

Accepted Manuscript

Title: Linear modeling of the soil-water partition coefficient normalized to organic carbon content by reversed-phase thin-layer chromatography

Author: Filip Andrić Sandra Šegan Aleksandra Dramićanin
Helena Majstorović Dušanka Milojković-Opsenica



PII: S0021-9673(16)30839-1
DOI: <http://dx.doi.org/doi:10.1016/j.chroma.2016.06.063>
Reference: CHROMA 357690

To appear in: *Journal of Chromatography A*

Received date: 19-2-2016
Revised date: 13-6-2016
Accepted date: 18-6-2016

Please cite this article as: Filip Andrić, Sandra Šegan, Aleksandra Dramićanin, Helena Majstorović, Dušanka Milojković-Opsenica, Linear modeling of the soil-water partition coefficient normalized to organic carbon content by reversed-phase thin-layer chromatography, *Journal of Chromatography A* <http://dx.doi.org/10.1016/j.chroma.2016.06.063>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Linear modeling of the soil-water partition coefficient normalized to organic carbon content by reversed-phase thin-layer chromatography

Filip Andrić^{1*}, Sandra Šegan², Aleksandra Dramićanin³, Helena Majstorović⁴, Dušanka Milojković-Opsenica¹

¹Faculty of Chemistry, University of Belgrade, Studentski trg 12-16, P.O. Box 51, 11158 Belgrade, Serbia

²Institute of Chemistry, Technology and Metallurgy, Njegoševa 12, 11000 Belgrade, Serbia

³Innovation center Faculty of Chemistry Ltd, Studentski trg 12-6, 11000 Belgrade, Serbia

⁴AIC „Tamiš“, Novoseljanski put 33, 26000 Pančevo, Serbia

*Correspondence should be addressed to:

Filip Lj. Andrić, Ph.D., assistant professor

Phone: +381 11 3336755

E-mail: andric@chem.bg.ac.rs

Highlights

- Thin-layer chromatographic methods for modeling soil-sorption are proposed
- Methods performs equally well as typical in silico estimators
- CN-silica and MeOH-water mixtures were selected as the most suitable systems

Abstract

Soil-water partition coefficient normalized to the organic carbon content (K_{OC}) is one of the crucial properties influencing the fate of organic compounds in the environment. Chromatographic methods are well established alternative for direct sorption techniques used for K_{OC} determination. The present work proposes reversed-phase thin-layer chromatography (RP-TLC) as a simpler, yet equally accurate method as officially recommended HPLC technique.

Several TLC systems were studied including octadecyl- (RP18) and cyano- (CN) modified silica layers in combination with methanol-water and acetonitrile-water mixtures as mobile phases. In total 50 compounds of different shape, molecular size, and various ability to establish specific interactions were selected (phenols, beznodiazepines, triazine herbicides, and polyaromatic hydrocarbons). Calibration set of 29 compounds with known $\log K_{OC}$ values determined by sorption experiments was used to build simple univariate calibrations, Principal Component Regression (PCR) and Partial Least Squares (PLS) models between $\log K_{OC}$ and TLC retention parameters. Models exhibit good statistical performance, indicating that CN-layers contribute better to $\log K_{OC}$ modeling than RP18-silica. The most promising TLC methods, officially recommended HPLC method, and four *in silico* estimation approaches have been compared by non-parametric Sum of Ranking Differences approach (SRD). The best estimations of $\log K_{OC}$ values were achieved by simple univariate calibration of TLC retention data involving CN-silica layers and moderate content of methanol (40 - 50% v/v). They were ranked far well compared to the officially recommended HPLC method which was ranked in the middle. The worst estimates have been obtained from *in silico* computation based on octanol-water partition coefficient.

Linear Solvation Energy Relationship study revealed that increased polarity of CN-layers over RP18 in combination with methanol-water mixtures through significant diminishing dipolar and proton accepting influence of the mobile phase as well as enhancing molar refractivity in excess of the chromatographic systems is the key to better modelling of $\log K_{OC}$.

Keywords: Soil-water partition coefficient; Reversed-phase thin-layer chromatography (RP-TLC); Benzodiazepines; Polycyclic aromatic hydrocarbons (PAHs); Multivariate regression methods; Sum of Ranking Differences (SRD)

1 Introduction

Sorption in the soil-water compartment is one of the most important factors that determine the fate of organic compounds in the environment. It controls their availability to transport and degradation processes in aqueous phase; involvement in catalytic reactions on the adsorption sites in the soil [1,2], influence volatilization, bioavailability, biodegradability, photolysis, and hydrolysis of pollutants [1]. It is one of the major factors considered in removal of toxic organic compounds in the waste water treatment facilities [3].

At present, the concept based on the soil-water partition coefficient normalized to organic carbon content (K_{OC}) has been widely used for the assessment of fate of compounds in the soil-water compartment including broad range of different soils and diverse classes of substances [1]. The reason for its widespread application lies in its direct link to the content of the soil organic matter (SOM) [2], and ability to account mostly for structural features of an analyte while significantly eliminating the effects of soil diversity.

Measurement of K_{OC} by direct sorption experiments in biphasic soil-water systems [4-6], is time and reagent consuming, tedious and subject to various difficulties and artifacts. Methods are inapplicable to compounds sparingly soluble in water, significant volatiles or those with strong sorption affinity towards soil organic matter or silicates [1,2]. Therefore, chromatographic methods have a central place in rapid indirect determination of K_{OC} . Mostly typical reversed-phase (RP) modalities have been used so far, with different combination of sorbents such as: chemically bonded octadecyl silica (RP-18) [7-9], cyano-modified silica (CN) [7-15], tetramethylammonium- (TMA) [9], aminopropyl- [11], as well as humic acid-modified silica [11,16-18]. In addition, soil materials have been used as sorbents with great success [19-21] providing valuable insights into mobility of analytes in different soils [22]. Chromatographic methods provide much coherent results, work in wider $\log K_{OC}$ range, and

are less time and reagent consuming. Therefore, they have been implemented in official guidelines for testing chemicals of the Organization for Economic Cooperation and Development (OECD) [23].

In addition to experimental techniques, computational estimation methods have been frequently employed. Most of them are relying on physicochemical properties such as octanol-water partition coefficient (K_{OW}) [24-26], water solubility ($\log S$) [24,27] or Quantitative Structure Property Relationships (QSPR) models based on structural, topological, electronic and other 1D molecular descriptors [10,28], as well as Linear Solvation Energy Relationships (LSER) [29,30]. While representing an efficient alternative for $\log K_{OC}$ assessment, computational approaches have significant drawbacks. They often do not differentiate among various forms of structural isomers, nor account for the influence of pH on the sorption of weak bases and acids [28]. Models based on $\log S$ and $\log K_{OW}$ are specific to compound subclass, while LSER models require experimental determination of solvatochromic parameters.

The ever increasing concern for the fate of pollutants and pharmaceuticals in the environment put constant demands on development of novel, cheaper, faster and yet reliable ways for determination of $\log K_{OC}$. In that sense TLC has much to offer. Being capable to process great number of samples in a short period of time, TLC does not require any advanced equipment, and consumes low amounts of solvents and reagents. Performed in a reversed-phase mode, and relying on the use of environmentally friendly mobile phases such as ethanol-water mixtures, TLC becomes favorable in terms of green chemistry.

So far, with exception of our previous study [31], no TLC methods have been systematically assessed for the determination of $\log K_{OC}$. However, our previous work [31] was focused on limited number of phenolic compounds. In the meantime, diversity and

number of compounds that we have investigated have significantly increased providing us with new insights, and possibility to make more general conclusions related to: (i) applicability of RP-TLC to compounds with various abilities to exhibit specific interactions and (ii) establishing the application limit values of the $\log K_{OC}$. In that sense, the aim of the present work is: (i) the systematic assessment of RP-TLC in the $\log K_{OC}$ modeling, including comparison with officially recommended HPLC method and *in silico* estimation approaches, and (ii) the determination of $\log K_{OC}$ values for environmentally and pharmaceutically important compounds for which there are no K_{OC} values determined yet.

2 Material and methods

2.1 Selection of the target set of compounds

A representative set of 50 compounds of various molecular structures, shapes and sizes, of significant environmental and pharmaceutical importance have been selected (Table 1). Out of that 29 being standard substances (5 phenols, 9 polyaromatic hydrocarbons (PAH), 4 triazine herbicides, and other aromatic amines, ketones, and aldehydes) with known, experimentally determined $\log K_{OC}$ values taken from the KOCWIN database (EPI Suite, EPA, USA) and 21 compounds with unknown values, among them 9 benzodiazepines and 7 phenols. Standard solutes have been chosen to cover a wide range of the soil-water sorption coefficient (1.10 - 6.22 log units), with diverse abilities to form hydrogen bonds, dipolar or polarizable interactions given as Abraham's solvatochromic parameters: *A* - hydrogen bond acidity: 0.00 – 0.94; *B* – hydrogen bond basicity: 0.15 – 1.63, *S* – dipolar interactions: 0.79 – 2.49, *E* – molar refractivity in excess: 0.80 – 3.43, and *V* – McGowan's molecular volume: 0.78 to 2.19 (Table S1, Supplementary material). Experimentally determined solvatochromic parameters have been collected through freely available Absolve online calculation tool, part

of the ACD/I-Lab (<https://ilab.acdlabs.com/iLab2/>).

All substances, with exceptions of 4-nitrophenol, 2,4,6-trichlorophenol, and 4-aminobenzoic acid, are considered to be present in the neutral form under the studied chromatographic conditions, *i.e.* ionization degree $\alpha < 5.00\%$ at pH = 6. The entire list of compounds followed by $\log K_{OC}$ values, Abraham's solvatochromic parameters, p*K*_a values, and retention factors is given in the Supplementary material (**Tables S1-S2**) as well as in the Supplementary Excel data sheets.

Table 1

2.2 The TLC experiments

Chromatographic experiments have been performed using commercially available RP18 and CN silica coated on alumina sheets and glass plates (Art. Nos. 5559 and 16464 respectively; Merck, Darmstadt, Germany). Methanol - water and acetonitrile - water binary mixtures were used as mobile phases, varying the content of organic solvents in the range 40 - 80% v/v with an increment of 5%. Size of the plates was 10 × 10 cm. Approximately 0.2 – 0.3 μL of freshly prepared solutions of target compounds in acetone ($C \approx 0.01 \text{ mol L}^{-1}$) were manually applied at 5 mm distance from the lower edge of the plate. Chromatograms were developed in a horizontal HPTLC chamber (CAMAG, Mutenz, Switzerland). The development distance was about 4.5 cm and position of each single zone was detected under UV-light (254 nm). The zone distances were manually measured and retention mobility values, R_M are calculated according to Eq. 1. All experiments were performed at ambient temperature ($22 \pm 2 \text{ }^\circ\text{C}$).

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

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

2.3 The HPLC experiment

The HPLC experiments were conducted in accordance with the current OECD procedure [23], using the Waters 1525 HPLC dual pump system, equipped with the Alltech, Select™ degasser system and dual λ 2487 UV–VIS detector. A cyanopropyl column (Waters Spherisorb S10 CN, 4.6×150 mm analytical column, Ser. No. 0103141281K03) was used as the stationary phase. Isocratic elution with a mixture of methanol (HPLC grade, Merck) and citric buffer, $C = 0.01 \text{ mol L}^{-1}$, $\text{pH} = 6$ in a ratio of 30:70, v/v, was employed. Although the OECD guideline [23] recommends the usage of 40% or 45% v/v of organic modifier in the mobile phase, the studied compounds have been eluted with relatively short retention times under such conditions; therefore, the elution strength of the mobile phase was lowered, providing more reliable retention data. The flow rate was maintained at 1 mL min^{-1} . Each run was performed at room temperature ($22 \pm 2 \text{ }^\circ\text{C}$). Sample solutions were prepared by dissolving the pure substance in the mobile phase, in a concentration of 1.0 mmol L^{-1} , with addition of 0.42 mmol L^{-1} of potassium bromide as the holdup volume marker ($\lambda = 220 \text{ nm}$, separate detection channel) and injected through the injection loop of $10 \text{ }\mu\text{L}$. Dual wavelength mode was used for detection. The retention time, t_R , and holdup time, t_0 , were determined for each compound as the mean values of triplicate measurements. Finally, the resulting retention factors were calculated according to the following equation:

$$k = \frac{t_R - t_0}{t_0} \quad (2)$$

2.4 *In silico* estimation of $\log K_{OC}$ values

We have decided to use the four most commonly employed computational approaches for estimation of $\log K_{OC}$ -s. Two are LSER models, one developed by Nguyen *et al.* [30] and the other reported by Poole and Poole [30] (Eqs. 3 and 4, respectively)

$$\log K_{OC} = (0.14 \pm 0.10) + (0.15 \pm 0.15)A - (1.98 \pm 0.14)B - (0.72 \pm 0.14)S + (1.10 \pm 0.10)E + (2.28 \pm 0.14)V \quad R^2 = 0.98, \text{ error} = 0.18, n = 75 \text{ points}, N = 356 \text{ measurements} \quad (3)$$

$$\log K_{OC} = (0.19 \pm 0.10) - (0.23 \pm 0.10)A - (2.33 \pm 0.12)B + (0.72 \pm 0.05)E + (2.12 \pm 0.11)V \quad R^2 = 0.954, S.D. = 0.249, F = 585, n = 119 \quad (4)$$

The other two approaches are implemented in the KOCWIN software which is part of the public database (EPI Suite, v. 4.1, U.S. EPA). The first one is based on the first order molecular connectivity indices (MCI) while the second one uses octanol-water partition coefficient (K_{OW}). MCI based method relies on already described methodology by Meylan *et al.* [29], with a difference that significantly larger database was used in the present version of the software compared to the originally published data. Both methods use the primary (uncorrected) models for non-polar compounds (Eqs. 5 and 7). Estimations of $\log K_{OC}$ for polar compounds require correction factors (Eqs. 6 and 8).

$$\log K_{OC} = 0.5213 MCI + 0.60, n = 69, r^2 = 0.967, S.D. = 0.247 \quad (5)$$

$$\log K_{OC} = 0.5213 MCI + 0.60 + \Sigma P_f N \quad (6)$$

$$\log K_{OC} = 0.8679 \log K_{OW} - 0.0004, n = 68, r^2 = 0.877, S.D. = 0.478 \quad (7)$$

$$\log K_{OC} = 0.55313 \log K_{OW} + 0.9251 + \Sigma P_f N \quad (8)$$

where $\Sigma P_f N$ is the summation of the products of all applicable correction factor coefficients (P_f) multiplied by the number of times (N) that factor is counted for the structure.

2.5 Modeling and data analysis

Linear modeling (Principal Component Regression – PCR, and Partial Least Squares regression – PLS) and Principal Component Analysis – PCA (part of exploratory data analysis) were done using PLS Tool Box (v. 7.02) for MATLAB (R2011). Prior to PCA, PCR, and PLS the data were standardized, *i.e.* mean-centered and expressed to the unit standard deviation. The number of principal components in PCA was decided based on the visual inspection of the scree plot (dependence between the data variance explained by each principal component and the number of component). A clear cut-off value that separates a region with steep decrease of variance from a rather flat part has been used as an indicator of the number of relevant PC-s that should be retained. The PCR and the PLS models accompanied with predictive performance parameters were built using double cross-validation (CV) approach as described by Varmuza and Filzmoser [34]. The entire data set has been split into four independent training and test sets as a part of the outer cross-validation loop using venetian blinds (VB) resampling strategy. Each training set has been further split into five calibration and validation sets as a part of the inner cross-validation loop, using the VB resampling. In that way each compound has been used as a part of the calibration, validation, and test set, however, never in the same time. The optimal model complexity was determined based on the model performance criteria obtained from the inner loop CV calculations, *i.e.*, the number of latent variables that lead to models with minimum of the root mean square cross-validation errors ($RMSE_{CV}$). The performance parameters calculated from the outer loop were used as the estimates of prediction ability ($RMSE_{PRED}$ and R^2_{PRED} , respectively).

Simple univariate and multivariate linear regressions were performed using Microsoft Excel 2010.

Comparisons and ranking of chromatographic and computational approaches were done

with the Sum of Ranking Differences (SRD), an approach that compares methods fairly [35,36]. User friendly SRD algorithm was provided in a form of Microsoft Excel visual basic macros freely available at <http://aki.ttk.mta.hu/srd/>. In order for methods to be compared by SRD a reference is required. Variables, in this case $\log K_{OC}$ determination approaches, and objects (compounds) are arranged in a form of a matrix, *i.e.* columns and rows, respectively. The objects are then ranked for each variable separately; ranks are subtracted from the reference ranks, and summed up giving unique SRD value for each variable. The SRD values are usually scaled to the range between 0 and 100, and variables are arranged in ascending order of SRD-s. The lower the SRD, the closer is the variable to the reference. The reference can be a standard method, or row-wise calculated minimum, maximum, or arithmetic mean average. In the present case we have decided to use the average. Such consensus based approach is justified from the two main points: (i) 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) and (ii) all random and systematic errors of the methods are canceled out to some extent by averaging.

3 Results and discussion

3.1 Exploratory data analysis

In order to reveal the presence of outlying effects in retention behavior of studied compounds, possible groupings and similarities among used chromatographic systems, as well as to illustrate representativeness of the selected sets of compounds, PCA has been performed on standardized retention data.

A relatively heterogeneous nature of the entire set of compounds can be noticed (**Figure 1a**). The *PC1* vs. *PC2* score plot show the presence of at least two groups. The first one is

located at the right side of the plot consisting of mostly polyaromatic hydrocarbons (compounds 7, 18-25), while the benzodiazepines and the rest of phenols and aromatic compounds are located at the left side of the graph. Two compounds: *p*-aminobenzoic acid (17) and *p*-anisidine (11) fall outside of the Hotelling 95% confidence curve. Although the outlying effect is not severe, in the case of *p*-aminobenzoic acid it could be attributed to extensive dissociation under chromatographic conditions ($\alpha = 95.6\%$).

Nevertheless, the standard compounds and unknowns are evenly distributed in the PC retention score space, which insures deduction of general conclusions applicable to entire data set. The loading plot of retention data (**Figure 1b**) reveals distinction in chromatographic conditions alongside the *PC2* direction which accounts for fine data variability (4.23%). Chromatographic systems based on RP18 silica have positive loadings, while the cyano-modified sorbents are negative. Obviously that different constellation of interactions alters the retention on stationary phase with polar moieties vs. hydrophobic long hydrocarbon chains.

Figure 1

3.2 Modeling of retention - $\log K_{OC}$ relationships

In order to build calibration models of $\log K_{OC}$ vs. retention data, two directions were considered. One is a simple univariate approach that provides a single calibration model for each chromatographic system. The other assumes building multivariate models based on the overall information from all chromatographic experiments. Univariate approach is simpler; however, the multivariate calibration has more power provided by simultaneous treatment of multiple variables. Because of its simplicity, a linear modeling is of primary concern. Also, linear addition of R_M values can be interpreted as a linear free energy relationship. When it comes to highly correlated chromatographic data PCR and PLS regressions are the most frequently applied methods [37-39]. Although, the methods are related to each other, they

have substantial differences. The PLS regression does not form latent variables only by maximizing variability of the scores in the independent variable domain, which is a key concept in PCR, but rather maximizing correlation between the projection scores in both, independent and dependent, variable domains. Therefore it was of particular interest to compare both approaches alongside with much simpler univariate calibration methods that do not require many chromatographic experiments.

All 29 standard compounds were used for building up calibration models.

In the case of PLS and PCR, a double cross-validation procedure resulted in two latent variable models (optimal complexity). A total of 19 univariate TLC calibrations are summarized in the Supplementary material (**Table S3**). Their prediction performance was estimated by 4 split cross-validation experiments in combination with VB resampling strategy. Only seven of them have satisfactory statistical parameters (errors 0.4 - 0.6 log units, and R^2_{Cal} and $R^2_{\text{PRED}} > 0.8$). They are included in the **Table 2** (models 3 – 9) together with the HPLC calibration, PCR and PLS regression models.

All chromatographic models (**Table 2**) demonstrate similar and fairly good predictive performance with the $RMSE$ values in the following ranges: $RMSE_{\text{Cal}} = 0.401 - 0.537$, $RMSE_{\text{CV}} = 0.496 - 0.517$; $RMSE_{\text{PRED}} = 0.441 - 0.608$ and coefficients of determination: $R^2_{\text{Cal}} = 0.8340 - 0.9124$, $R^2_{\text{CV}} = 0.8640 - 0.8585$; $R^2_{\text{PRED}} = 0.8140 - 0.8878$. All residual values are normally distributed without presence of any trends, (tested by Kolmogorv-Smirnov, Shapiro-Wilk's, and Lilliefors' tests, $p > 0.20$ in all cases). This confirms that linear approach is able to correctly model the K_{OC} data in approximate range of 1 – 6 log units, with the prediction error not exceeding 0.4 – 0.6 log units.

Table 2

Statistical performance parameters show that TLC models perform equally well as the officially recommended HPLC method. Furthermore, the qualities of selected TLC models are comparable with HPLC univariate calibrations published by Koerdel and coworkers [18], Gawlik et al. [15], and Szabo and coworkers [11].

PCR and PLS variable diagnostic plots, *i.e.*, regression vector plots, sensitivity ratio graphs and variable importance to projection (VIP) diagrams are included in the Supplementary material (Figures S1 and S2). They illustrate stronger contribution of CN-silica layers over the RP18-modified ones to the overall modeling of $\log K_{OC}$. Short discussion is appended in the Supplementary material, page 8.

3.3 Determination of $\log K_{OC}$ values and comparison of chromatographic and in silico approaches

The $\log K_{OC}$ values for 31 compounds that lack experimental data as well as 29 standards were determined using chromatographic models presented in **Table 2**. The values are presented together with computationally estimated $\log K_{OC}$ -s in the **Table S4** (Supplementary material).

In order to identify the best and the worst $\log K_{OC}$ determination method non-parametric comparison by the SRD was applied on the entire set of $\log K_{OC}$ values. The SRD method has been already successfully employed to rank and group variables, finding statistically significant differences even if the variables are highly correlated [40-45], which is the case with the present set of $\log K_{OC}$ values.

In a consensus based comparison methods are stacked in the narrow range of SRD scores (9 – 22, out of 100). Validation of ranking done by comparison with random number distribution of SRD-s (CRRN) shows that all methods are positioned far from the 95% confidence range

of the bell-shaped random SRD distribution curve, which implies that all of them rank studied far better than a statistical chance.

Two univariate calibration TLC models corresponding to CN-modified silica in combination with moderate methanol content of 40% and 50% v/v, are selected as the best $\log K_{OC}$ estimates (closest to the average) (**Figure 2a**). They are followed by the LSER model proposed by Poole, univariate TLC calibrations obtained on RP18-silica using higher content of methanol (80% and 70% v/v) and CN-silica combined with moderate methanol content (60% v/v).

Figure 2

In order to assess statistical significance of differences in SRD-s, variability was introduced into dataset by the sevenfold jack-knife resampling procedure. Approximately 1/7 of objects are removed and the ranking is performed on the remaining data set. Procedure is repeated seven times producing seven SRD values for each $\log K_{OC}$ estimation method. Results are plotted in a form of a box-and-whisker diagram using median, maximum, minimum, and interquartile ranges of SRD values (**Figure 2b**). The median SRD values, with exception of the CN/MeOH-70, follow increasing order (**Tables S5a** and **S5b**, Supplementary material). Statistical difference among each pair of variables is then tested by applying the Wilcoxon's matched pair test, and the sign test, which are able to group (separate) methods into four distinct sections (denoted by vertical dashed lines). Univariate TLC calibration models obtained on CN-modified silica in combination with moderate content of methanol in the mobile phase (40% and 50% v/v, respectively) have the lowest and statistically indistinguishable SRD medians (the first group). From at least two points, these methods should be considered the best ones and the simplest to perform. First of all, by providing $\log K_{OC}$ -s that are closest to the consensus values (the average arithmetic mean), they are the

most accurate. Second of all, they can efficiently substitute the rest of compared approaches. This is particularly important when it comes to PLS and PCR models that require several chromatographic experiments in order to increase the estimation accuracy and precision. The four methods that follow, *i.e.*, LSER (Poole), TLC (RP18/MeOH 70), TLC (RP18/MeOH 80), and TLC (CN/MeOH 60), are placed together in the second group according to both significance tests. The recommended HPLC method is in the third group and performs equally well as the remaining TLC univariate and multivariate (PLS and PCR) calibrations, LSER (Nguyen) approach, and EPI Suite estimations based on MCI-s. Therefore, it can be effectively replaced with the simplest method to perform, *e.g.* TLC experiment involving CN-modified silica in combination with higher content of methanol (70% v/v), or *in silico* method.

The lowest ranked is the EPI Suite method based on K_{OW} estimations. The method differs statistically significantly from the rest of approaches, and should be considered as the least trusted.

3.4 Impact of stationary and mobile phase on modeling $\log K_{OC}$ (CN- vs. RP18-modified silica layers)

According to the SRD ranking it is obvious that different chromatographic systems result in variously reliable $\log K_{OC}$ estimates. Among them CN-modified silica in combination with moderate content of methanol provides the best estimates.

In order to get insights into interactions governing retentions in the studied chromatographic systems LSER models using Abraham's solvatochromic parameters of all 50 compounds, as independent variables, and retention factors as dependent ones, were built for each TLC system by multiple linear regression. Models are expressed in the following form.

$$R_M = I + aA + bB + eE + sS + vV \quad (9)$$

Intercept, and regression coefficients a , b , e , s , and v are summarized in the **Table S6** (Supplementary material) with accompanying statistics. They describe the overall effects of hydrogen bond basicity - a , hydrogen bond acidity - b , excess molar refractivity/polarizability - e , dipolarity - s , and ability to form of a vacant space - v in the chromatographic system.

In all TLC systems regression coefficients follow pattern that is typical for reversed-phase environment, *i.e.* negative values of a , b , and s , positive values of e and v , and the highest absolute values of a , b , and v . [46,47] This clearly demonstrates the significant role of aqueous mobile phase and organic modifier, their strong contribution to formation of hydrogen bonds with solute molecules, and significant resistance to formation of an empty space for incorporation of a transferring molecule from stationary phase [30,46,47].

Figure 3

PCA score plot of LSER regression coefficients a , b , s , e , and v of all TLC systems, as well as Nguyen's and Pool's $\log K_{OC}$ models indicates that TLC systems based on CN-silica, especially those with 40 - 60% v/v of methanol, as well as those involving RP18 in combination with higher content of methanol (70 - 80% v/v), are the closest to Nguyen's and Pool's $\log K_{OC}$ models (**Figure 3a**). Diagram of loadings (**Figure 3b**) suggests that polarizability in excess, e , and proton basicity, a , are the most responsible factors for such disposition of LSER models.

Dissimilarities and similarities among two contrasting LSER models i and j are further calculated using Euclidian distance [49] and Ishihama's and Asakawa's [48,49] cosine similarity measure ($\cos \theta$) (Eqs. 10 and 11, respectively).

$$d = \sqrt{(a'_i - a'_j)^2 + (b'_i - b'_j)^2 + (s'_i - s'_j)^2 + (e'_i - e'_j)^2 + (v'_i - v'_j)^2} \quad (10)$$

$$\cos \theta_{ij} = a'_i a'_j + b'_i b'_j + s'_i s'_j + e'_i e'_j + v'_i v'_j \quad (11)$$

Where a', b', s', e', v' means regression coefficients, each divided by the vector length ($l = [a^2 + b^2 + s^2 + e^2 + v^2]^{1/2}$).

It is clear that the two parameters are interrelated by a simple trigonometry:

$$\cos \theta_{ij} = \cos (2\arcsin(d_{ij}/2)) \quad (12)$$

Obviously, the closer the d is to zero and $\cos \theta$ to unit, the more similar LSER models are. In this particular case chromatographic systems resulting in the LSER models the most similar to the general models of Nguyen and Poole, are governed with the same constellation of interactions responsible for the modeling of the soil-water partitioning ($\log K_{OC}$). Their values of d and $\cos \theta$ are summarized in the **Table 3**. According to Lazaro *et al.* [48] values of d lying in the range of 0 – 0.2 units suggest a negligible difference among LSER systems.

TLC systems closest to the Poole's general LSER model ($d < 0.2$) are those involving CN-silica and methanol content of 40 and 50% v/v, as well as those based on RP18-silica and higher content of methanol (70 - 80%). In addition, CN-beds in combination with moderate content of acetonitrile 40 – 50% v/v are among the closest to the Nguyen's general $\log K_{OC}$ model ($d = 0.144$ and 0.177 , respectively).

Examination of LSER coefficients of above mentioned TLC systems indicates that polarizability in excess, e , makes these systems statistically significantly different from the others (with considerably higher values: 0.352, 0.349, and 0.313, respectively).

Similar findings were reported by Poole and coworkers [50] who find a TLC system using CN-silica and methanol-water mixture (3:2 v/v), among several promising chromatographic candidates for modeling soil-water partition.

4 Conclusions

Modeling of $\log K_{OC}$ with TLC retention parameters by simple univariate calibration method, PCR and PLS regression resulted in statistically well established relationships with good predictive performance (prediction errors not exceeding the value of 0.6 log units).

Comparing TLC models with officially recommended HPLC approach, as well as with the most frequently used *in silico* estimation approaches results in the highly coherent SRD values ranging from 9 to 22 (out of 100) units, and falling far away from the random distribution. The best approaches, resulting in the $\log K_{OC}$ values the closest to the consensus values, are the simple univariate TLC calibrations involving CN-silica in combination with moderate methanol content (40 and 50% v/v). Being the closest to the average arithmetic mean, these methods have the lowest bias and can substitute all others, including officially recommended HPLC method, as well as *in silico* approaches.

The main advantages of the proposed methods compared to HPLC method recommend by OECD are significantly reduced costs; possibility of high throughput sample analyzes; remarkably lower consumption of reagents and solvents which makes them much favorable in terms of green analytical chemistry.

The methods are simple to perform and are applicable to small organic molecules that have $\log K_{OC}$ values in the range of 1 – 6 log units. Compounds should not be present in ionized form (< 5%) under the chromatographic conditions (pH \approx 5-6), and are supposed to be able to establish specific interactions within the following limits of the Abraham's solvatochromic parameters: $A = 0.00 - 1.00$; $B = 0.15 - 1.7$; $S = 0.80 - 2.50$; $E = 0.80 - 3.40$; and $V = 0.80 - 2.20$.

LSER analysis revealed that increased polarity of CN-modified layers over RP18 in combination with methanol-water mixtures, especially regarding the ability to significantly diminish dipolar and proton accepting influence of the mobile phase on the overall retention

process and enhances molar refractivity in excess of the chromatographic system, is most likely the key to better modelling of $\log K_{OC}$.

5 Acknowledgments

This work has been supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, Grant no. 172017. Dr. Filip Andrić would like to express his gratitude to Prof. Károly Héberger for constructive suggestions regarding data analysis.

References

- [1] W. J. Doucette, Quantitative structure-activity relationships for predicting soil-sediment sorption coefficient for organic chemicals, *Annu. Rev., Environ. Toxicol. Chem.* 22 (8) (2003) 1771-1788.
- [2] A. Delle Site, Factors affecting sorption of organic compounds in natural sorbent/water systems and sorption coefficients for selected pollutants, *Annu. Rev., J. Phys Chem. Ref. Data*, 30 (2001) 187-439.
- [3] R. A. Dobs, L. Wang, R. Govind, Sorption of toxic organic compounds on wastewater solids; Correlation with fundamental properties, *Environ. Sci. Technol.* 23 (1989) 1092-1097.
- [4] Organization for Economic Co-Operation and Development, OECD Guidelines for the Testing of Chemicals, Adsorption - Desorption Using a Batch Equilibrium Method, Test No. 106: OECD, (2001).
- [5] Organization for Economic Co-Operation and Development, OECD Guidelines for the Testing of Chemicals, Leaching in Soil Columns, Test No. 312: OECD, (2004).
- [6] H. de Jonge, T. J. Heimovaara, J. M. Verstraten, Naphthalene sorption to organic soil materials studied with continuous stirred flow experiments, *Soil Sci. Soc. Am. J.* 63 (1999) 297–306.
- [7] P. D. Vowles, R. F. C. Mantoura, Sediment-water partition coefficients and HPLC retention factors of aromatic hydrocarbons, *Chemosphere.* 16 (1) (1987) 109–116.
- [8] J. Hodson, N. A. Williams, The estimation of the adsorbent coefficient (KOC) for soils by high performance liquid chromatography, *Chemosphere.* 17 (1) (1988) 67–77.

- [9] W. Koerdel, J. Stutte, G. Kuthoff, HPLC- screening method for the determination of the adsorption coefficient on soil – comparison of different stationary phases, *Chemosphere*. 27 (1993) 2341–2352.
- [10] M. Müller, W. Kördel, Comparison of screening methods for the estimation of adsorption coefficients on soil, *Chemosphere*. 32 (1996) 2493–2504.
- [11] G. Szabo, J. Guzzi, W. Kordel, A. Zsolnay, V. Major, P. Keresztes, Comparison of different HPLC stationary phases for determination of soil-water distribution coefficient, K_{oc} , values of organic chemicals in RP-HPLC system, *Chemosphere*. 39 (3) (1999) 431–442.
- [12] W. Kordel, D. Hennecke, M. Hermann, Application of the HPLC-screening method for the determination of the adsorption coefficient on sewage sludges, *Chemosphere*. 35 (1/2) (1997) 121–128.
- [13] W. Kordel, G. Kotthoff, J. Muller, HPLC-screening method for the determination of the adsorption coefficient on soil – results of a ring test, *Chemosphere*. 30 (7) (1995) 1373–1384.
- [14] B. M. Gawlik, E. A. Feicht, W. Karcher, A. Kettrup, H. Muntau, Application of the European reference soil set (EUORSOILS) to a HPLC-screening method for the estimation of soil adsorption coefficients of organic compounds, *Chemosphere*. 36 (14) (1998) 2903–2919.
- [15] B. M. Gawlik, A. Kettrup, H. Muntau, Estimation of soil adsorption coefficients of organic compounds by HPLC screening using the second generation of the European reference soil set, *Chemosphere*. 41 (2000) 1337–1347.
- [16] G. Szabo, S. Lesley Prosser, R. A. Bulman, Determination of the adsorption coefficient (K_{oc}) of some aromatics for soil by RP-HPLC on two immobilized humic acid phases, *Chemosphere*. 21 (6) (1990) 777–788.

- [17] G. Szabo, S. Lesley Prosser, R. A. Bulman, Prediction of the adsorption coefficient (K_{OC}) for soil by a chemically immobilized humic acid column using RP-HPLC. *Chemosphere*. 21 (6) (1990) 729–739.
- [18] W. Kordel, J. Stutte, G. Kotthoff, HPLC-screening method to determine the adsorption coefficient in soil-comparison of immobilized humic acid and clay mineral phases for cyanopropyl columns, *Sci. Total Environ*. 162 (1995) 119–125.
- [19] F. Xu, X. Ling, B. Lin, K.-W. Schramm, A. Kettrup, Estimation of soil organic partition coefficients: from retention factors measured by soil column chromatography with water as eluent, *J. Chromatogr. A*. 968 (2002) 7–16.
- [20] R. Guo, X. Liang, J. Chen, W. Wu, Q. Zhang, D. Martens, A. Kettrup, Prediction of soil organic carbon partition coefficients by soil column liquid chromatography, *J. Chromatogr. A*. 1035 (2004) 31–36.
- [21] F. Xu, X. Liang, B. Lin, F. Su, K.-W. Schramm, A. Kettrup, Soil column chromatography for correlation between capacity factors and soil organic partition coefficients for eight pesticides, *Chemosphere*. 39 (13) (1999) 2239–2248.
- [22] M. Mastrocicco, N. Colombani, A. Cavazzini, L. Pasti, A green and fast chromatographic method for determining organic compound mobility in soils, *J. Chromatogr. A*. 1216 (2009) 6802–6809.
- [23] Organization for Economic Co-operation and Development (OECD). OECD Guidelines for Testing of Chemicals, Estimation of the Adsorption Coefficient (K_{oc}) on Soil and on Sewage Sludge using High Performance Liquid Chromatography (HPLC), Test 121, (2001).
- [24] B. M. Gawlik, N. Sotiriou, E. A. Feicht, S. Schulte-Hostede, A. Kettrup, Alternatives for the determination of the soil adsorption coefficient, K_{OC} , of non-ionic organic compounds – a review, *Chemosphere*. 34 (12) (1997) 2525–2551.

- [25] J. R. Baker, J. R. Mihelcic, D. C. Luehrs, J. P. Hickey, *Water Environ. Res.* 69 (1997) 136.
- [26] A. Sabljčić, H. Gusten, H. Verhaar, J. Joop Hermens, QSAR modeling of soil sorption improvements and systematics of log KOC vs. log KOW correlations, *Chemosphere*. 31 (1995) 4489–4514.
- [27] Z. Gerstl, Estimation of organic chemical sorption by soil, *J. Contam. Hydrol.* 6 (1990) 357–375.
- [28] W. Meylan, P. H. Howard, R. S. Boethling, Molecular topology/fragment contribution method for predicting soil sorption coefficients, *Environ. Sci. Technol.* 26 (1992) 1560–1567.
- [29] T. H. Nguyen, K. Goss, P. W. Ball, Polyparameter linear free energy relationships for estimating the equilibrium partition of organic compounds between water and the natural organic matter in soils and sediments, *Environ. Sci. Technol.* 39 (2005) 913–924.
- [30] S. K. Poole, C. F. Poole, Chromatographic models for the sorption of neutral organic compounds by soil from water and air, *J. Chromatogr. A.* 845 (1999) 381–400
- [31] F. Lj. Andrić, J. Đ. Trifković, A. D. Radoičić, S. B. Šegan, Ž. Lj. Tešić, D. M. Milojković-Opsenica, Determination of the soil–water partition coefficients (log KOC) of some mono and poly-substituted phenols by reversed-phase thin-layer chromatography, *Chemosphere*. 81 (2010) 299–305.
- [32] G. Schüürmann, R. Ebert, R. Kühne, Prediction of the sorption of organic compounds into soil organic matter from molecular structure, *Environ. Sci. Technol.* 40 (22) (2006) 7005–7011.

- [33] B. Von Oepen, W. Kördel, W. Klein, G. Schüürmann, Predictive QSPR models for estimating soil sorption coefficients: potential and limitations based on dominating processes, *Sci. Total Environ.* 109/110 (1991) 343–354.
- [34] K. Varmuza, P. Filzmoser, Introduction to multivariate statistical analysis in chemometrics. CRC Press, Taylor & Francis Group, Boca Raton, Florida, USA, (2009).
- [35] K. Heberger, K. Kolar-Hunek, Sum of ranking differences for method discrimination and its validation: comparison of ranks with random numbers, *J. Chemom.* 25 (2011) 151–158.
- [36] K. Heberger, Sum of ranking differences compares methods fairly, *TrAC, Trends Anal. Chem.* 29 (2010) 101–109.
- [37] T. Cserhádi, *Multivariate Methods in Chromatography: A Practical Guide*, Chichester, England, John Wiley & Sons Ltd., 2008, pp. 219–220.
- [38] R. Kaliszan, *Structure and Retention in Chromatography*, Amsterdam, Harwood Academic Publishers, 1997, pp. 56–57.
- [39] U. Normder, Analysis of reversed-phase liquid chromatographic separations using the data-reduction methods PCR and PLS, *Anal. Chim. Acta.* 259 (1992) 105–108.
- [40] C. West, M. A. Khalikova, E. Lesellier, K. Héberger, Sum of ranking differences to rank stationary phases used in packed column supercritical fluid chromatography, *J. Chromatogr. A.* 1409 (28) (2015) 241–250.
- [41] F. Andrić, K. Héberger, Chromatographic and computational assessment of lipophilicity using sum of ranking differences and generalized pair-correlation, *J. Chromatogr. A.* 1380 (6) (2015) 130–138.
- [42] F. Andrić, K. Héberger, Towards better understanding of lipophilicity: Assessment of in silico and chromatographic logP measures for pharmaceutically important compounds by nonparametric rankings, *J. Pharm. Biomed. Anal.* 115 (2015) 183–191.

- [43] D. Bajusz, A. Rácz, K. Héberger, Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations, *J. Cheminf.* 7 (1, 8) (2015) 20.
- [44] A. Rácz, N. Papp, E. Balogh, M. Fodor, K. Héberger, Comparison of antioxidant capacity assays with chemometric methods, *Anal. Methods.* 7 (10) (2015) 4216–4224.
- [45] A. Rácz, D. Bajusz, K. Héberger, Consistency of QSAR models: Correct split of training and test sets, ranking of models and performance parameters, *SAR QSAR Environ. Res.* 26 (7-9) (2015) 683–700.
- [46] H. C. Tülp, K. Goss, R. P. Schwarzenbach, K. Fenner, Experimental determination of LSER parameters for a set of 76 diverse pesticides and pharmaceuticals, *J. Environ. Sci. Technol.* 42 (6) (2008) 2034–2040.
- [47] M. H. Abraham, A. Ibrahim, A. M. Zissimos, Determination of sets of solute descriptors from chromatographic measurements, *J. Chromatogr. A.* 1037 (1-2) (2004) 29–47.
- [48] E. Lazaro, C. Rafols, M. H. Abraham, M. Rose, Chromatographic Estimation of Drug Disposition Properties by Means of Immobilized Artificial Membranes (IAM) and C18 Columns, *J. Med. Chem.* 49 (2006) 4861–4870.
- [49] Y Ishihama, N. Asakawa, Characterization of lipophilicity scales using vectors from solvation energy descriptors. *J. Pharm. Sci.* 88 (1999) 1305–1312.
- [50] C. F. Poole, T. C. Ariyasena, N. Lenca, Estimation of the environmental properties of compounds from chromatographic measurements and the solvation parameter model, *J. Chromatogr. A.* 1317 (2013) 85–104.

Figure captions

Fig. 1 PCA of standardized retention data; Score plot of $PC1$ vs. $PC2$ (a) showing characteristic grouping of compounds and loading plot (b) illustrating similarities and differences among chromatographic systems

Fig. 2 SRD comparison with random numbers (CRRN) of chromatographic and in silico approaches for determination of $\log K_{OC}$ (a); x axis and left side y axis represent scaled SRD values (%), right side y axis represent relative frequencies of random numbers (%); Additional ranking and grouping based on sevenfold “jack knife like” SRD cross-validation procedure (b); Dashed lines denote statistically significant difference among variables (approaches) at $p = 0.05$.

Fig. 3 PCA of LSER regression coefficients: a , b , s , e , and v ; Score plot (a) and loading diagram (b) illustrating characteristic pattern of chromatographic systems vs. Nguyen’s and Poole’s general LSER $\log K_{OC}$ model and influence of LSER coefficients on such disposition. Dashed circle line represent 95% confidence limit.

Tables

Table 1 List of standard compounds and unknowns accompanied with the list of $\log K_{OC}$ values directly measured by the soil-sorption experiments and recommended by U.S. EPA

No	Compound	$\log K_{OC}$	Ref.	No	Compound
<i>Standard compounds</i>				<i>Unknowns</i>	
1	Phenol	1.43	[32]	30	3-Nitrophenol
2	4-Nitrophenol	2.37	[32]	31	2-Naphthol
3	Benzyl Alcohol	1.10	[28]	32	4-Hydroxybenzaldehyde
4	1-Naphthylamine	3.51	[32]	33	2-Aminophenol
5	1-Naphthol	3.10	[32]	34	4- <i>t</i> -Buthylphenol
6	2,4-Dichlorophenol	2.81	[32]	35	2,6-Dimethylphenol
7	Anthracene	4.31	[32]	36	4-Methoxyphenol
8	Acetophenone	1.80	[28]	37	Methyl- <i>p</i> -hydroxybenzoate
9	2,4,6-Trichlorophenol	3.03	[32]	38	2-Nitrobenzaldehyde
10	Ethyl- <i>p</i> -hydroxybenzoate	2.21	[32]	39	3-Nitrobenzaldehyde
11	<i>p</i> -Anisidine	1.93	[29]	40	Phthalimide
12	1,2,3-Benzotriazole	1.69	[32]	41	Oxazepam
13	Diphenylamine	2.78	[32]	42	Lorazepam
14	2,2'-Dipiridyne	1.60	[32]	43	Clonazepam
15	4-Bromoaniline	1.96	[32]	44	Bromazepam
16	Benzophenone	2.63	[32]	45	Diazepam
17	4-Aminobenzoic acid	1.70	[33]	46	Nitrazepam
18	Pyrene	4.90	[32]	47	Chlordiazepoxide
19	Benzo[<i>a</i>]pyrene	5.95	[28]	48	Clobazam
20	Fluorene	3.70	[32]	49	Medazepam
21	Acenaphthene	3.59	[32]	50	Chrysene
22	Naphthalene	2.96	[32]		
23	Phenanthrene	4.35	[32]		
24	Dibenz[<i>a,h</i>]anthracene	6.22	[32]		
25	Benz[<i>a</i>]anthracene	5.30	[32]		
26	Simazine	2.10	[32]		
27	Propazine	2.40	[32]		
28	Ametryn	2.59	[32]		
29	Prometryn	2.85	[32]		

Table 2 Selected chromatographic models for prediction of $\log K_{OC}$ accompanied by statistical parameters

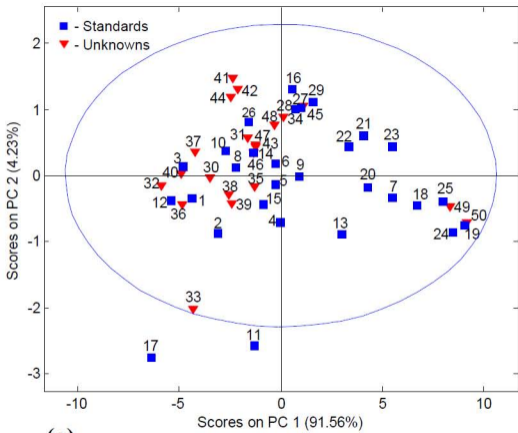
No.	Chrom. method	Model type	Statistical parameters
1	TLC	PCR	$n(PC) = 2$; $RMSE_{Cal} = 0.433$, $RMSE_{CV} = 0.517$, $RMSE_{PRED} = 0.494$, $\max RMSE_{PRED} = 0.568$; $R^2_{Cal} = 0.8921$, $R^2_{CV} = 0.8460$, $R^2_{PRED} = 0.8596$; Percent of variance captured by each principal component in the X and Y domain, respectively: $PC1$: 91.88%, and 86.40% $PC2$: 4.49%, and 2.70%
2		PLS	$n(LV) = 2$; $RMSE_{Cal} = 0.401$, $RMSE_{CV} = 0.496$, $RMSE_{PRED} = 0.492$, $\max RMSE_{PRED} = 0.608$; $R^2_{Cal} = 0.9124$, $R^2_{CV} = 0.8585$, $R^2_{PRED} = 0.8605$; Percent of variance captured by each latent variable in the X and Y domain, respectively: $PLS1$: 91.92%, and 86.73% $PLS2$: 4.33%, and 3.94%
3		OLS*	RP18-silica, MeOH 70% v/v $\log K_{OC} = (2.00 \pm 0.14) + (2.11 \pm 0.19) R_M$, $RMSE_{Cal} = 0.500$, $RMSE_{PRED} = 0.527$, $R^2_{Cal} = 0.8557$, $F = 160.17$, $p < 1 \cdot 10^{-4}$, $R^2_{PRED} = 0.8397$, $n = 29$
4			RP18-silica, MeOH 80% v/v $\log K_{OC} = (2.57 \pm 0.12) + (2.60 \pm 0.25) R_M$, $RMSE_{Cal} = 0.516$, $RMSE_{PRED} = 0.557$, $R^2_{Cal} = 0.8465$, $F = 148.90$, $p < 1 \cdot 10^{-4}$, $R^2_{PRED} = 0.8210$, $n = 29$
5			CN-silica, MeOH 40% v/v $\log K_{OC} = (1.55 \pm 0.16) + (2.21 \pm 0.19) R_M$, $RMSE_{Cal} = 0.471$, $RMSE_{PRED} = 0.506$, $R^2_{Cal} = 0.8722$, $F = 184.31$, $p < 1 \cdot 10^{-4}$, $R^2_{PRED} = 0.8522$, $n = 29$
6			CN-silica, MeOH 50% v/v $\log K_{OC} = (2.27 \pm 0.10) + (2.67 \pm 0.19) R_M$, $RMSE_{Cal} = 0.403$, $RMSE_{PRED} = 0.441$, $R^2_{Cal} = 0.9062$, $F = 260.93$, $p < 1 \cdot 10^{-4}$, $R^2_{PRED} = 0.8878$, $n = 29$
7			CN-silica, MeOH 60% v/v $\log K_{OC} = (2.94 \pm 0.10) + (3.36 \pm 0.26) R_M$, $RMSE_{Cal} = 0.439$, $RMSE_{PRED} = 0.474$, $R^2_{Cal} = 0.8892$, $F = 216.65$, $p < 1 \cdot 10^{-4}$, $R^2_{PRED} = 0.8716$, $n = 29$
8			CN-silica, MeOH 70% v/v $\log K_{OC} = (3.69 \pm 0.13) + (4.13 \pm 0.40) R_M$, $RMSE_{Cal} = 0.531$, $RMSE_{PRED} = 0.556$, $R^2_{Cal} = 0.8373$, $F = 138.92$, $p < 1 \cdot 10^{-4}$, $R^2_{PRED} = 0.8220$, $n = 29$
9			CN-silica, ACN 50% v/v $\log K_{OC} = (2.72 \pm 0.12) + (4.43 \pm 0.44) R_M$, $RMSE_{Cal} = 0.537$, $RMSE_{PRED} = 0.568$, $R^2_{Cal} = 0.8340$, $F = 135.68$, $p < 1 \cdot 10^{-4}$, $R^2_{PRED} = 0.8140$, $n = 29$
10	HPLC	OLS*	$\log k = -0.45 (\pm 0.10) + 0.274 (\pm 0.031) \log K_{OC}$

$$R^2_{\text{Cal}} = 0.865, \text{ S.D.} = 0.219, F = 80.18, P = 1.4 \cdot 10^{-9}, n = 29$$

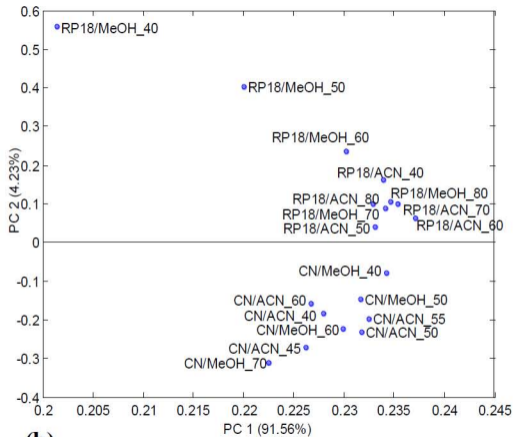
*Univariate ordinary least squares (OLS) regression

Table 3 Distances between LSER vectors (normalized to the unit length) of individual TLC systems and LSER $\log K_{OC}$ models of Pool and Nguyen; D' represents the absolute difference among two vector lengths.

No.	Chrom. system		<i>Distances from the Poole's $\log K_{OC}$ LSER model</i>			<i>Distances from the Nguyen's $\log K_{OC}$ LSER model</i>		
	St. phase	Mob. phase	d	D'	$\cos\theta$	d	D'	$\cos\theta$
1	RPI8 - silica	MeOH 40 % v/v	0.406	1.61	0.797	0.283	1.67	0.858
2		MeOH 50 % v/v	0.320	1.48	0.840	0.203	1.54	0.899
3		MeOH 60 % v/v	0.218	1.58	0.891	0.140	1.64	0.930
4		MeOH 70 % v/v	0.178	1.72	0.911	0.105	1.78	0.948
5		MeOH 80 % v/v	0.175	2.01	0.913	0.117	2.07	0.941
6		ACN 40 % v/v	0.254	1.64	0.873	0.217	1.70	0.892
7		ACN 50 % v/v	0.396	1.79	0.802	0.329	1.85	0.836
8		ACN 60 % v/v	0.374	1.82	0.813	0.325	1.88	0.837
9		ACN 70 % v/v	0.398	1.89	0.801	0.334	1.94	0.833
10		ACN 80 % v/v	0.459	1.94	0.770	0.376	2.00	0.812
11	CN -silica	MeOH 40 % v/v	0.153	1.81	0.924	0.125	1.87	0.938
12		MeOH 50 % v/v	0.142	2.12	0.929	0.146	2.18	0.927
13		MeOH 60 % v/v	0.220	2.39	0.890	0.203	2.44	0.898
14		MeOH 70 % v/v	0.419	2.63	0.791	0.408	2.69	0.796
15		ACN 40 % v/v	0.180	2.25	0.910	0.144	2.31	0.928
16		ACN 45 % v/v	0.221	2.28	0.890	0.177	2.34	0.911
17		ACN 50 % v/v	0.355	2.57	0.822	0.294	2.63	0.853
18		ACN 55 % v/v	0.373	2.59	0.813	0.290	2.65	0.855
19		ACN 60 % v/v	0.331	2.66	0.835	0.304	2.71	0.848



(a)



(b)

