete absence of symmetry is expected). This has significant consequences on the interpretation of GPCM-COVAT matrices. Basically the benchmark variables are arranged in columns in an ascending order of the column-wise total sums of the scores. However, row-wise summation leads to different results, therefore the arrangement of GPCM-COVAT matrices demands a compromise.

We compared four approaches: a) classical correlation matrix based on Pearson's correlation coefficient, b) non-parametric correlation matrix based on Spearman's rank correlation coefficient (rho), c) SRD-COVAT, and d) GPCM-COVAT matrices.

In order to identify similarities and dissimilarities among lipophilicity scales, the relative coloring scheme was applied, consisting of ten different colors. The most similar variables (the maximum similarity or the minimum dissimilarity measure value) are colored in red, while the most dissimilar ones (the minimum similarity or the maximum dissimilarity value) are marked with dark blue.

All matrices show similar patterns but the classical Pearson's and Spearman's correlation matrices are more similar to each other (Figure 5a and 5b as compared to Figure 6a and 6b). The highly correlated lipophilicity measures are located in the upper left corner (square marked as L1 which is mostly composed of chromatographic indices obtained on C18-silica, with a few *in silico* estimates and the recommended values $\log K_{OW}$). In the lower-intermediate parts (L2 and L3) of both heatmaps somewhat dissimilar lipophilicity indices are located, mainly chromatographic ones derived from experiments on CN-silica, with a few computational measures (orange, ochre, and yellow colored). An important difference between the Pearson's and Spearman's heatmaps are the different portions of orange, and ochre colors (70-80% of the maximum correlation values), which are dominant in the Pearson's map. Therefore, it is obvious that the Pearson's heatmap has a slightly lower discriminatory power.

Highly dissimilar (orthogonal) variables, colored in dark green and blue, are located along the bottom and right edge of the heatmaps – parts L4 and L5 (slopes *b* (C18-MeOH) and *b* (C18-MeCN)).

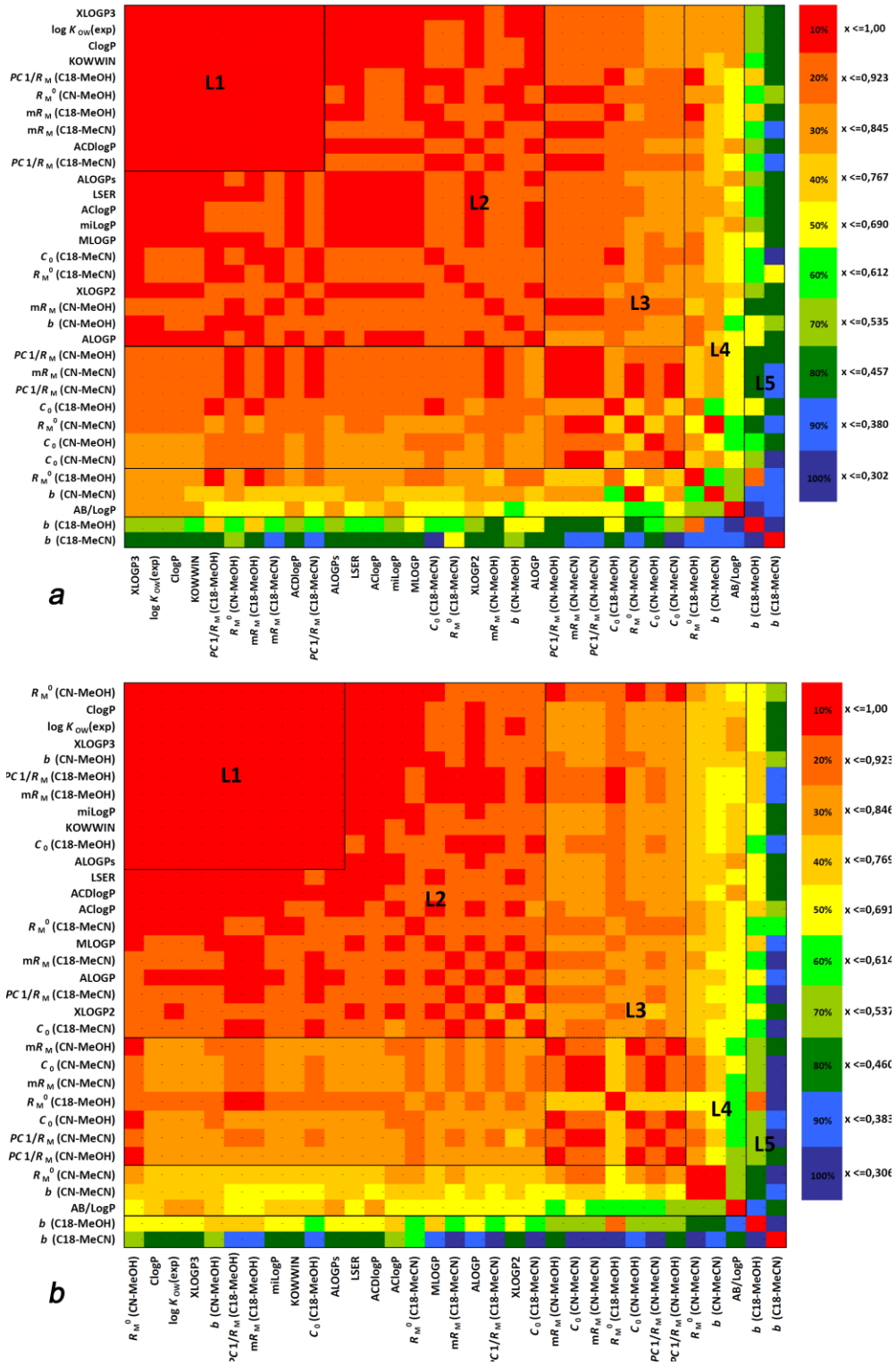

Figure 5 Relative colored heatmap representation of Pearson (a) and Spearman (b) correlation matrices. Red color represents the highest correlation values while blue marks the lowest one. Color codes are provided on the right side with absolute and relative (%) values.

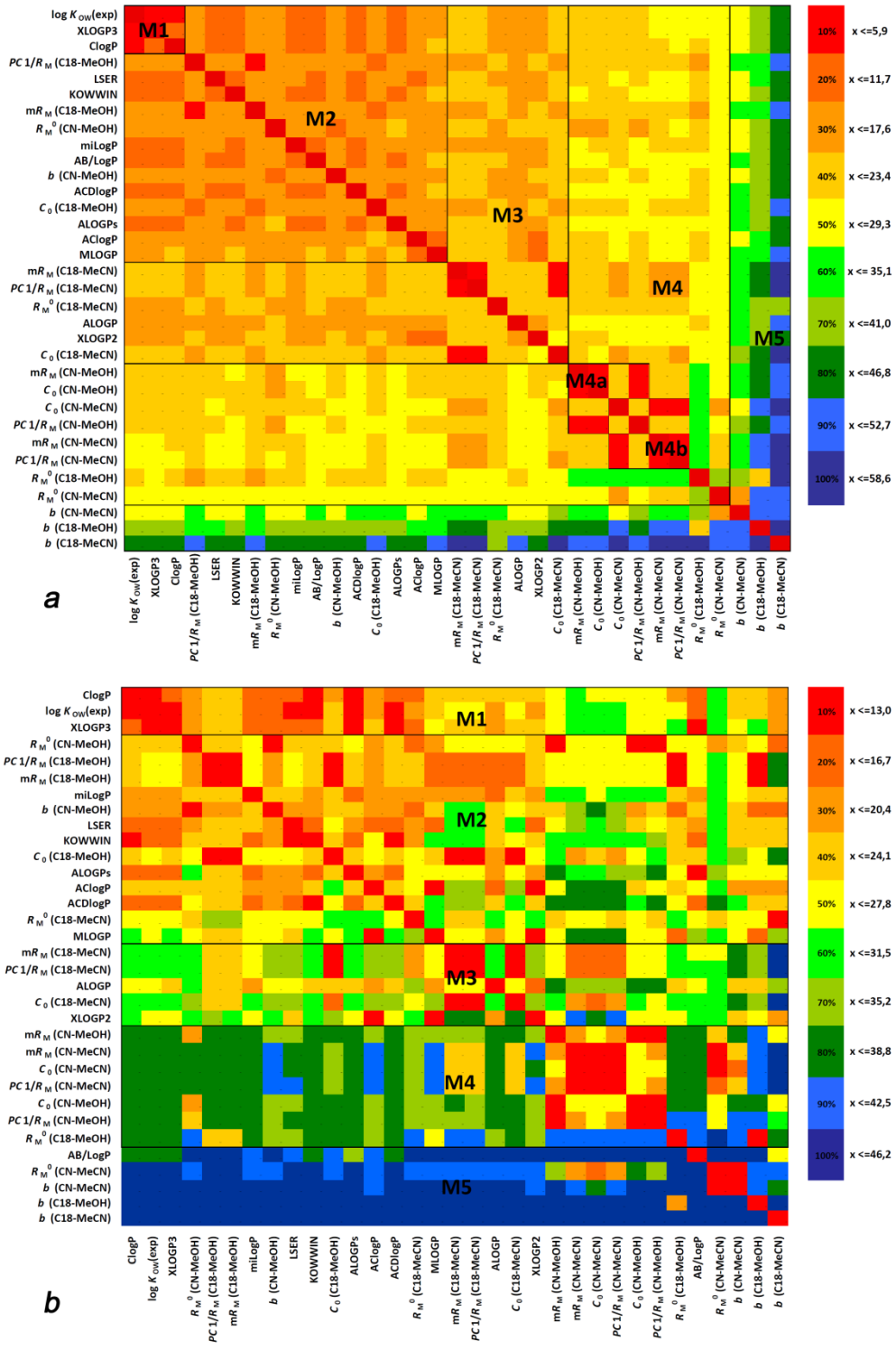

Figure 6 Heatmap representations of SRD (a) and GPCM-CEPW (b) COVAT matrices. Red color represents the lowest score value (the highest similarity), while blue marks the highest one (the lowest similarity). Color codes are provided on the right side with absolute and relative (%) values. CEPW stands for probability weighted ranking (PW) based on Fisher's conditional exact test (CE)

Instead of the current misuse of Pearson's correlation matrices [**24,25,44]** we would like to encourage the implementation of: (a) adequate arrangement of variables and (b) coloring schemes which enables patterns among variables to be easily perceivable. The choice of the best variable/lipophilicity measure following the aforementioned matrix arrangement is straightforward, *i.e.* the variable that correlates the best with the majority

(the upper left corner) is the best choice: XLOGP3, logK_{OW}(exp), and ClogP, along with the rest of the lipophilicities belonging to the cluster A (Figure 5a and b), and can replace the rest of them.

The SRD-COVAT heatmap provides similar patterns as the Spearman based-one, but with significantly greater discrimination power (Figure 6a). Practically, red, orange, ochre, dark and light yellow colored squares (regions) that cover variables of different similarity are well defined and easily noticeable (clusters M1-M5). Also, the upper left red square (M1) of highly similar variables is extremely narrowed to only XLOGP3, logK_{OW}(exp), and ClogP. Orange and ochre regions gathers mostly computationally estimated log*P*-s, mixed with few C18- and CN- derived lipophilicity indices (M2). The rest of C18-silica based descriptors can be found in the darker yellow region located in the middle parts of the heatmap (M3), while CN-based lipophilicity scales are predominantly colored with light yellow parts located in the lower parts of a heatmap space (M4). CN-based lipophilicity scales are here distinctively differentiated according to the use of methanol or acetonitrile (M4a and M4b red colored regions). The most different lipophilicity measures can be found at the matrix margins colored in dark green and blue (*b* (C18-MeOH) and *b* (C18-MeOH)).

Also, it can be concluded that the mR_M, *PC*1/R_M and C_0 measures are highly correlated for each stationary phase-mobile phase combination: they can be detected as smaller clusters along the diagonal. (*b* values on the other hand are not necessarily present in these clusters.)

The GPCM-COVAT heatmap, based on probability weighted ranking using Fisher's conditional exact significance testing, results in a similar pattern and variable arrangement, with some insignificant differences (Figure 6b). GPCM has the greatest discriminatory power of the above mentioned cases. Row-wise summation shows a more easily distinguishable pattern. The coloring scheme suggest that the variables that are the most similar with the rest of the studied lipophilicity scales should be found at the top of the heatmap, colored in red and yellow (the best ones, since they can replace most of the others). In this particular case those are: ClogP, logK_{OW}(exp), and XLOGP2 (belonging to M1), as well as R_M^0(CN-MeOH), *PC*1/R_M(C18-MeOH) and mR_M(C18-MeOH) (belonging to M2). The most orthogonal ones, on the other hand, are located at the bottom of the heatmap (colored in blue, M5): R_M^0(CN-MeCN), *b*(CN-MeCN), *b*(C18-MeOH), and *b*(C18-MeCN).

Although the coloring is somewhat arbitrary, it is astonishing that methods based on completely different concepts provide so similar patterns for ordering lipophilicity indices.

4 Conclusions

Many chromatographic methods in addition to *in silico* estimation approaches have been developed so far in order to measure/quantify the lipophilic character of compounds. Now we provide a unique systematic approach to select the most appropriate lipophilicity measures available. Many of the chemometric methods applied are misused, leading often to wrong conclusions. Sum of ranking differences (SRD) leads to the selection of the closest and farthest lipophilicity measure to the reference, in a straightforward manner, compared to principal component analysis (PCA) and hierarchical cluster analysis (HCA). While being based on completely different concepts, generalized pairwise correlation method (GPCM) provides a similar ordering of lipophilicity scales. Comparison with "random numbers" in the case of SRD provides information regarding the statistical significance of the obtained ranking (which cannot be obtained from PCA and HCA). Furthermore, uncertainties among SRD and GPCM scores, introduced by sevenfold cross-validation experiments enables to test statistical significance among studied lipophilicity scales, as well as different factors by analysis of variance (ANOVA) (data pretreatment approaches, ordering and ranking procedures). Two factors, namely the way of ranking (SRD and GPCM) and the type of lipophilicity measures have been identified as statistically significant by ANOVA. SRD generally results in lower scores than GPCM. The shake-flask method provides the lowest scores (the closest to the average) and therefore it can be considered as the best one. Computational estimates closely follow. Chromatographic indices obtained on octadecyl-modified silica do not differ significantly in terms of the use of methanol or acetonitrile as the mobile phase component. However, the situation is different when it comes to cyano-modified silica, in which

case acetonitrile exhibits different effects compared to methanol, which can be explained with the strong dipolar properties of cyano groups of the stationary phase and acetonitrile as a constituent of the mobile phase, that cancel each other out.

Introduction of a *relative* coloring scheme to correlation matrices and their adequate arrangement enables the discovery of intricate relationships among lipophilicity scales and the selection of the most similar and dissimilar ones. SRD-COVAT matrix has more discriminating power than Pearson and Spearman based-ones. The window that grasps the lipophilicity scales that are mostly correlated with others are significantly narrowed down (in this case to only three recommended: logK_{OW}(exp), XLOGP3 and ClogP). However, although based on completely different concepts, GPCM-COVAT heatmaps discriminate lipophilicity scales the most.

Acknowledgments

This work has been supported by the Ministry of Education, Science and Technological development, of the Republic of Serbia, grant No. 172017. KH is indebted to the support from OTKA (Hungary), contract No. K112547. FA would like to express gratitude to prof. Dušanka Milojković-Opsenica for the support in studying chromatographic mechanisms and developing of chromatographic methods for assessing lipophilicity over many years.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jpba.2016.04.001

References

[1] S.K. Poole, C.F. Poole, Separation methods for estimating octanol–water partition coefficients, J. Chromatogr. B 797 (2003) 3–19.
[2] M.J. Waring, Defining optimum lipophilicity and molecular weight ranges for drug candidates - molecular weight dependent lower logD limits based on permeability, Bioorg. Med. Chem. Lett. 19 (2009) 2844–2851.
[3] M.P. Gleeson, Generation of a set of simple, interpretable AMDET rules of thumb, J. Med. Chem. 51 (2008) 817–834.
[4] D.A. Price, J. Blagg, L. Jones, N. Greene, T. Wager, Physicochemical drug properties associated with in vivo toxicological outcomes: a review, Expert Opin. Drug Metab. Toxicol. 5 (2009) 921–931.
[5] M.J. Waring, Lipophilicity in drug discovery, Expert Opin. Drug Discov. 5 (2010) 235–248.
[6] H. Meyer, Zur Theorie der Alkoholnarkose I. Welche Eigenschaft der Anästhetika bedingt ihre narkotische Wirkung? Arch. Pharmakol. Exp. Pathol. 42 (1899) 109–118.
[7] E. Overton, Studien über die Narkose, zugleich ein Beitrag zur allgemeinen Pharmakologie, (1901) Jena, Gustav Fischer, Switzerland.
[8] IUPAC, Compendium of Chemical Terminology, second ed. (the Gold Book),Compiled by A. D. McNaught and A. Wilkinson, (1997) Blackwell Scientific Publications, Oxford. XML online corrected version: http://goldbook.iupac.org (2006) created by M. Nic, J., Jirat, B. Kosata; updates compiled by A. Jenkins.
[9] OECD Guideline for the testing of chemicals, Test No. 117, Partition coefficient(n-octanol/water), HPLC method (2004) OECD, Paris, <www.oecd.org/>.
[10] M. Janicka, D. Pietras-Ozga, Chromatographic evaluation of the lipophilicity of N-phenyltrichloroacetamide derivatives using micellar TLC and OPLC, JPC, J. Planar Chromatogr. Mod. TLC 23 (2010) 396–399.
[11] M.J. Medina-Hernandez, S. Sagrado, Chromatographic evaluation of hydrophobicity using micellar mobile phases, J. Chromatogr. A 718 (1995) 273–282.
[12] H. Nishi, Pharmaceutical application of micelles in chromatography and electrophoresis, J. Chromatogr. A 780 (1997) 243–264.

[13] S.S. Shtykov, E.G. Sumina, N.V. Tyurina, Calculation of partition coefficients of organic reagents in micellar thin-layer chromatography, J. Anal. Chem. 57 (2002) 322–325.
[14] L. Xu, L. Li, J. Huang, S. Yu, J. Wang, Determination of the lipophilicity (logPo/w) of organic compounds by microemulsion liquid chromatography, J. Pharm. Biomed. Anal. 102 (2015) 409–416.
[15] M. Janicka, Correlations between chromatographic parameters and bioactivity predictors of potential herbicides, J. Chromatogr. Sci. 52 (2014) 676–684.
[16] F. Pehourcq, C. Jarry, B. Bannwarth, Potential of immobilized artificial membrane chromatography for lipophilicity determination of arylpropionic acid non-steroidal anti-inflammatory drugs, J. Pharm. Biomed. Anal. 33 (2003) 137–144.
[17] D. Vrakas, D. Hadjipavlou-Litina, A. Tsantili-Kakoulidou, Retention of substituted coumarins using immobilized artificial membrane (IAM) chromatography: a comparative study with n-octanol partitioning and reversed-phase HPLC and TLC, J. Pharm. Biomed. Anal. 39 (2005) 908–913.
[18] K. Valkó, Application of high-performance liquid chromatography based measurements of lipophilicity to model biological distribution, J. Chromatogr. A 1037 (2004) 299–310.
[19] M. Starek, L. Komsta, J. Krzeka, Reversed-phase thin-layer chromatography technique for the comparison of the lipophilicity of selected non-steroidal anti-inflammatory drugs, J. Pharm. Biomed. Anal. 85 (2013) 132–137.
[20] S. Segan, J. Trifkovic, T. Verbic, D. Opsenica, M. Zlatovic, J. Burnett, B. Solaja, D.Milojkovic-Opsenica, Correlation between structure, retention, property, and activity of biologically relevant 1,7-bis(aminoalkyl) diazachrysene derivatives, J. Pharm. Biomed. Anal. 72 (2013) 231–239.
[21] C. Sarbu, D. Casoni, M. Darabantu, C. Maiereanu, Quantitativestructure–retention and retention–activity relationships of some1,3-oxazolidine systems by RP-HPTLC and PCA, J. Pharm. Biomed. Anal. 35 (2004) 213–219.
[22] L. Komsta, R. Skibinski, A. Berecka, A. Gumieniczek, B. Radkiewicz, M. Radon, Revisiting thin-layer chromatography as a lipophilicity determination tool—a comparative study on several techniques with a model solute set, J. Pharm. Biomed. Anal. 53 (2010) 911–918.
[23] D. Casoni, C. Sarbu, Comprehensive evaluation of lipophilicity of biogenic amines and related compounds using different chemically bonded phases and various descriptors, J. Sep. Sci. 35 (2012) 915–921.
[24] R.D. Briciu, A. Kot-Wasik, A. Wasik, J. Namiesnik, C. Sarbu, The lipophilicity of artificial and natural sweeteners estimated by reversed-phase thin-layer chromatography and computed by various methods, J. Chromatogr. A 1217 (2010) 3702–3706.
[25] R.D. Nascu-Briciu, C. Sarbu, A comparative study concerning the chromatographic behaviour and lipophilicity of certain natural toxins, J. Sep. Sci. 35 (2012) 1059–1067.
[26] E. Soczewinski, G. Matysik, Two types of RM-composition relations in liquid partition chromatography, J. Chromatogr. 32 (1968) 458-471.
[27] F. Murakami, Retention behaviour of benzene derivatives on bonded reversed-phase columns, J. Chromatogr. A 178 (1979) 393–399.
[28] C. Horvath, W. Melander, and I. Molnar, Solvophobic interactions in liquid chromatography with nonpolar stationary phases, J. Chromatogr A 125 (1976) 129–156.
[29] M.L. Bieganowska, A. Doraczynska-Szopa, and A. Petruczynik, The retention behavior of some sulfonamides on different TLC plates. 2. Comparison of the selectivity of the systems and quantitative determination of hydrophobicity parameters, J. Planar Chromatogr. Mod. TLC 8 (1995) 122–128.
[30] C. Sarbu, K. Kuhajda, and S. Kevresan, Evaluation of the lipophilicity of bile acids and their derivatives by thin-layer chromatography and principal component analysis, J. Chromatogr. A 917 (2001) 361–366.
[31] C. Onisor, M. Posa, S. Kevresan, K. Kuhajda, C. Sarbu, Estimation ofchromatographic lipophilicity of bile acids and their derivatives by reversed-phase thin layer chromatography, J. Sep. Sci. 33 (2010) 3110–3118.
[32] R. Mannhold, G.I. Poda, C. Ostermann, I. V. Tetko, Calculation of Molecular Lipophilicity: State-of-the-Art and Comparison of LogP Methods on More Than 96,000 Compounds, J. Pharm. Sci., 98 (2009) 861–893.

[33] C. Hansch, and A. Leo, Substituent Constants for Correlation Analysis in Chemistry and Biology, (1979) Wiley, New York, NY.
[34] A. K. Ghose, A. Pritchett, G.M. Crippen, Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships III: Modeling hydrophobic interactions, J. Comput. Chem. 9 (1988) 80.
[35] Calculation of Molecular Physicochemical Properties - Log*P* (octanol/water partition coefficient) - miLogP2.2, http://www.molinspiration.com/services/logp.html (Data accessed on December 2016)
[36] W. M. Meylan, P. H. Howard, Atom/fragment contribution method for estimating octanol-water partition coefficients, J. Pharm. Sci., 84 (1995) 83-92.
[37] R. Wang, Y. Gao, L. Lai, Calculating Partition Coefficient by Atom-additive Method, Perspect. Drug Discov. Des., 19 (2000) 47-66.
[38] T- Cheng, Y. Zhao, X. Li, F. Lin, Y. Xu, X. Zhang, Y. Li, R. Wang, Computation of Octanol-Water Partition Coefficients by Guiding an Additive Model with Knowledge, J. Chem. Inf. Model., 47 (2007) 2140-2148.
[39] M.H. Abraham, H.S. Chadha, G.S. Whiting, R.C. Mitchell, Hydrogen Bonding. 32. An Analysis of Water-Octanol and Water-Alkane Partitioning and the Δlog*P* parameter of Seiler, J. Pharm. Sci. 83 (1994) 1085-1100.
[40] M. Hornig, A. Klamt. COSMOfrag: A novel tool for high-throughput ADME property prediction and similarity screening based on quantum chemistry, J. Chem. Inf. Model., 45 (2005) 1169–1177.
[41] I. Moriguchi, S. Hirono, Q. Liu, I. Nakagome, Y. Matsushita, Simple Method of Calculating Octanol/Water Partition Coefficient, Chem. Pharm. Bull., 40 (1992) 127-130.
[42] I. V. Tetko, V. Y. Tanchuk, Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program, J. Chem. Inf. Comput. Sci., 42 (2002) 1136-1145.
[43] H. Barchanska, M. Sajdak, A. Wrobel, and C. Pieszko, Chemometric Characterization of the Liphophilicity Parameters of Triketone Herbicides and Their Degradation Products, J. Planar Chromatogr. Mod. TLC 28 (2015) 323–332.
[44] T. Djakovic-Sekulic, A. Smolinski, N. Perisic-Janjic and M. Janicka, Chemometric characterization of (chromatographic) lipophilicity parameters of newly synthesized s-triazine derivatives, J. Chemometr. 22 (2008) 195–202.
[45] M. Janicka, K. Stepnik, A. Pachuta-Stec, Quantification of Lipophilicity of 1,2,4-Triazoles Using Micellar Chromatography, Chromatographia 75 (2012) 449–456.
[46] F. Andric, K. Heberger, Chromatographic and computational assessment of lipophilicity using sum of ranking differences and generalized pair-correlation, J. Chromatogr. A 1380 (2015) 130–138.
[47] F. Andric, K. Heberger, Towards better understanding of lipophilicity: Assessment of *in silico* and chromatographic log*P* measures for pharmaceutically important compounds by nonparametric rankings, J. Pharm. Biomed. Anal. 115 (2015) 183–191.
[48] C. Hansch, A. Leo, D. Hoekman, Exploring QSAR. Hydrophobic, Electronic, and Steric Constants. ACS Professional Reference Book, American Chemical Society, Washington, DC, 1995
[49] P. De Voogt, G.A. Van Zijl, H. Govers and U.A.T. Brinkman, Reversed-phase TLC and structure-activity relationships of polycyclic (hetero) aromatic hydrocarbons, J. Planar Chromatogr.-Mod. TLC 3(1-2) (1990) 24-33.
[50] P.G.J. De Maagd, D.Th.E.M. ten Hulscher, H. van den Heuvel, A. Opperhuizen and D.T.H.M. Sijm, Physicochemical properties of polycyclic aromatic hydrocarbons: aqueous solubilities,n-octanol/water partition coefficients, and Henry's law constants, Environ. Toxicol. Chem. 17 (1998) 251–257.
[51] J. Sangster, LOGKOW Database. A databank of evaluated octanol-water partition coefficients (Log P) on microcomputer diskette, Sangster Research Laboratories, Montreal, Quebec, Canada, 1993
[52] L. Wang, X. Wang, O. Xu and L. Tian, Determination of the *n*-octanol/water partition coefficients of polycyclic aromatic hydrocarbons and estimation of their aqueous solubilities, Huanjing Kexue Xuebao, 6(4) (1986) 491-497.

[53] J. Sangster, Octanol-Water Partition Coefficients: Fundamentals and Physical Chemistry. Wiley, New York, 1994

[54] M. Henczi, J. Nagy, D.F. Weaver, Determination of octanol-water partition coefficients by an HLPC method for anticonvulsant structure-activity studies, J. Pharm. Pharmacol. 47(4) (1995) 345-347.

[55] I. V. Tetko, J. Gasteiger, J., R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V. A. Palyulin, E. V. Radchenko, N. S. Zefirov, A. S. Makarenko, V. Y. Tanchuk, V. V. Prokopenko, Virtual computational chemistry laboratory - design and description, J. Comput. Aid. Mol. Des., 19 (2005) 453-463

[56] I. V. Tetko, Computing chemistry on the web, Drug Discov. Today, 10 (2005) 1497-1500.

[57] K. Héberger, Sum of ranking differences compares methods or models fairly, Trends Anal. Chem. 29 (2010) 101-109.

[58] K. Héberger, K. Kollár-Hunek, Sum of ranking differences for method discrimination and its validation: comparison of ranks with random numbers, J. Chemometr. 25 (2010) 151–158.

[59] T. Hastie, R. Tibshirani and J. Friedman, in The Elements of Statistical Learning; Data Mining, Inference, and Prediction, Springer, New York, 2001, p. 31.

[60] D. Bajusz, A. Rácz, Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? J. Cheminformatics. 7 (2015), 20.

[61] K. Héberger, R. Rajkó, Generalization of pair correlation method (PCM) for non-parametric variable selection, J. Chemometr. 16 (2002) 436–443.

[62] A. Rácz, D. Bajusz, K. Héberger, Consistency of QSAR models: Correct split of training and test sets, ranking of models and performance parameters, SAR and QSAR in Environmental Research 26 (2015) 683–700.

[63] P. Gramatica, External Evaluation of QSAR Models, in Addition to Cross-Validation: Verification of Predictive Capability on Totally New Chemicals, Mol. Inf. 33 (2014) 311–314.