Classification of biphasic solvent systems according to Abraham descriptors for countercurrent chromatography

Léa Marlot, Magali Batteau, Karine Faure

Classification of biphasic solvent systems according to Abraham descriptors for countercurrent chromatography

Léa Marlot[1], Magali Batteau[1], Karine Faure[1]

[1] Université de Lyon, CNRS, Université Claude Bernard Lyon 1, Institut des Sciences Analytiques, UMR 5280, 5 rue de la Doua, F-69100 VILLEURBANNE, France

Abstract

The method development of liquid-liquid chromatography, either countercurrent chromatography or centrifugal partition chromatography, is slowed down by the selection of the biphasic solvent system that constitutes its column. This paper introduces a classification of 19 solvent systems, including the most popular systems based on heptane/ethyl acetate/methanol/water, some non-aqueous systems and some greener systems. This classification is based on Abraham descriptors determined through the partition coefficients of 43 probes. Among 21 determined models, nine of them allow an accurate prediction of partition coefficients from solute descriptors and another ten provide a description of the chromatographic interactions at the 5% significance level. A graphical tool (spider diagram) is built for the comparison of the chromatographic columns previously characterized with the solvation parameter model. The position of a solvent system in this spider diagram relates to the interactions at stake, thus the selection of columns offering similar or orthogonal interactions is facilitated, with no previous knowledge of the solute required. This semi-empirical strategy cannot fully predict the retention behavior but can judiciously orientate the user towards a limited number of solvent systems to be experimentally tested.

Keywords: Countercurrent chromatography, centrifugal partition chromatography, Spider diagram; Abraham descriptors; Classification methods;

* Corresponding author : Karine Faure, Institut des Sciences Analytiques, 5 rue de la Doua, F-69100 VILLEURBANNE, France ; phone : +33-437423686 ; e–mail: karine.faure@isa-lyon.fr

1. Introduction

Support-free liquid-liquid chromatography (LLC), also known as countercurrent chromatography (CCC) or centrifugal partition chromatography (CPC), has emerged as an important instrumental approach for the purification and isolation of active compounds from relatively complex samples, mainly in the natural product industry. The stationary and mobile phases are the two-phases of a biphasic solvent system. This characteristic offers a wide range of operating modes as recently reviewed [1]. It also provides a versatility in the nature of the columns that can be engaged in the process, leading usually to much larger selectivity than observed in solid-phase chromatography. The number of biphasic systems (called solvent systems) is tremendous, being generated from the mixture of up to four different solvents in various proportions and the selection process is hence labor-intensive and highly time-consuming in the overall method development. The most efficient strategy is to select a limited number of solvent systems and to conduct partition experiments using the actual sample to be purified. It is often recommended that the target compound exhibits a partition coefficient in the range $0.25 < K < 4$ in elution mode, but alternative modes such as elution-extrusion, dual-mode or co-current mode have now spread the range of partition coefficient spectrum that can be separated. Hence, the driving force for solvent selection should now focus on selectivity.

Once the best solvent system is identified, adjustments of its composition have to be made to obtain the maximal selectivity between compounds of interest and the matrix, using factorial designs for example [2] or scanning through the biphasic composition diagram. Hence, it is clear that the primary investigation of the most suitable solvent system candidates has to be conducted with minimal effort yet guided by chemical considerations.

Solvent system screening can be facilitated by the use of databases, solvent system families and thermodynamic models, as summarized in 2015 by Liu et al. [3]. Skalicka-Wozniak and Garrard built a comprehensive database containing the solvent systems used from 1984 to 2014 in natural product purification by LLC [4]. Since natural products represent more than 80 % of the CCC applications, this database can be considered as a good overlook of the columns used. After compiling 2322 isocratic solvent systems, they found out that these were constituted of 29 different solvents. This number is rather elevated compared to the few number of solvent used in preparative and reflect the large degree of freedom still available in LLC development. These databases can be of help to orientate the solvent screening towards certain solvent families, the HEMWat (heptane/ethyl acetate/methanol/water) families being the most used in natural products of medium polarity [5].

Another approach based on the “best solvent approach” consists in finding a solvent that easily dissolves the compound of interest and in building around a biphasic system with two partially miscible solvents [6]. This approach is, unfortunately, highly related to the scientist’experience, and limited to binary or ternary mixtures of solvents.

Some theoretical strategies tend to simulate the partition coefficient of a given solute. Its solubility in each LLC phase is calculated relating to its chemical structure combining solvent descriptors with solutes characteristics. While considerably decreasing the experimental work, these theoretical strategies such as COSMO-RS [7, 8] or UNIFAC [9] require the previous knowledge of the chemical structure of the compound of interest or at least its descriptors, but also of the impurities that may be present, as well as an expertise on computational tools.

The modelling approach based on quantitative structure activity relationships (QSAR) has been extensively detailed on HEMWat solvent systems [10] with models based on 196

descriptors for each test compound. The authors were hence able to predict log K within 0.5 log unit. One of the main advantages of a QSAR model is that it can be run simply using Excel.

The aim of the present work is to supply a visual mean to compare a large number of solvent systems and the interactions they involve. For this purpose, a classification of the most common solvents systems is performed using Abraham descriptors. While being less precise than a complete QSAR model, with only 5 descriptors per solvent system, the model could be used to estimate the partition coefficient of a neutral solute if its Abraham descriptors are known. More importantly this classification aims at illustrating similar (correlated) and complementary (orthogonal) columns by their respective position. It provides a faster decision support in CCC/CPC development method for unknown compounds. The classification presents here 21 biphasic solvent systems (hence providing 42 columns depending on the role of each phase), amongst which 3 totally organic solvent systems.

2. Experimental

2.1 Chemicals

The solvents used were analytical grade provided by Acros organic (1-butanol, methyl tert-butyl ether) or Sigma-Aldrich (methanol, acetonitrile, ethanol, limonene, heptane, ethyl acetate). The initial compositions (v/v) of each studied solvent system are listed in Table 1. The composition of the upper and lower phases of each solvent system has been described in several papers [8; 11-14]

[Table 1]

A set of 27 compounds was extracted from a previous publication by West et al. [15] who use them to classify SFC columns. They represent homologous series that easily describe some LSER descriptors. Fourteen compounds, from the GUESS list [16], were added to the set since they are very often encountered in LLC column testing, covering a wide range of polarity, and supplying a better description of the terms A and B. Their LSER descriptors *E*, *S*, *A*, *B* and *V* were extracted from ACD web database. All solutes were purchased from Sigma-Aldrich. These 43 solutes were pooled into 7 groups, mainly by chemical families, to ensure their chromatographic separation (Table 2). A second set comprising 5 solutes was gathered to validate the models (Table 2: validation set) and their partition coefficients were measured using partition experiments (shake-flasks) and CCC and CPC separations.

[Table 2]

2.2 Determination of partition coefficient: shake-flask procedure

Each solvent system (Table 1) was prepared using thermostated solvents and allowed to settle for three hours in a thermostated bath at 24 °C (± 1 °C). A shake-flask constituted of 2 mL of each phase was arranged for each group of solutes. This means that each solvent system to be characterized requires 7 shake-flasks. Then 2 mg of each solute of the group of concern was added to its shake-flask. It is noteworthy that these partition tests have to be performed at high dilution to avoid any saturation of a phase. After dissolution through ultrasonic bath, the shake flasks were allowed to sit for 30 min in a thermostated bath at 24 °C (±1 °C) to ensure that thermodynamic equilibrium was reached.

2.3 Analytical conditions

Aliquots of the upper and lower phases of the shake-flasks were diluted by a factor 10 in methanol. The UHPLC analysis was performed using a Waters Acquity instrument, equipped with a 20 µL injection loop and for which the dwell volume is estimated at 0.11 mL (loop excluded). The column Acquity CSH Phenylhexyl (100 x 2.1 mm i.d., 1.7 µm) was employed at 30 °C. The mobile phase was composed of (A) water + 0.1 % formic acid and (B) acetonitrile + 0.1 % formic acid. The flow rate was 0.4 mL/min and UV detection was set at 254 nm. The injection volume was 1 µL. For each group a UHPLC gradient method was optimized using OSIRIS software (Datalys, Grenoble). Resulting methods are providing in Table S1. Peak areas were monitored thanks to Empower software and partition coefficients were established as the ratio of the upper phase peak area on the lower phase peak area, considering that the sample solvent does not influence the detector response.

2.4 Data analysis

For the determination of the coefficients *e*, *s*, *a*, *b* and *v* of each solvent system, multilinear regression was performed using Excel, as well as statistical analysis and graphic representations. The number of solutes used for the multilinear regression of each solvent system is indicated in Table 3.

2.5 Centrifugal partition chromatography and countercurrent chromatography

The LLC separation was conducted on the validation set solutes with three different columns: #SS2 (heptane/ethyl acetate/methanol/water 1/1/1/1), #SS7 (octanol/water) and #SS15 (heptane/methanol/water 50/33.5/16.5), the first two in descending mode; i.e. with the upper phase used as stationary phase, the latter in ascending mode, i.e. with the lower phase used as stationary phase.

142 The CPC experiments were set with a Spotprep II pumping system from Gilson (Saint-Avé,
143 France) connected to a FCPC-A frame equipped with a 33.1 mL rotor from Kromaton
144 Rousselet-Robatel (Annonay, France) thermostated at 25 °C (± 3 °C) and detection was set up
145 at 254 or 280 nm. Operating conditions were in elution mode, at 2000 rpm and 10 mL/min,
146 with an injection volume of 0.5 mL. Because of the large expected K values for #SS7, an
147 artificial stationary phase ratio of 10 % of the volume column was induced.

148 The CCC experiments were set with a Waters 600E pumping system (Waters, Milford,
149 Massachusetts, USA) on a CCC apparatus (Spectrum, Dynamic Extractions, Slough, UK) with a
150 65.5 mL column (polytetrafluoroethylene bore tubing of 3.2 mm, β range 0.52-0.86)
151 thermostated at 25 °C (± 2 °C) and a SPD20A detector (Shimadzu, Noisel, France) set up at 210
152 nm. The rotational speed was 1000 rpm. A classical elution at 5 mL/min of mobile phase for
153 the first 6 minutes was followed by a co-current mode, with the addition of a flow of
154 “stationary” phase of 1 mL/min, in order to elute compounds with very large K values.

155

156 3. Results and discussion

157 The linear solvation energy relationships (LSERs) framework relates the properties of a solute
158 and a solvent to the strength of the interactions that generate retention mechanism in such
159 an environment [17,18]. In a partition process, molecular interactions take place in each phase
160 and the overall interaction strength provides the partition coefficient K. Hence, the LSER
161 approach for neutral solutes can be transposed as equation 1.

162 $\log K = c + eE + sS + aA + bB + vV$ Eq. (1)

163 *K* is the partitioning coefficient of a solute in the biphasic system. The capital letters E, S, A, B
164 and V are the solute descriptors highlighting its ability to participate in a given interaction. The
165 lowercase letters c, e, s, a, b and v are the coefficients that describe, in a specific biphasic
166 system, the difference in the strength of interactions with the solutes in the two immiscible
167 phases. While c is a system constant, the descriptor E represents the excess molar refractivity,
168 i.e. the electron lone pair interaction, the descriptor S measures the dipole-type interaction
169 solute, A and B are the hydrogen bond acidity and basicity respectively, and V is the solute's
170 McGowan characteristic volume.

171 It is very important to consider that the system parameters (noted as lowercase letters)
172 reflects the differences in the properties of the two phases and not the properties of a specific
173 phase. Hence, a coefficient with a zero value for a biphasic solvent system that does not mean
174 that this type of interaction is absent, but that the same strength of interaction is actually
175 provided in each phase, meaning that this specific interaction does not contribute to the
176 overall free energy of transfer and hence does not influence the partition coefficient.

177 3.1 Descriptors

178 To properly describe the constants of a given system, a minimum number of 20 solutes is
179 recommended, providing a wide range of interactions, preferentially with at least 4 solutes
180 per studied interaction [17]. Figure S1a in supplementary material presents the repartition of
181 the 43 solutes according to their polarity (log P) and molecular weight. The solutes are mostly
182 below 300 g/mol, and polarity log P exhibits a wide range of 0–6 with few polar compounds
183 (chlorogenic acid log P = -0.36). This set of solutes is representative of most applications that
184 are dealt with in centrifugal partition chromatography. The selected compounds provide a

185 good general cover of LSER descriptor space (Figure S1b) between 0-3.9 for E, 0-5 for S and V,
186 0-2.0 for A and 0-3.0 for B (41 values for each descriptor, except A that has 19 values). The
187 adjunction of GUESS solutes [16] provides A and B values above 1 which was not the case if
188 only conventional LC/SFC test solutes were used. Absence of cross –correlation between the
189 descriptors was proven through a covariance matrix with values far below 1 (Table S2).

190 3.2 Partition coefficients

191 The partition coefficient K of each solute was determined in every studied biphasic systems
192 using the shake-flask procedure (see experimental section for details) and calculated
193 conventionally according to equation 2.

194 $K = [A]_{up}/[A]_{low}$ Eq. (2)

195 Where $[A]_{up}$ is the solute concentration in the upper phase and $[A]_{low}$ is the concentration in
196 the lower phase. This K value corresponds to the retention factor in CPC when used in
197 descending mode, i.e. the stationary phase is the upper phase. When using the ascending
198 mode, i.e. the lower phase as stationary phase, the retention factor is the inverse of K. Hence
199 it would be possible to measure K values of solutes in a solvent system directly during a CPC
200 procedure [19].

201 Repeatability experiments (3 shake-flasks and 3 HPLC determinations per shake-flask) were
202 carried out on group 7 (five solutes) in solvent system #2 (heptane/methanol/ethyl
203 acetate/water 1/1/1/1) and it was found that relative standard deviation was below 2 %.
204 While errors can occur when the concentration in a phase is very low, the partition coefficient
205 measured in shake-flask is considered as reliable for log K between -3 and +3 [20]. In the set
206 we selected, only 1 % of the log K values are over +3 and 2 % below -3. When solutes did not

properly partition i.e. when HPLC peaks were too small to be quantified or a solubility issue occurred, the solute descriptors were discarded from the model.

3.3 Multilinear regression

For each solvent system a multilinear regression was performed and the 21 solvent system descriptors *e*, *s*, *a*, *b* and *v* are presented in Table 3, along with the number of solutes *n* that were considered in the regression. Indeed, some solutes were discarded due to solubility issues.

[Table 3]

n number of solutes considered in the multilinear regression; u vector length; p-value probability value

Over the 21 studied solvent systems, 9 models exhibit R^2adjusted over 0.8, meaning that these models can be used to accurately predict the partition coefficient of any solute for which the LSER coefficients are known. For 10 models, the endogeneous variance is not fully explained by the model (R^2adjusted below 80%), but the models are globally significant at the 5% level (p-value below 0.05). For these models, there is a significant linear relationship between the response and the prediction, with a 5% risk of false positive. These models cannot be used to accurately predict the partition coefficient of a known compound, but they can illustrate the global chromatographic interactions. In practice, these models can be used to compare two solvent systems in terms of selectivity and increase the chance to select orthogonal solvent systems without any knowledge of the solutes to be separated.

Two models were found not significant at the 5% level. Systems #19 and #21 exhibit p-values around 8.6 % and 6.5 % which means that there is over 5 % risk that the correlation is not

229 significant. The residuals for these models were found to vary with the retention value, a
230 statistic phenomenon called heteroscedasticity, which indicates that the LSER model is not
231 adequate, probably due to the lack of a variable. These two solvent systems are hence
232 discarded from the rest of the study.

233

234 To illustrate the correlation between the log K values calculated through the model and the
235 experimental log K values and the residuals, Figure 1 presents the results obtained with the
236 most common solvent system, heptane/ethyl acetate/methanol/water 1/1/1/1, here named
237 system #2. The analysis of residuals [21] shows that they have a normal distribution (the mean
238 value -1.10^{-16} being close to the median value -1.10^{-2}). Two outliers (quercetine and
239 propriophenone) were identified above the residuals distribution range (-0.426; +0.394) and
240 two others (umbelliferone and ferulic acid) were found below this range. These solutes were
241 not properly modeled with the LSER model due to either experimental error or inadequate
242 model.

243 [Figure 1]

244 The root mean square error (RMSE) is calculated over the 39 experimental measurements
245 using equation 3. Its value that should be as low as possible was found to be 9.10^{-16}, showing
246 a good accuracy of the model. The open circles in Figure 1 represents the results obtained for
247 the validation set that was not included to build the model. For these solutes, it was found a
248 RMSE of 0.06, which is lower than the prediction accuracy reported for 580 compounds of the
249 heptane-water partition coefficients (RMSE = 1.45) [22]

250 $$RMSE = \sqrt{\frac{\sum(\log K_c - \log K_{exp})^2}{n}}$$ Eq. (3)

where K_c is the calculated value and K_{exp} the experimental value and n is the number of observations.

3.4 Representation of solvent systems with spider diagram

The system constants provide information on the nature and extent of the interactions between the solute and the solvent. However, comparing two systems requires the comparison of the five coefficients of each system. While this is still achievable for two systems, it becomes almost impossible for the comparison of several systems because of the number of data to compare. Many representations exist [23] but the representation on a 5-axis diagram (called spider diagram) [24] is the easiest way to visualize similarity and differences simply according to the distance between dots. These diagrams are simple to represent and very illustrative to interpret the results. The mathematical treatment to build spider diagrams was described elsewhere [15]. Each column (the combination of stationary and mobile phases) is represented as a vector, where the coordinates of the vector are the values provided by the system constants. The overall strength of interactions is estimated through the length *u* of the solvatation vector (equation 4) and represented as the size of the dot [25]. This vector length is a useful tool to compare the strength of the interaction capabilities for each solvent system.

$$u = \sqrt{e^2 + s^2 + a^2 + b^2 + v^2} \qquad \text{Eq. (4)}$$

It is important to note that in the case of CPC, a single biphasic system can generate two opposite chromatographic columns, depending on the phase selected as stationary phase. Hence, in descending mode, the stationary phase is the upper phase of the biphasic system. Its system constants are indicated in Table 3. In ascending mode, the lower phase is selected

as the stationary phase, and hence the ascending system constants exhibit opposite values to the corresponding descending column constants. In Figure 2 are represented the descending solvent systems numbered as in Table 1, and their ascending counterparts (same colour but no numbering).

[Figure 2]

Two close dots represent solvent systems that exhibit similar interactions for a given solute, a distant point provides different interactions. Hence we can expect that in CPC, the selection of two close solvent systems may provide similar elution order, i.e. that the retention mechanisms are correlated, while two distant solvent systems should offer different types of interactions and hence orthogonal separations with different elution order. This was illustrated by comparing solvent systems #2, #7 and #15asc (solvent system #15 used in ascending mode, meaning the aqueous phase is used as stationary phase). A closer look on these solvent systems coefficients on a radar plot (Figure 3a) clearly indicates the similarities and differences in the strength of the interactions that lead the separations. When running these columns for the CCC separation of validation set solutes (Figure 3b), the shift in retention order and hence the non-correlation between solvents system #15asc and the two other is obvious, while some correlation exist between solvent systems #2 and #7.

[Figure 3]

3.5 Validation set

294 The LSER models using Abraham descriptors have limited predictive ability. However, it was
295 interesting to verify, with a set of solutes that are not homologous to the model solutes, how
296 predictable their partition coefficients or their CPC/CCC retention are. To do this, five solutes
297 were tested (group 8; Abraham descriptors in Figure S2), with their partition coefficient
298 measured via shake-flask in the 19 solvent systems, and their retention factors measured
299 using centrifugal partition chromatography (CPC) and countercurrent chromatography (CCC)
300 on system solvents #2, #7 and #15asc. Figure 4a reports the results from the 5 x 19
301 experimental partition coefficients plotted versus their predicted values. The large majority of
302 them fall close to the first bisector, with only 9 % of the values outside the ±0.5 log unit. The
303 RMSE is calculated at 0.448, a value of less than 0.5 being considered as acceptable [10, 22].
304 The validation set was injected in CPC and CCC instruments with 3 solvent systems. Catechol
305 was not injected in #15asc as its retention factor was predicted at 353, which was considered
306 as not feasible. Because some predicted values of retention factors reach nearly 20, the LLC
307 methods were adapted to reduce the run duration, which is a great benefit of the liquid nature
308 of the stationary phase. The CPC instrument was operated with a reduced amount of
309 stationary phase. A 10 % column volume of stationary phase was introduced before
310 equilibration with the mobile phase and a conventional elution mode was performed. The
311 partition coefficient K was deduced from the observed retention volume V_R (Equation 5).

312 $$V_R = Vc - (K-1)*V_S \qquad \text{Eq. (5)}$$

313 with Vc the column volume and V_S the stationary phase volume.

314 The CCC separation was run in a co-current mode, which consists in slowly pushing the
315 stationary phase out while the elution goes on. In order to analyze low-retained and highly

316 retained compounds in the same run, the co-current mode was started after a few minutes
317 delay. Equation 6 is derived from the co-current equation [26] to consider this delay.

318 $$K = \frac{t_R x F_M - V_M}{V_S - (t_R - t_{delay}) x F_S}$$ Eq. (6)

319 With t_R the retention time, F_M the mobile phase flowrate, V_M the mobile phase volume, t_{delay}
320 the starting time for co-current mode and F_s the stationary phase flowrate during this co-
321 current mode.

322 The RMSE was found to be 0.36 for CPC measurements and 0.47 for CCC measurement (n=12).
323 Once again, the experimental values are close to the calculated values as seen in Figure 4b
324 with a deviation of ±0.5 log units, with the exception of catechol, for which the experimental
325 value was always far below the calculated value. Despite Marsden-Jones's statement [10], a
326 0.5 log unit deviation does not seem acceptable to predict the most suitable solvent system
327 in CCC or CPC, but it may be sufficient to select few of them on which further studies can be
328 conducted to confirm solvent system selectivity. The concordance between the partition
329 coefficients observed in LLC techniques and the one measured using shake-flask procedure is
330 notoriously limited [27,28]. While CCC provides usually closer values than CPC to the expected
331 shake-flasks values, we suspect the temperature control in LLC instruments being a general
332 source of deviation from the prediction.

333

334 4. Conclusion

335 Based upon the Abraham descriptors of a set of probes, we have presented here the LSER
336 classification of biphasic solvent systems that can act as CCC or CPC columns for neutral
337 species. The Abraham model was found accurate for 9 out of 21 biphasic solvent systems

and can provide a significant correlation for 10 other solvent systems. A validation set of five solutes was used to confirm the strong correlation between the model and the experimental values. While the partition experiments in a static environment follow the model, the observed partition coefficients in CPC or CCC experiments show some deviation. Nonetheless, the visualization as a spider diagram contributes to the selection of similar or orthogonal columns. While the exploration of various compositions within a chosen family must still be performed, this semi-empirical strategy reduces the workload by directing the selection towards orthogonal systems, hence aiming at a faster method development or an easier investigation of 2D configurations, including LLC techniques alone or in combination with LC techniques.

Acknowledgements.

The authors wish to thank Pavel Smak for his help in preliminary studies.

Funding.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Morley, R. and M. Minceva, *Operating mode and parameter selection in liquid–liquid chromatography.* J Chromatogr A, 2019: p. 460479.
2. Dubant, S., et al., *Practical solvent system selection for counter-current separation of pharmaceutical compounds.* J Chromatogr A, 2008. **1207**(1): p. 190-192.
3. Liu, Y., et al., *Solvent System Selection Strategies in Countercurrent Separation.* Planta Medica, 2015. **81**(17): p. 1582-1591.
4. Skalicka-Wozniak, K. and I. Garrard, *A comprehensive classification of solvent systems used for natural product purifications in countercurrent and centrifugal partition chromatography.* Nat. Prod. Rep., 2015. **32**(11): p. 1556-1561.
5. Friesen, J.B., et al., *Countercurrent Separation of Natural Products: An Update.* J. Nat. Prod., 2015. **78**(7): p. 1765-1796.
6. Foucault, A.P. and L. Chevolot, *Counter-current chromatography: instrumentation, solvent selection and some recent applications to natural product purification.* J Chromatogr A, 1998. **808**(1–2): p. 3-22.
7. Frey, A., E. Hopmann, and M. Minceva, *Selection of Biphasic Liquid Systems in Liquid-Liquid Chromatography Using Predictive Thermodynamic Models.* Chem. Eng. Technol., 2014. **37**(10): p. 1663-1674.
8. Hopmann, E., W. Arlt, and M. Minceva, *Solvent system selection in counter-current chromatography using conductor-like screening model for real solvents.* J Chromatogr A, 2011. **1218**(2): p. 242-250.
9. Ren, D.B., et al., *Systematic and practical solvent system selection strategy based on the nonrandom two-liquid segment activity coefficient model for real-life counter-current chromatography separation.* J Chromatogr A, 2015. **1393**: p. 47-56.
10. Marsden-Jones, S., et al., *Using quantitative structure activity relationship models to predict an appropriate solvent system from a common solvent system family for countercurrent chromatography separation.* J Chromatogr A, 2015. **1398**: p. 66-72.
11. Foucault, A.P., Centrifugal Partition Chromatography. Chromatography Science series. Vol. 68. 1994, Ed. Marcel Dekker: New York.
12. Faure, K., et al., *Limonene in Arizona liquid systems used in countercurrent chromatography. I Physicochemical properties*. Analytical and Bioanalytical Chemistry, 2014. **406**(24): p. 5909-5917.

13. Garrard, I.J., L. Janaway, and D. Fisher, *Minimising Solvent Usage in High Speed, High Loading, and High Resolution Isocratic Dynamic Extraction*. J Liq Chromatogr Rel Technol, 2007. **30**(2): p. 151-163.

14. Berthod, A., M. Hassoun, and M.J. Ruiz-Angel, *Alkane effect in the Arizona liquid systems used in countercurrent chromatography*. Anal Bioanal Chem, 2005. **383**(2): p. 327-40.

15. West, C. and E. Lesellier, *A unified classification of stationary phases for packed column supercritical fluid chromatography.* J Chromatogr A, 2008. **1191**(1-2): p. 21-39.

16. Friesen, J.B. and G.F. Pauli, *GUESS - A generally useful estimate of solvent systems for CCC.* J. Liq. Chromatogr. Related Technol., 2005. **28**(17): p. 2777-2806.

17. Vitha, M. and P.W. Carr, *The chemical interpretation and practice of linear solvation energy relationships in chromatography.* J Chromatogr A, 2006. **1126**(1): p. 143-194.

18. Abraham, M.H., A. Ibrahim, and A.M. Zissimos, *Determination of sets of solute descriptors from chromatographic measurements.* J Chromatogr A, 2004. **1037**(1): p. 29-47.

19. Faure, K., et al., *Solvent Selection in Countercurrent Chromatography Using Small-Volume Hydrostatic Columns.* LCGC North America, 2013. **31**(2): p. 132-143.

20. Poole, C.F., et al., *Determination of solute descriptors by chromatographic methods.* Anal. Chim. Acta, 2009. **652**(1): p. 32-53.

21. Taraji, M., et al., *Error measures in quantitative structure-retention relationships studies.* J Chromatogr A, 2017. **1524**: p. 298-302.

22. Wittekindt, C. and A. Klamt, *COSMO-RS as a Predictive Tool for Lipophilicity*. QSAR & Combinatorial Science, 2009. **28**(8): p. 874-877.

23. West, C. and E. Lesellier, *Characterisation of stationary phases in subcritical fluid chromatography by the solvation parameter model: II. Comparison tools.* J Chromatogr A, 2006. **1110**(1): p. 191-199.

24. Lesellier, E., *Sigmapider diagram: a universal and versatile approach for system comparison and classification: application to solvent properties.* J Chromatogr A, 2015. **1389**: p. 49-64.

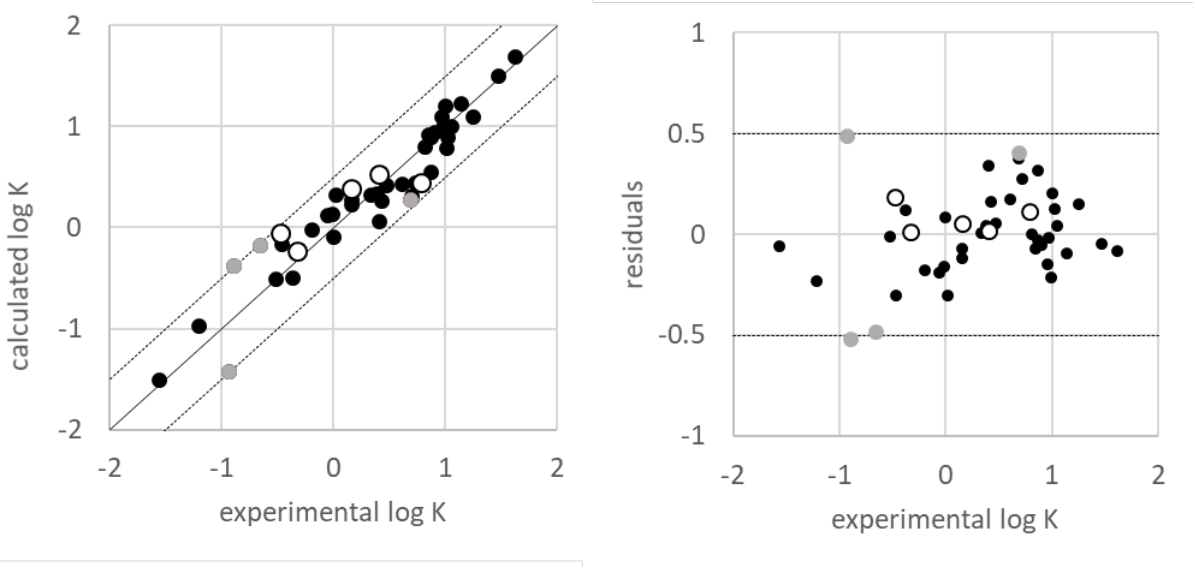
25. Khater, S., C. West, and E. Lesellier, *Characterization of five chemistries and three particle sizes of stationary phases used in supercritical fluid chromatography.* J Chromatogr A, 2013. **1319**: p. 148-159.

26. Berthod, A. and M. Hassoun, *Using the liquid nature of the stationary phase in countercurrent chromatography - IV. The cocurrent CCC method.* J Chromatogr A, 2006. **1116**(1-2): p. 143-148.

27. Ignatova, S., et al., *A new non-synchronous preparative counter-current centrifuge—the next generation of dynamic extraction/chromatography devices with independent mixing and settling control, which offer a step change in efficiency.* J Chromatogr A, 2010. **1217**(1): p. 34-39.
28. Marlot, L., *KD-values in CPC: is prediction reliable ?,* oral presentation, the 9th International Countercurrent Chromatography Conference CCC2016, Chicago July 30th-August 3rd, 2016.

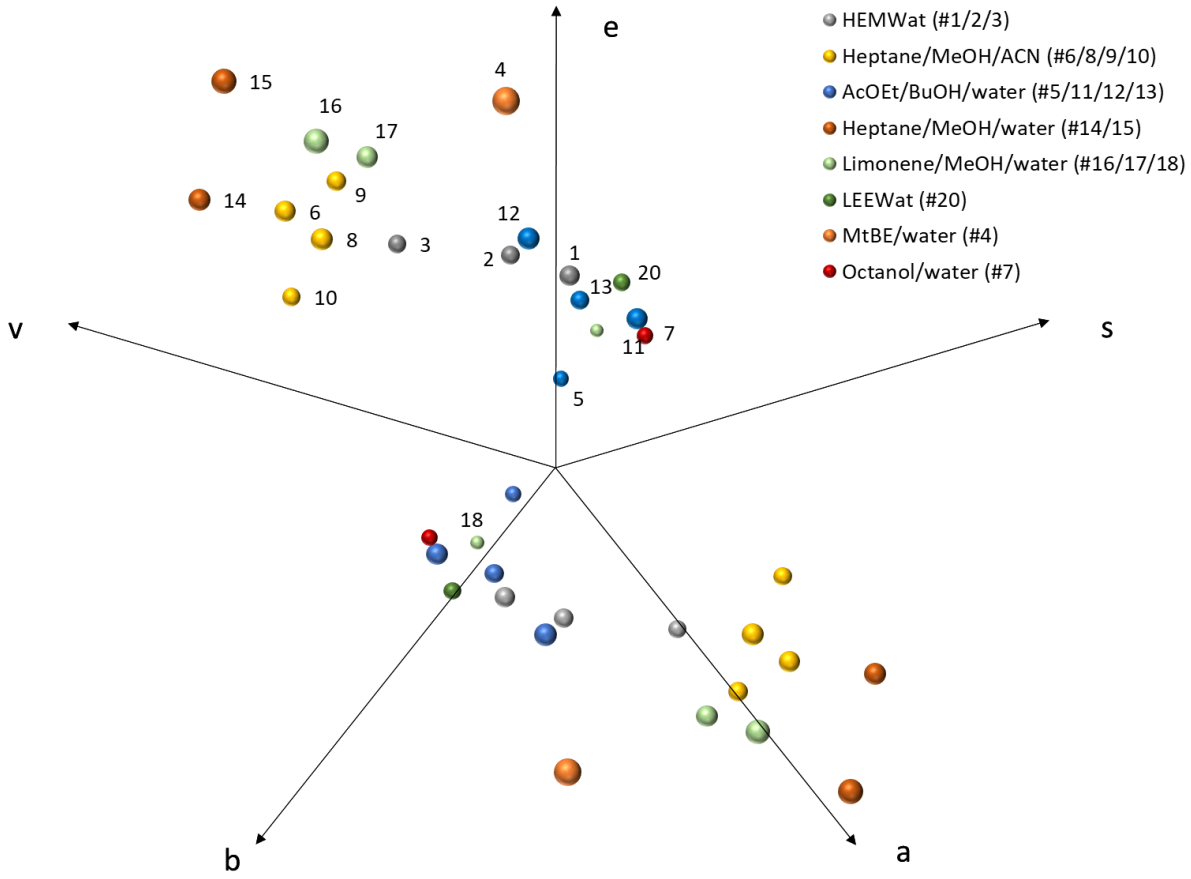
29.

Figure 1. Comparison of experimental and predicted log K values for the solvent system heptane/ethyl acetate/methanol/water 1/1/1/1 (system #2). Dashed lines represent log K ± 0.5 log unit, grey dots are the outliers, open circles are the validation set.



438 Figure 2 Spider diagram of the solvent systems in descending (numbered dots) and ascending

439 mode



440

441

Figure 3: Radar plot comparison of the solvent systems coefficients comparison and CCC chromatograms of syringol (1), 1-indanone (2), 2-naphtol (3) using the three solvent systems. #2: heptane/ethylacetate/methanol/water 1/1/1/1; #7: octanol/water; #15asc: heptane/methanol/water.

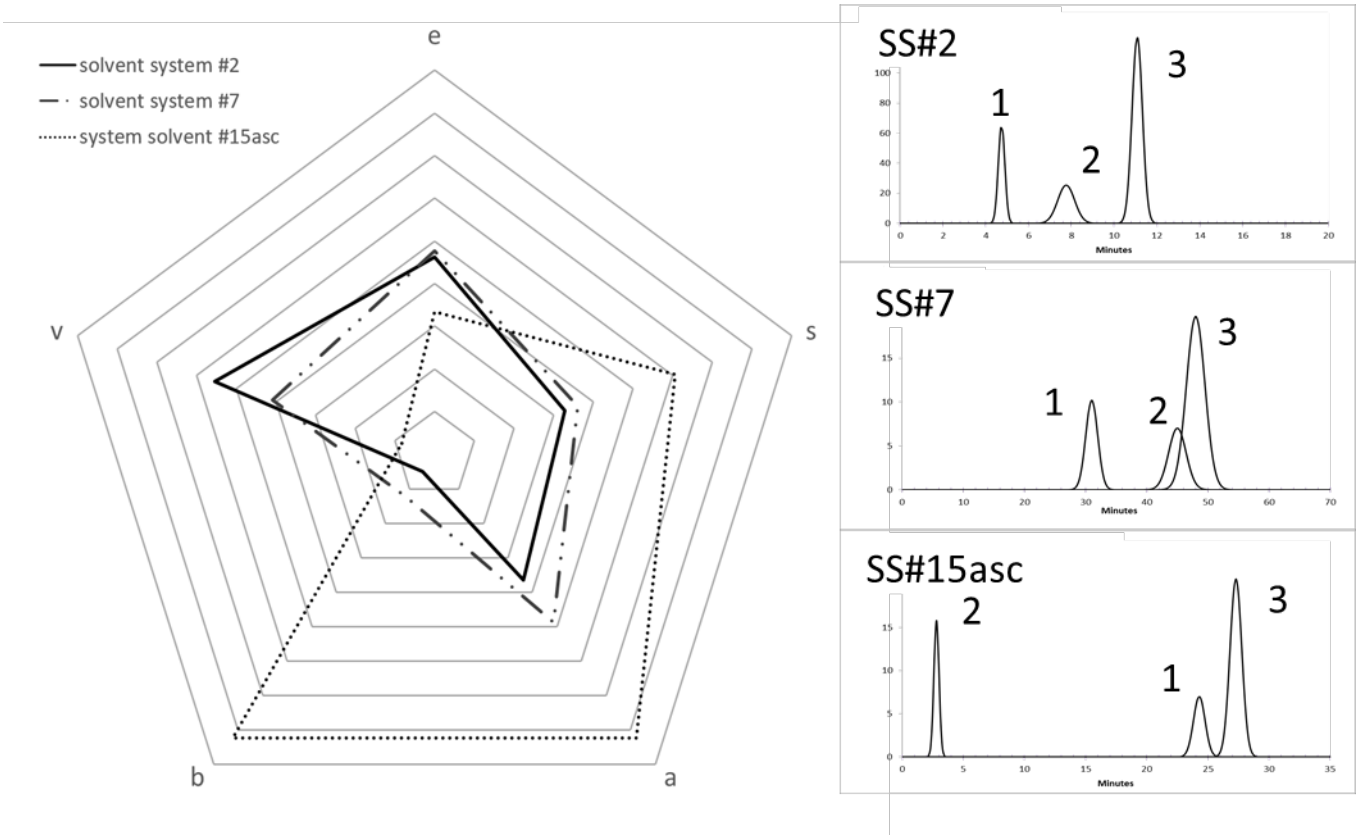
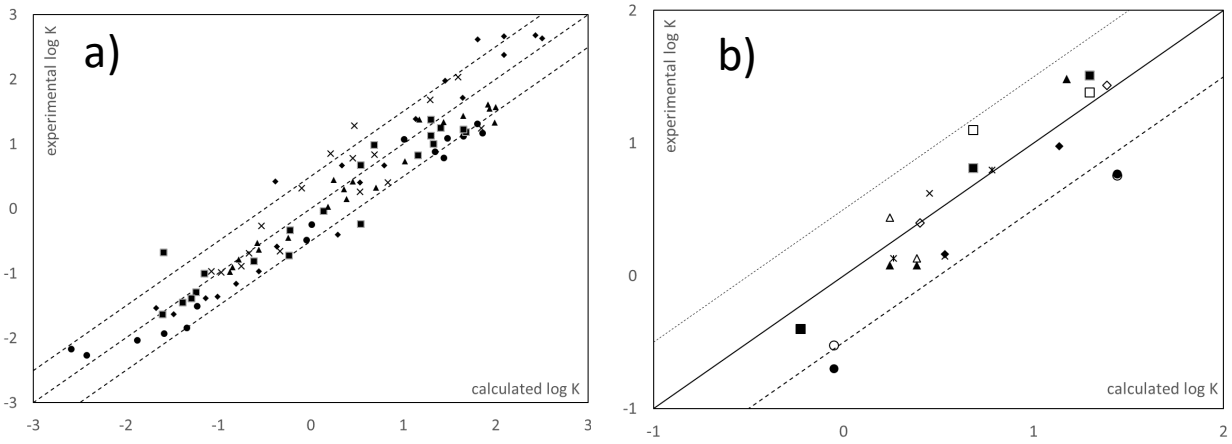


Figure 4: Comparison of experimental and calculated log K values for the validation set solutes in the 19 solvent systems. Validation set: catechol (circle), syringol (square); 1-indanone (triangle), 2-naphtol (diamond) and trimethylphenol (cross). Experimental values from a) partition experiments and b) CPC (full marks) or CCC (open marks) experiments.



461 **List of Tables and Figures**

462

466

483