I&EC research
Industrial & Engineering Chemistry Research

pubs.acs.org/IECR

Article

Predicting Partition Coefficients in Organic Biphasic Systems Using COSMO-RS

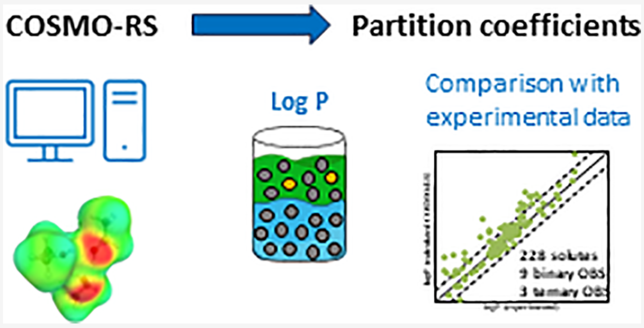
Rubén Santiago, Filipe Hobi Bordón Sosa, Ismael Díaz, María González-Miquel, and João Araújo Pereira Coutinho*

Cite This: https://doi.org/10.1021/acs.iecr.3c02303

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: To achieve a successful liquid−liquid extraction, the selection of the biphasic system is crucial for the proper separation of the solutes of interest. The number of biphasic systems that can be employed is vast, requiring a large amount of time and cost to experimentally determine the most suitable for each application. The use of computational methods to predict partition coefficients in biphasic systems is of great interest to design and select the most appropriate. COSMO-RS, a quantum chemical computational tool that requires only the chemical structure of the compounds for calculations, would be an ideal tool for that purpose. Here, we present a systematic evaluation of COSMO-RS as a predictive tool for partition coefficients. Its performance was evaluated for the partitioning of 228 solutes in 9 binary and 3 ternary organic biphasic systems (OBS). The results show that the use of COSMO-RS with TZVPD_FINE parametrization allows for very good predictions. They also show that predictions are greatly dependent on an accurate description of the compositions of the phases in equilibrium. Thus, the use of experimental mutual solubilities (binary OBS) or tie lines (ternary OBS) when available is the most suitable option for this purpose. In the case of using a total predictive tool, TZVPD_FINE parametrization can properly predict both mutual solubilities and tie lines, so it can also be used for the estimation of partition coefficients in an OBS. Therefore, the COSMO-RS method is demonstrated here to be a useful and reliable tool to predict partition coefficients in binary and ternary OBS, which can be used for the screening and selection of the most appropriate system to be used in a separation process.



■ INTRODUCTION

Countercurrent chromatography (CCC) and centrifugal partition chromatography (CPC) techniques mainly consist of a differential partitioning of solutes on a liquid−liquid biphasic system.[1] Instead of the use of a solid support to hold the stationary phase, in CPC the phase remains stationary by means of centrifugal forces while the mobile phase is pumped.[2] The compounds to be separated are differentially partitioned between the two phases and transferred at different velocities along the column. This enables the collection of the separated solutes, in solution, at the outlet stream. Traditionally, CPC has been used to separate natural compounds of very different natures from very rich matrixes.[3] They were reported for the purification of proteins,[4] peptides,[2] fatty acids,[5] and carotenoids,[6] among others. The optimal operation of a CCC or CPC requires partition coefficients in the range of 0.4 and 2.5.[7] Therefore, the identification of the most adequate biphasic system for each application is crucial.[2] In most applications, water is used as one compound of the biphasic system since it permits the formation of a very large number of biphasic systems because of its high cohesive energy.[8] However, it presents some limitations because all water-based biphasic systems are dominated by the extreme solvation properties of water, since no other solvent can compete with those.[9] Therefore, it is necessary to look for alternative systems to complement water-based biphasic systems.

The nonaqueous biphasic systems evaluated in the literature include organic binary and ternary systems such as *n*-hexane-acetonitrile,[10] hexane-trifluoroethanol,[11] heptane-methanol,[12] heptane-*N*,*N*-dimethylformamide,[13] and heptane-ethanol-glycerol/ethylene glycol/levulinic acid,[14] among others. Considering not only the vast number of organic biphasic systems but also the different initial concentrations that can be evaluated, the partition coefficients play a crucial role in characterizing the biphasic systems and in its selection. Some *a priori* methods have been evaluated to predict the partition coefficients and

Received: July 7, 2023
Revised: August 28, 2023
Accepted: October 3, 2023

Downloaded via UNIV OF AVEIRO 00300 on October 20, 2023 at 08:15:37 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

ACS Publications © XXXX American Chemical Society

select the proper biphasic system for a particular application.[15,16]

The COnductor like Screening MOdel for Real Solvents (COSMO-RS)[17] has been one of the most used for this purpose.[15,16,18] The great advantage of using COSMO-RS to calculate partition coefficients in biphasic systems is that the only information required is the chemical structure of each compound involved in the computation. It was shown that different approaches for calculating the partition coefficients can be used. Their principal difference is the way that the user defines the composition of each phase. It can be achieved in different ways: (i) assuming only the most abundant compound of each phase; (ii) using the experimental compositions of each phase; or (iii) calculating the phase compositions using COSMO-RS. Vilas-Boas et al.[16] reported high deviations when using the predictive COSMO-RS model with an average RMSD of 1.17, concluding that NRTL-SAC (semipredictive method) was a better way of predicting partition coefficients. Therefore, the authors suggested the use of COSMO-RS only as a preliminary solvent screening since no experimental data is required. Frey et al.[15] compared the capability of using COSMO-RS with UNIFAC for partition coefficient estimations. They examined the impact of incorporating experimental liquid−liquid equilibrium (LLE) data on the accuracy of predictions, comparing the results obtained when LLE data were used or not in the calculations. In all cases, the introduction of experimental data was shown to improve the results. They concluded that the UNIFAC model predicted the LLE with higher accuracy, resulting in better predictions. However, COSMO-RS was the only alternative for new systems as well as for functional groups for which UNIFAC binary interaction parameters are missing. Hopmann et al.[19] employed the COSMO-RS model to successfully select a biphasic system along with initial concentrations for piperine and coumarin extraction. However, to the best of our knowledge, no systematic studies have been proposed by considering different approaches to calculate partition coefficients in organic biphasic systems using COSMO-RS and evaluating them against a large database of experimental data.

The goal of this work is to systematically evaluate the capability of the COSMO-RS method to predict partition coefficients of a variety of solutes in 12 total organic biphasic systems (OBS) including binary and ternary mixtures. A collection of 228 solutes with experimental partition coefficients in the OBS was compiled for this purpose. For the COSMO-RS calculations, the influence of the different parametrization levels (TZVP and TZVPD_FINE) was analyzed. Moreover, the description of the compositions of each phase (infinite dilution and experimental and calculated mutual solubilities in the case of binary mixtures) was studied to elucidate the possibility of using the COSMO-RS method as a full predictive tool for partition coefficient estimation in the OBS. In the case of ternary mixtures, different approaches were adopted to describe the binodal curve and tie lines by defining ethanol as the most stable monomer, dimer, and trimer conformations and a mixture of them. The compositions of each phase were defined by using both experimental and calculated data. The results of this work aim to be a guideline for the most accurate and reliable approach to predicting partition coefficients in OBS.

■ COMPUTATIONAL DETAILS

Geometry Optimization. Turbomole v4.5.2 software was employed to optimize the structure of the ethanol trimer conformation at two different computational levels, BP86/TZVP and BP86/TZVPD_FINE, by using the COSMO continuum solvation method (solvent effect). From these calculations, two different files were obtained (two *.cosmo files with the COSMO continuum solvation method in the two different computational levels) to be used in the COSMOtherm software. Finally, a single point calculation was carried out in order to ensure that no vibrational frequencies are found; therefore, the most stable conformer is employed.

COSMO-RS Computations. The COSMO-RS method was used for all of the calculations carried out in this work. To do so, COSMOtherm v.21 along with its graphical interface COSMOthermX was employed for computations. The two parametrizations included in COSMOtherm, i.e., BP_TZVP_21 and BP_TZVPD_FINE_21, were evaluated to give guidelines for the prediction of partition coefficients in organic biphasic systems. The main differences between these two parametrizations are for BP_TZVPD_FINE_21: a larger basis set for calculations (TZVPD), a novel hydrogen bond interaction term, and a novel type of molecular surface cavity construction (FINE cavity as a tetrahedron cavity) creating σ-surfaces with more uniform segments.[20] In all cases, the calculated data were compared to the experimental data available in the literature. All of the compounds studied in this work are present in the COSMObase21 (extended) database, and their most stable conformers (lowest energy) were the only compounds considered for calculations. The only exception was carried out while using levulinic acid in which they were considered both the most stable conformer (−OH group faced with ketone) and the second most stable conformer (−OH group completely available) for comparison purposes (Figure S5 and S6 in the Supporting Information). Moreover, in the case of the ternary biphasic systems, we attempted to model the binodal curves by the inclusion of an ethanol trimer instead of the ethanol most stable conformer (Figure S4 in the Supporting Information).

Prediction of Partition Coefficients. The partition coefficients (P_i) in this work were calculated by using COSMO-RS. They are directly related to the activity coefficients at infinite dilution (γ_i^{∞}) of the species i involved in the calculations by eq 1

$$P_i = \frac{x_i^{I}}{x_i^{II}} = \frac{\gamma_i^{II,\infty}}{\gamma_i^{I,\infty}} \quad (1)$$

where the subscripts I and II refer to the upper and lower phases of the biphasic system, respectively, and x_i is the mole fraction of this compound "i". By default, the COSMO-RS method will compute the chemical potentials of the solutes at infinite dilution for each phase. Moreover, it is possible to define finite concentrations (of each phase) at which the chemical potentials of the compounds are calculated. This is crucial if the two solvent phases are partially soluble in each other. To do so, in this work, different alternatives were adopted in order to find out which option is the best for partition coefficient calculations. The partition coefficients at infinite dilution were calculated by specifying each phase composition with both experimental and calculated data (at BP_TZVP_21 and BP_TZVPD_FINE_21 parametrization levels). Different calculations must be done to define the

https://doi.org/10.1021/acs.iecr.3c02303
Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

Table 1. Types of Calculations and Nomenclature Followed in This Work for Binary OBS

Nomenclature	Calculation type
TZVP_ID	BP_TZVP_21 at infinite dilution
TZVP_EXP_MS	BP_TZVP_21 with experimental mutual solubilities
TZVPDFINE_PRED_FINE_MS	BP_TZVPD_FINE_21 with predicted mutual solubilities (BP_TZVPD_FINE_21)
TZVPDFINE_PRED_TZVP_MS	BP_TZVPD_FINE_21 with predicted mutual solubilities (BP_TZVP_21)
TZVPDFINE_EXP_MS	BP_TZVPD_FINE_21 with experimental mutual solubilities

Table 2. Types of Calculations and Nomenclature Followed in This Work for Ternary OBS

Nomenclature	Calculation type
TZVPD_FINE_Binodal Curve	LLE with BP_TZVPD_FINE_21 parametrization
TZVP_Binodal Curve	LLE with BP_TZVP_21 parametrization
Tie-Line_TZPD_FINE	tie line with BP_TZVPD_FINE_21 parametrization
Tie-Line_Exp	experimental tie line
EtOH_Dimer/Trimer_TZVPD_FINE	LLE with ethanol in dimer or trimer conformation with BP_TZVPD_FINE_21 parametrization
14%EtOH_Monomer_86%EtOH_Trimer	LLE with a mixture of 14% ethanol monomer and 86% ethanol trimer (BP_TZVPD_FINE_21 parametrization)

Table 3. Number of Experimental Solute Data for Biphasic Systems Evaluated

Binary systems		Ternary systems	
System	No. of exp data	System	No. of exp data
n-Hexane-Acetonitrile	129	*n*-Heptane-Ethanol-Ethylene glycol	4
n-Heptane-Methanol	73		
n-Heptane-*N*,*N*-Dimethylformamide	78		
n-Heptane-Dimethyl Sulfoxide	110	*n*-Heptane-Ethanol-Glycerol	4
n-Heptane-Ethylene Glycol	138		
n-Heptane-2,2,2-Trifluoroethanol	149		
n-Heptane-Propylene Carbonate	120	*n*-Heptane-Ethanol-Levulinic Acid	4
n-Heptane-Formamide	121		
n-Heptane-Ethanolamine	65		

compositions of upper and lower phases with respect to whether they have a binary or ternary biphasic system.

The deviation of predicted partition coefficients from experimental values was calculated for each case by means of the root-mean-square deviations (RMSD) as follows

$$\text{RMSD} = \sqrt{\left[\frac{\sum_i (\log P_i^{exp} - \log P_i^{calc})^2}{n}\right]} \quad (2)$$

where superscripts *exp* and *calc* referd to the experimental and calculated values, respectively, n is the total number of data points and i includes all of the systems evaluated.

Binary Biphasic Systems. The composition of each phase of the binary OBS was specified in partition coefficient calculations in different ways. The nomenclature employed in the whole work depending on the approach followed is detailed in Table 1. First, the infinite dilution approach was evaluated by defining the pure compounds in each phase using TZVP parametrization (TZVP_ID). Then, experimental mutual solubilities of the nine different OBS were collected and included in partition coefficients computations with both TZVP (TZVP_EXP_MS) and TZVPD_FINE (TZVPDFINE_EXP_MS) parametrizations. Finally, to elucidate the possibility of using a fully predictive tool, mutual solubilities were also calculated and included in partition coefficient estimations (with TZVPD_FINE parametrization). These mutual solubilities of the nine OBS were calculated by using TZVP (TZVPDFINE_PRED_TZVP_MS) and TZVPD_FINE (TZVPDFINE_PRED_FINE_MS) approaches. In all cases, the most stable conformers were used for calculations.

Ternary Biphasic Systems. In the case of ternary OBS, the first step was to attempt to model the liquid−liquid equilibrium (LLE) data (binodal curve and tie lines) for the initial compositions of the evaluated systems. To do so, different approaches were made whose nomenclature can be found in Table 2. First, TZVP and TZVPD_FINE parametrizations were used with the most stable conformers of each ternary system. Then, by using the initial compositions of the collected experimental data, tie lines were also calculated with TZVPD_FINE parametrization (Tie-Line TZVPD_FINE). Different ways of defining ethanol, as dimer or trimer (based on literature), were carried out and compared with the experimental data (EtOH_Dimer or Trimer_TZVPD_FINE). Finally, a mixture of 86% trimer and 14% monomer (based on the literature[21]) was also modeled with its corresponding tie line at the initial composition reported in the literature.[14] For all of the solutes and compounds of the ternary systems, the most stable conformers were used for calculations except for levulinic acid in which both the most stable, denoted as c0 (−OH group faced with ketone), and the second one, named as c1 (−OH group completely available), as included in the COSMObase21 database were used.

■ RESULTS

The evaluation of the capability of COSMO-RS to predict partition coefficients in binary and ternary organic biphasic systems was carried out. To do so, the first step was to collect

https://doi.org/10.1021/acs.iecr.3c02303
Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

the available experimental data in the literature.[22] Table 3 includes the nine binary OBS with the number of solutes measured for each system. It also shows the three ternary OBS with four natural compounds' experimental partition coefficients for each system. The name of each solute and the experimental and calculated data are shown in Tables S2−S10 of the Supporting Information along with their experimental and calculated data. The total number of solutes evaluated is 228, covering almost the whole range of partition coefficients for each OBS. Table 3 details the number of experimental data points for each OBS. The binary systems are mainly composed of an aliphatic compound (*n*-hexane or *n*-heptane) and polar compounds of very different natures. Meanwhile, the ternary systems present an aliphatic molecule (*n*-heptane), a short-chain alcohol (ethanol), and an alcohol-based compound such as ethylene glycol or glycerol or a carboxylic acid such as levulinic acid.

Binary Organic Biphasic Systems. The calculation of partition coefficients with COSMO-RS requires the composition definition of each phase. Therefore, different approaches were adopted to analyze the most accurate methodology to predict partition coefficients. Therefore, in order to do so, experimental mutual solubilities[23] for the systems evaluated (see Table 3) were collected and compared with those calculated with the COSMO-RS method (both TZVP and TZVPD_FINE approaches). Figure 1 shows the obtained

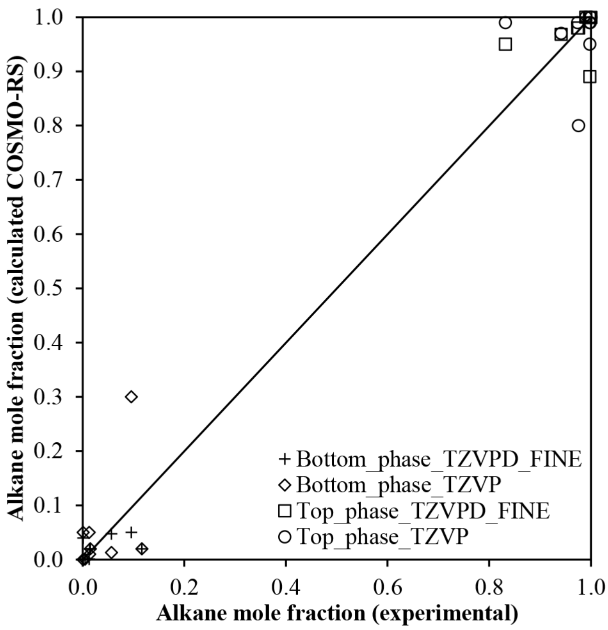


Figure 1. Calculated (TZVP and TZVPD_FINE parametrizations) vs experimental alkane (*n*-heptane or *n*-hexane) mole fraction in both phases (top and bottom) at 25 °C.

results for each phase, comparing the *n*-alkane composition. All of the data were compiled in Table S1 of the Supporting Information with their RMSDs and mean absolute deviations (MADs) for comparison purposes. As shown, in all cases the predictions were better for the top phase, which is the *n*-alkane-rich phase. The obtained RMSD in the case of using TZVPD_FINE parametrization was 0.04, and 0.08 was obtained for TZVP. Slightly worse predictions were found for the bottom phase, in which some discrepancies with the experimental data were found. The calculated RMSD by using TZVPD_FINE was 0.08, the same value found for TZVP. Overall, it can be seen that TZVPD_FINE parametrization offered slightly better predictions when compared with TZVP, so it seems to be the one to be selected for this kind of calculation. When estimating mutual solubilities in binary systems, it was observed that there were no notable differences in calculation time between the TZVP and TZVPD_FINE parametrization levels. However, the TZVPD_FINE parametrization was identified as the most accurate method for describing the behavior of the evaluated binary mixtures. It was shown that reasonably good predictions are found. However, it can be improved by including all of the polar compound conformers (when available) in calculations.

Once the mutual solubilities of all of the systems were calculated and compared with experimental data, the next step was the calculation of the partition coefficients of all of the solutes in the nine OBS in the various ways described in Table 1. Figure 2 shows the experimental versus calculated partition coefficients for the *n*-heptane-methanol system (A) and *n*-heptane-3,3,3-trifluoroethanol (B) with the different calculation types proposed in Table 1. Figure 2A demonstrates that the partition coefficient range of the evaluated solutes is relatively narrow, spanning log P_i − 2 to 1. Moreover, higher deviations are found by using the infinite dilution compositions (RMSD = 0.89). This is attributed to the high mutual solubilities between *n*-heptane and methanol (Table S1 in the Supporting Information). Thus, the compositions of the phases in equilibrium are not well described by the infinite dilution approximation, and thus the partition coefficient predictions are not accurate. The introduction of the experimental composition of each phase reveals a great enhancement in predictions, being TZVPD_FINE parametrization, the one exhibiting the best performance (RMSD = 0.24). Finally, in order to have a fully predictive tool, mutual solubilities were also calculated with TZVP and TZVPD_FINE parametrizations. As previously revealed, TZVPD_FINE parametrization can predict mutual solubilities better than TZVP. This influences the log P_i estimations because this approach is the one with the lowest deviation (RMSD = 0.46). All in all, in the case of *n*-heptane-methanol OBS, TZVPD_FINE parametrization offers the best predictions by including the experimental mutual solubilities. In the case of Figure 2B, a very different polar compound was evaluated containing three fluorine atoms on its structure. As seen, the log P_i range is wider than that of methanol with values of between −3 and 2 with a huge number of solutes (149). Starting with the infinite dilution approach, again it offers the worst results (RSMD = 0.74) from those evaluated; however, the results are not as far as in the case of methanol since mutual solubilities between the two compounds of the OBS are much lower. The use of experimental mutual solubilities leads to the best log P_i predictions (RMSD = 0.41), but in this case, TZVPD_FINE parametrization can predict mutual solubilities with good accuracy, translating into log P_i estimations (RMSD = 0.41) that are as good as those based on experimental data. Therefore, in this OBS, the COSMO-RS method can be used to estimate log P_i in a fully predictive way without using any experimental data. Regarding the RMSD found, there is a lack of consensus about under what deviation may be admissible. For instance, Kenney et al.[24] reported RMSD = 1 as a good value for octanol−water partition coefficient predictions. Instead, Vilas-Boas et al. claimed that acceptable RMSD values should be lower than 0.5. In any case, the results reported here show that COSMO-

https://doi.org/10.1021/acs.iecr.3c02303
Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

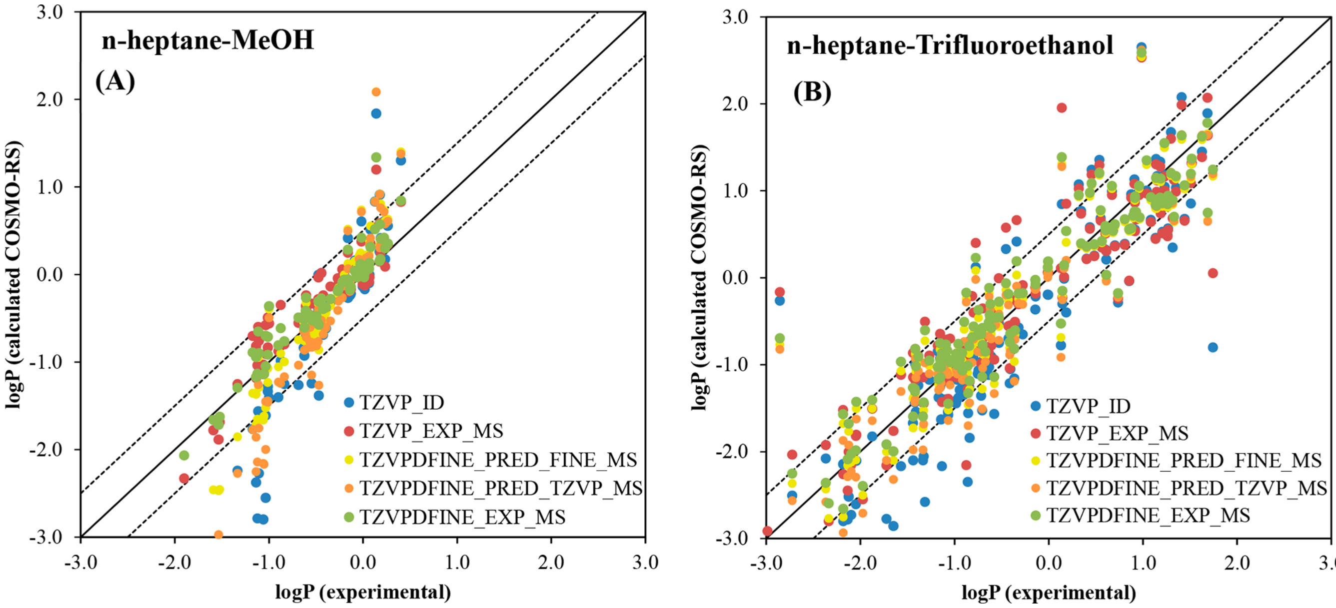


Figure 2. Calculated vs experimental partition coefficients of (A) the *n*-heptane-methanol system and (B) *n*-heptane-3,3,3-trifluoroethanol at 25 °C by using different parametrizations and mutual solubilities definition (see Table 2). Dotted lines are log *P* ± 0.5. Blue (TZVP_ID) means both phases are defined as the major pure compound of each phase and log *P* calculated with TZVP; red (TZVP_EXP_MS) phases defined with experimental mutual solubilities and log *P* calculated with TZVP; yellow (TZVPDFINE_PRED_FINE_MS) calculated mutual solubilities with TZVPD_FINE and log *P* with TZVPD_FINE; orange (TZVPDFINE_PRED_TZVP_MS) calculated mutual solubilities with TZVP and log *P* with TZVPD_FINE; green experimental mutual solubilities and log *P* calculated with TZVPD_FINE.

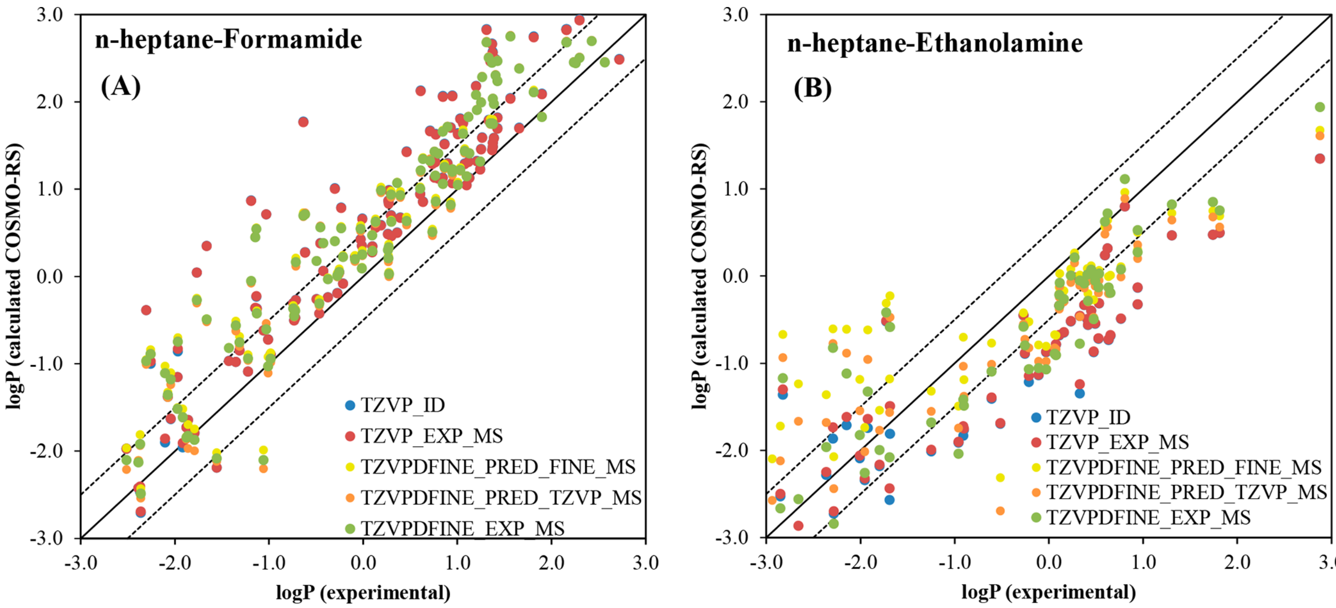


Figure 3. Calculated vs experimental partition coefficients of (A) *n*-heptane-formamide system and (B) *n*-heptane-ethanolamine at 25 °C by using different parametrizations and mutual solubilities definition (see Table 2). Blue (TZVP_ID) means both phases are defined as the major pure compound of each phase and log P calculated with TZVP; red (TZVP_EXP_MS) phases defined with experimental mutual solubilities and log P calculated with TZVP; yellow (TZVPDFINE_PRED_FINE_MS) calculated mutual solubilities with TZVPD_FINE and log P with TZVPD_FINE; orange (TZVPDFINE_PRED_TZVP_MS) calculated mutual solubilities with TZVP and log P with TZVPD_FINE; green experimental mutual solubilities and log P calculated with TZVPD_FINE.

RS predictions may be accurate enough to select the OBS for separations.

Figure 3 shows the predicted versus experimental partition coefficients of different solutes in *n*-heptane-formamide (A) and *n*-heptane-ethanolamine (B) OBS. In Figure 3A, 121 solutes were analyzed in a wide log *P* range (between −3 and 3). In this case, since the mutual solubilities between *n*-heptane and formamide are very low (see Table S1 in the Supporting Information), the infinite dilution (TZVP_ID) approximation offers a very good alternative. Therefore, partition coefficient estimations are very similar when comparing the hypothesis of infinite dilution (RMSD = 0.88) with the use of experimental solubilities (RMSD = 0.87). Once again, the use of TZVPD_FINE parametrization clearly enhances the predic-

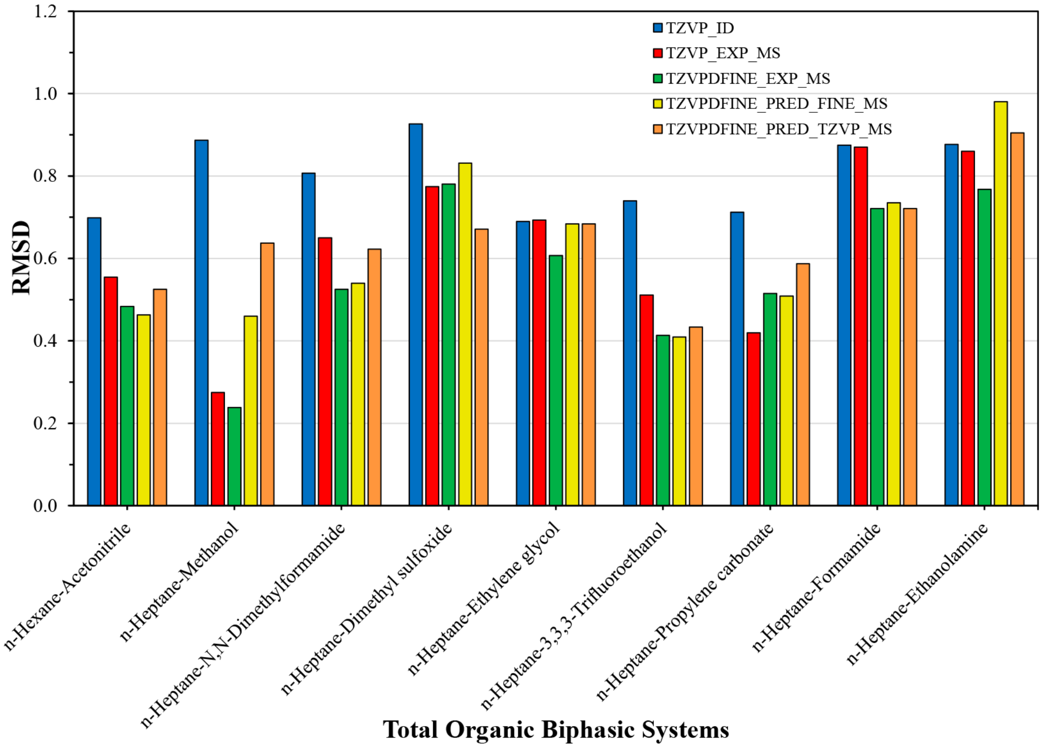


Figure 4. RMSD of predicted partition coefficient of the nine total organic biphasic systems evaluated in this work by using different parametrizations and the mutual solubilities definition (see Table 2). Blue (TZVP_ID) means both phases are defined as the major pure compound of each phase and log P calculated with TZVP; red (TZVP_EXP_MS) phases are defined with experimental mutual solubilities and log P calculated with TZVP; yellow (TZVPDFINE_PRED_FINE_MS) represents calculated mutual solubilities with TZVPD_FINE and log P with TZVPD_FINE; orange (TZVPDFINE_PRED_TZVP_MS) represents calculated mutual solubilities with TZVP and log P with TZVPD_FINE; green represents experimental mutual solubilities and log P calculated with TZVPD_FINE.

tions revealing RMSD = 0.72. Regarding the fully predictive approach, the COSMO-RS method can predict quite well the mutual solubilities, and therefore log P_i are in good agreement with the experimental data with RMSD = 0.74, very similar to that obtained by introducing experimental data. In this case, with a solvent containing an amine group, predictions are worse compared with previous OBS evaluated, but trends are well described with this methodology. In this case, both infinite dilution and the fully predictive approaches provide results as good as those obtained by introducing experimental data.

The results for the OBS *n*-heptane-ethanolamine reported in Figure 3B show a huge dispersion in the predicted partition coefficient values. In this case, as before, the infinite dilution approximation offers equivalent results (RMSD = 0.88) to those obtained with the inclusion of experimental data (RMSD = 0.86). Again, these results are from the very low mutual solubilities between *n*-heptane and ethanolamine. In this case, the COSMO-RS method cannot properly predict their mutual solubilities, and thus, it leads to worse predictions (RMSD = 0.98). Analyzing the results of two OBS-containing compounds with amine and amide groups, it can be concluded that the COSMO-RS method cannot predict the partition coefficients as well as in the other systems discussed above. Moreover, their very low mutual solubility offers the possibility of using the infinite dilution approximation to predict the log P_i. Regarding the rest of the OBS systems, the results are detailed in Figures S1−S3 of the Supporting Information. The same general conclusions can be drawn from those figures. In order to analyze all the systems together, Figure 4 shows the RMSD of the nine OBS with the various approaches evaluated here. In summary, in all of the evaluated systems, by using the best approach reported, more than 94% of the evaluated solutes present deviations lower than 1 in log units.

The results reported in Figure 4 show that the use of the infinite dilution approach does not provide log P predictions with good accuracy except for the case of the compounds containing amino groups (formamide and ethanolamine) and ethylene glycol. In these cases, their mutual solubilities with *n*-heptane are very low, implying that predictions are good enough. The mean RMSD values obtained for the nine OBS is 0.80. Therefore, this approach can be used when mutual solubilities of the OBS are low. Overall, the introduction of experimental mutual solubilities in the calculations enhances the predictions since the compositions of the two phases in equilibrium are properly described (mean RMSD = 0.62). Moreover, in general terms, the use of TZVPD_FINE parametrization shows better predictions than TZVP (RMSD = 0.56). Regarding the use of a fully predictive approach, TZVPD_FINE and TZVP parametrizations were used to calculate mutual solubilities between the two solvents conforming the OBS. In this case, the use of TZVPD_FINE is the best option with a mean RMSD of 0.62. As discussed, the results are slightly worse compared with the use of experimental solubilities. However, it seems to be a good alternative when lacking mutual solubility experimental data. Overall, COSMO-RS can be used as a fully predictive tool to calculate partition coefficients in binary OBS. Nevertheless, in the cases in which experimental mutual solubilities are available, their inclusion in calculations improves the predictions.

https://doi.org/10.1021/acs.iecr.3c02303
Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

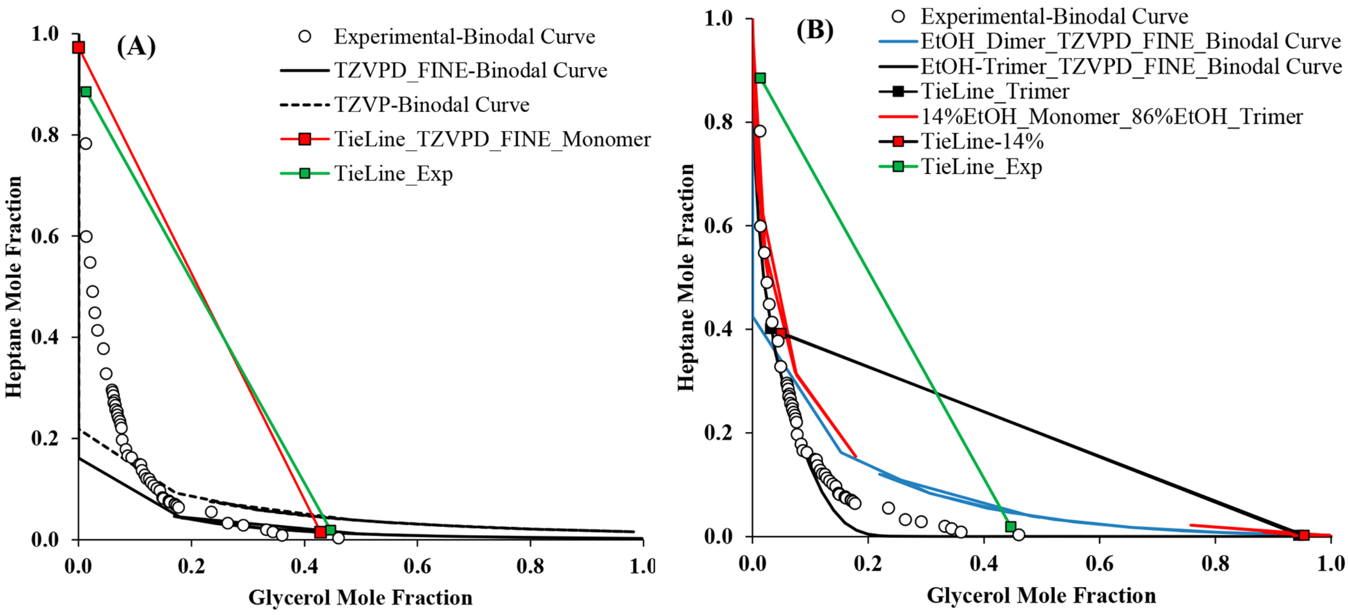


Figure 5. Experimental versus calculated binodal curves and tie lines by using ethanol monomer (A) and the ethanol trimer approximation (B) with TZVP and TZVPD_FINE parametrizations at 25 °C in *n*-heptane/ethanol/glycerol OBS.

Ternary Organic Biphasic Systems. The next step is the evaluation of the capability of COSMO-RS to predict log P_i in ternary OBS. The methodology previously used was employed, but in this case, instead of calculating and defining mutual solubilities, the binodal curves and tie lines of the ternary system were obtained. Three ternary OBS were evaluated as reported in Table 3.[16] In all cases, it involves an aliphatic compound (*n*-heptane), a short alkyl chain alcohol (ethanol), and a third polar compound such as glycerol, ethylene glycol, or levulinic acid. Regarding the experimental tie lines, we measured them for the three ternary OBS with initial concentrations of 37.5 wt % (*n*-heptane), 25 wt % (ethanol), and 37.5 wt % (polar compound) with the methodology described in the Supporting Information. The experimental tie line results can be found in Tables S11−S13 in the Supporting Information. Figure 5 shows the experimental vs calculated (TZVP and TZVPD_FINE) binodal curve and tie lines in different ways: (A) defining ethanol as a monomer and (B) dimer, trimer, and mixtures in the *n*-heptane/ethanol/glycerol OBS system *n*-heptane/ethanol/glycerol. First, starting with Figure 5A, it can be seen how, by using the most stable conformer of ethanol to model the OBS, the COSMO-RS method can predict with higher accuracy the composition of the binodal curve in the alcohol-rich phase (bottom phase). However, both TZVP and TZVPD_FINE parametrizations predict a partial miscibility between ethanol and *n*-heptane, which leads to the wrong estimation of the binodal curve in the *n*-heptane-rich phase (top phase). However, despite the wrong estimation of the binodal curve in the top-phase region, it can be observed how the calculated tie line matches with high accuracy the experimental one. In this case, TZVPD_FINE parametrization offers the best estimations when compared with the experimental data.

This poor description of the mutual solubilities of these compounds can be attributed to the formation of clusters by the ethanol molecule. Balanay et al.[21] conducted a study revealing that density functional theory (DFT) results suggest the thermodynamic preference for cyclic clusters (trimers and tetramers) of ethanol molecules over chain-shaped clusters. Furthermore, a combination of Fourier-transform infrared spectroscopy (FT-IR) and proton nuclear magnetic resonance (^{1}H NMR) techniques provided substantial evidence supporting the existence of a cyclic-cluster-dominated microscopic structure in pure ethanol and ethanol solutions within nonpolar solvents. At this point, we tried to improve COSMO-RS method predictions of this OBS by modeling ethanol as a trimer, based on previous DFT-experimental measurements in the literature.[21] To do so, we optimized the most stable trimer conformation proposed in the literature, in which all the −OH groups are completely blocked (Figure S4 in the Supporting Information).

Figure 5B shows the results obtained. First, it can be observed how the definition of ethanol as a dimer conformation (blue line) cannot properly describe the solubility between *n*-heptane and ethanol (almost the same as using an ethanol monomer). In the case of using the ethanol trimer conformation (black line), the top-phase region is perfectly described. This is because with this conformation the oligomer is highly nonpolar (see Figure S5 in the Supporting Information for the σ-profile of monomer and trimer) and thus more like *n*-heptane, better describing their mutual solubility. However, it cannot properly predict the bottom phase. This is directly translated into the tie lines. Therefore, the predicted tie line when using the trimer conformation is far from the experimental one, calculating almost pure glycerol in the bottom phase and a mixture of ethanol and *n*-heptane in the top phase. Following the literature and based on DFT calculations, it is suggested that ethanol is defined as a mixture of 14% monomer and 86% trimer.[21] For this reason, we tried to model that mixture in order to properly estimate the whole binodal curve and tie lines. Using this approach, the top-phase region is well described by the mixture model. However, we were not able to model the bottom phase with this approach, so the phase composition of the tie line is, again, far from the experimental measurement.

The predictions of the binodals for the two other ternary OBS are presented in Figures S7 and S8 of the Supporting Information, with it being possible to draw the same

https://doi.org/10.1021/acs.iecr.3c02303
Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

conclusions as in the case of the *n*-heptane/ethanol/glycerol system. All in all, it can be concluded that by using the most stable monomer conformation of ethanol, the binodal curves are partially well described (good predictions in the bottom phase but bad in the top phase). Regarding the tie lines, TZVPD_FINE can provide satisfactory predictions of the experimental values. The inclusion of the trimer in the calculations corrects the predictions of the top phase but then the bottom-phase region is not properly described, implying that the tie line is far from the experimental behavior.

Once the liquid−liquid equilibria were modeled in different ways, the next step was the estimation of the partition coefficients of the different solutes. Figure 6 shows the

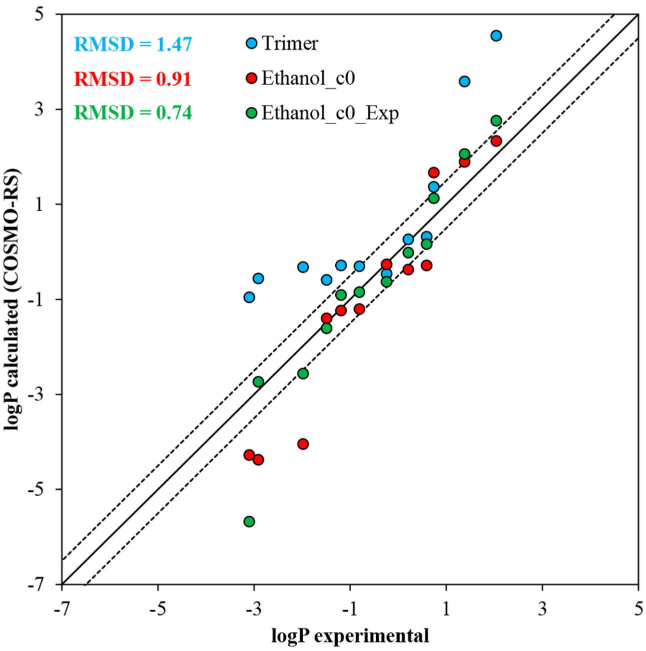


Figure 6. Calculated (TZVPD_FINE) versus experimental partition coefficients of the solutes in the three ternary biphasic systems at 25 °C by using different ways of defining ethanol and the compositions of both phases (predicted and experimental). Blue (Trimer) denotes that calculations were made by using the trimer conformation with the total predictive tool (tie line). Red (Ethanol_c0) shows the calculations while using the most stable ethanol monomer conformation with the total predictive (tie-line) tool, and green (Ethanol_c0_Exp) refers to calculations that were made by using the ethanol most stable monomer conformation and experimental tie line.

experimental versus the calculated log P_i by using TZVPD_FINE parametrization with the ethanol trimer (Trimer in the legend) and monomer (Ethanol_c0 in the legend), which are conformations for estimating both the tie line and log P_i and the ethanol monomer for log P_i estimations including the experimental tie line (Ethanol_c0_Exp in the legend) for calculations.

The results reported in Figure 6 show that the best predictions for the partition coefficients were obtained by using the ethanol monomer and the experimental compositions of the two phases. The results using the trimer conformation (in blue) present a deviation from the experimental results caused by the wrong definition of the tie line. Results using the ethanol monomer in a total predictive way (in red) provide a better description of the experimental data because the corresponding OBS tie line is much better described. Finally, the use of the experimental tie line makes the predictions (in green) even better, concluding that the accurate description of the compositions of the two phases is essential for the reliable predictions of log P_i in ternary OBS. However, when lacking LLE experimental data, the use of TZVPD_FINE parametrization offers satisfactory results that allow us to compare the performance of the different solutes in OBS.

■ CONCLUSIONS

This work presents a systematic evaluation of the COSMO-RS method in the prediction of partition coefficients in binary and ternary organic biphasic systems. A total number of partition coefficients of 228 solutes in 9 binary and 3 ternary OBS were collected from the literature. Regarding binary OBS, it was demonstrated that TZVPD_FINE parametrization offered the most accurate results for all of the solute systems evaluated. We found that the adequate description of the mutual solubilities is crucial, and thus, only for systems with very low mutual solubilities (such as ethanolamine and formamide in *n*-heptane) would the infinite dilution approach be suitable. Therefore, whenever possible, the inclusion of experimental mutual solubilities significantly improves the COSMO-RS method predictions (RMSD = 0.56). A fully predictive methodology by also predicting the mutual solubilities was also demonstrated to be possible while using TZVPD_FINE parametrization with a mean RMSD of 0.62. In the case of ternary OBS, the capability of predicting tie lines was the key to obtaining good partition coefficient estimations. However, in all other cases, the estimations were observed to deviate significantly from the experimental data. It was shown that using the most stable ethanol monomer with TZVPD_FINE parametrization provided an adequate description of the experimental tie line, which means that the two phases were well described. All of this was directly translated into the estimation of the partition coefficients, with an RMSD of 0.91 when compared with the trimer conformer (RMSD = 1.47). Again, the inclusion of the experimental data enhanced the predictions, obtaining an RMSD of 0.74. Therefore, the capability of using COSMO-RS with TZVPD_FINE parametrization was demonstrated for the prediction of the partition coefficients in binary and ternary OBS. Moreover, the use of experimental data to define the compositions of both phases improves calculations, but a complete predictive tool can also be used with good estimations.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.iecr.3c02303.

Figures and tables containing all of the experimental and calculated data for all of the OBS evaluated in the work; experimental and calculated mutual solubilities; σ-surfaces and σ-profiles of the ethanol monomer and trimer and two levulinic acid conformers; and HPLC methodology followed and experimental data obtained of tie lines in ternary OBS (PDF)

■ AUTHOR INFORMATION

Corresponding Author

João Araújo Pereira Coutinho − *CICECO* − *Aveiro Institute of Materials, Department of Chemistry, University of Aveiro, 3810-193 Aveiro, Portugal*; Email: jcoutinho@ua.pt

Authors

Rubén Santiago − *Departamento de Ingeniería Química, Universidad Autónoma de Madrid, Madrid 28049, Spain; CICECO − Aveiro Institute of Materials, Department of Chemistry, University of Aveiro, 3810-193 Aveiro, Portugal;* orcid.org/0000-0002-6877-9001

Filipe Hobi Bordón Sosa − *CICECO − Aveiro Institute of Materials, Department of Chemistry, University of Aveiro, 3810-193 Aveiro, Portugal*

Ismael Díaz − *Departamento de Ingeniería Química Industrial y del Medio Ambiente, ETSI Industriales, Universidad Politécnica de Madrid, Madrid 28006, Spain*

María González-Miquel − *Departamento de Ingeniería Química Industrial y del Medio Ambiente, ETSI Industriales, Universidad Politécnica de Madrid, Madrid 28006, Spain*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.iecr.3c02303

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was developed within the scope of project CICECO-Aveiro Institute of Materials (UIDB/50011/2020, UIDP/50011/2020, and LA/P/0006/2020), financed by national funds through the FCT/MCTES (PIDDAC). Rubén Santiago thanks Ministerio Universidades for his Margarita Salas contract (CA1/RSUE/2021-00585). Filipe H. B. Sosa acknowledges FCT − Fundação para a Ciência e a Tecnologia, I.P. for researcher contract CEECIND/07209/2022 under the Scientific Employment Stimulus - Individual Call. Authors from Universidad Politécnica de Madrid acknowledge Community of Madrid (Spain) for funding through the Multiannual Agreement with Universidad Politécnica de Madrid in the line Excellence Programme for University Professors, in the context of the V PRICIT (Regional Programme of Research and Technological Innovation).

REFERENCES

(1) Murayama, W.; Kobayashi, T.; Kosuge, Y.; Yano, H.; Nunogaki, Y.; Nunogaki, K. J. J. o. C. A. A new centrifugal counter-current chromatograph and its application. *J. Chromatogr., A* **1982**, *239*, 643−649.

(2) Marchal, L.; Legrand, J.; Foucault, A. Centrifugal partition chromatography: A survey of its history, and our recent advances in the field. *Chem. Record* **2003**, *3* (3), 133−143.

(3) Chagnoleau, J.-B.; Rocha, I. L.; Khedher, R.; Coutinho, J. A.; Michel, T.; Fernandez, X.; Papaiconomou, N. J. J. o. C. A. Separation of natural compounds using eutectic solvent-based biphasic systems and centrifugal partition chromatography. *J. Chromatogr. A* **2023**, 463812, DOI: 10.1016/j.chroma.2023.463812.

(4) Bezold, F.; Goll, J.; Minceva, M. Study of the applicability of non-conventional aqueous two-phase systems in counter-current and centrifugal partition chromatography. *J. Chromatogr A* **2015**, *1388*, 126−132.

(5) Marchal, L.; Intes, O.; Foucault, A.; Legrand, J.; Nuzillard, J. M.; Renault, J. H. Rational improvement of centrifugal partition chromatographic settings for the production of 5-n-alkylresorcinols from wheat bran lipid extract. I. Flooding conditions-optimizing the injection step. *J. Chromatogr A* **2003**, *1005* (1−2), 51−62.

(6) Goncalves de Oliveira-Junior, R.; Grougnet, R.; Bodet, P.-E.; Bonnet, A.; Nicolau, E.; Jebali, A.; Rumin, J.; Picot, L. Updated pigment composition of Tisochrysis lutea and purification of fucoxanthin using centrifugal partition chromatography coupled to flash chromatography for the chemosensitization of melanoma cells. *Algal Res.* **2020**, *51*, 102035.

(7) Adelmann, S.; Baldhoff, T.; Koepcke, B.; Schembecker, G. Selection of operating parameters on the basis of hydrodynamics in centrifugal partition chromatography for the purification of nybomycin derivatives. *J. Chromatogr A* **2013**, *1274*, 54−64.

(8) Abraham, M. H.; Acree, W. E. Gas-solvent and water-solvent partition of trans-stilbene at 298K. *J. Mol. Liq.* **2017**, *238*, 58−61.

(9) Poole, C. F.; Karunasekara, T.; Ariyasena, T. C. Totally organic biphasic solvent systems for extraction and descriptor determinations. *J. Sep. Sci.* **2013**, *36* (1), 96−109.

(10) Isidorov, V. A.; Krajewska, U.; Vinogorova, V. T.; Vetchinnikova, L. V.; Fuksman, I. L.; Bal, K. Gas chromatographic analysis of essential oil from buds of different birch species with preliminary partition of components. *Biochemical Systematics and Ecology* **2004**, *32* (1), 1−13.

(11) Zenkevich, I. G. Chromato-mass-spectrometric identification using partition coefficients in the system of hexane-2,2,2-trifluoroethanol. *Russian Journal of General Chemistry* **2012**, *82* (8), 1391−1399.

(12) Marchal, L.; Legrand, J.; Foucault, A. Mass transport and flow regimes in centrifuga partition chromatography. *AIChE J.* **2002**, *48*, 1692−1704.

(13) Ramil Criado, M.; Rodrıguez Pereiro, I.; Cela Torrijos, R. Selective determination of polychlorinated biphenyls in waste oils using liquid-liquid partition followed by headspace solid-phase microextraction and gas chromatography with atomic emission detection. *Journal of Chromatography A* **2004**, *1056* (1), 263−266.

(14) Chagnoleau, J.-B.; Papaiconomou, N.; Jamali, M.; Abranches, D. O.; Coutinho, J. A. P.; Fernandez, X.; Michel, T. Toward a Critical Evaluation of DES-Based Organic Biphasic Systems: Are Deep Eutectic Solvents so Critical? *ACS Sustainable Chem. Eng.* **2021**, *9* (29), 9707−9716.

(15) Frey, A.; Hopmann, E.; Minceva, M. Selection of Biphasic Liquid Systems in Liquid-Liquid Chromatography Using Predictive Thermodynamic Models. *Chem. Eng. Technol.* **2014**, *37* (10), 1663−1674.

(16) Vilas-Boas, S. M.; Cordova, I. W.; Kurnia, K. A.; Almeida, H. H. S.; Gaschi, P. S.; Coutinho, J. A. P.; Pinho, S. P.; Ferreira, O. Comparison of two computational methods for solvent screening in countercurrent and centrifugal partition chromatography. *J. Chromatogr A* **2022**, *1666*, 462859.

(17) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99* (7), 2224−2235.

(18) Hopmann, E.; Arlt, W.; Minceva, M. Solvent system selection in counter-current chromatography using conductor-like screening model for real solvents. *J. Journal of Chromatography A* **2011**, *1218* (2), 242−250.

(19) Hopmann, E.; Frey, A.; Minceva, M. A priori selection of the mobile and stationary phase in centrifugal partition chromatography and counter-current chromatography. *Journal of Chromatography A* **2012**, *1238*, 68−76.

(20) Paduszyński, K. An overview of the performance of the COSMO-RS approach in predicting the activity coefficients of molecular solutes in ionic liquids and derived properties at infinite dilution. *Phys. Chem. Chem. Phys.* **2017**, *19* (19), 11835−11850.

(21) Balanay, M. P.; Kim, D. H.; Fan, H. Revisiting the formation of cyclic clusters in liquid ethanol. *J. Chem. Phys.* **2016**, *144* (15), 154302.

(22) Poole, C. F. Partition constant database for totally organic biphasic systems. *Journal of Chromatography A* **2017**, *1527*, 18−32.

(23) Poole, C.; Karunasekara, T. Solvent classification for chromatography and extraction. *Journal of Planar Chromatography - Modern TLC* **2012**, *25* (3), 190−199.

(24) Kenney, D. H.; Paffenroth, R. C.; Timko, M. T.; Teixeira, A. R. Dimensionally reduced machine learning model for predicting single component octanol-water partition coefficients. *Journal of Cheminformatics* **2023**, *15* (1), 9.