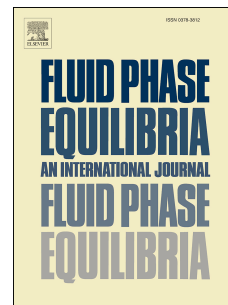


# Journal Pre-proof

An extended COSMO-SAC method for the prediction of carboxylic acid solubility

Sung Shin Kang, Jonghwi Lee, Jeong Won Kang



PII: S0378-3812(20)30219-3

DOI: <https://doi.org/10.1016/j.fluid.2020.112673>

Reference: FLUID 112673

To appear in: *Fluid Phase Equilibria*

Received Date: 24 February 2020

Revised Date: 26 May 2020

Accepted Date: 1 June 2020

Please cite this article as: S.S. Kang, J. Lee, J.W. Kang, An extended COSMO-SAC method for the prediction of carboxylic acid solubility, *Fluid Phase Equilibria* (2020), doi: <https://doi.org/10.1016/j.fluid.2020.112673>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. 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.

© 2020 Published by Elsevier B.V.

**Credit Author Statement**

**Sung Shin Kang** – Conceptualization, Methodology, Investigation, Writing-Original Draft

**Jongwhi Lee** – Review and Editing, Methodology

**Jeong Won Kang** – Supervision, Writing-Review & Editing,

**An extended COSMO-SAC method for the prediction of  
carboxylic acid solubility**

Sung Shin Kang<sup>a</sup>, Jonghwi Lee<sup>b</sup> and Jeong Won Kang<sup>a,\*</sup>

<sup>a</sup>*Department of Chemical and Biological Engineering, Korea University, 145 Anam-ro,  
Seongbuk-gu, Seoul 136-701, Republic of Korea*

<sup>b</sup>*Department of Chemical Engineering and Material Science, Chung-Ang University, Seoul  
156-756, Republic of Korea*

<sup>\*</sup>**Corresponding author.** Tel.: +82-2-3290-3305, fax: +82-2-926-6102, e-mail:

[jwkang@korea.ac.kr](mailto:jwkang@korea.ac.kr)

**Abstract**

The COSMO-SAC model provides a convenient method for predicting the phase behavior of components with minimal information about molecules using quantum-mechanical principles. During the past few years, several researchers tried to improve the prediction capability with modifications to hydrogen bonding terms and other methods of calculation. In this contribution, we focused on carboxylic acid groups, which may exhibit different hydrogen-bond behavior as compared with alcohols. The calculation scheme in COSMO-SAC (2017 version) was slightly modified to encompass a hydrogen-bond term specific to carboxylic acid groups. The parameters for the extended COSMO-SAC model were fitted using 559 data points. The characteristics of the extended model were investigated for the prediction capability of binary solubility, ternary solubility, and vapor-liquid equilibrium data. The results show that the extension correctly predicts the phase behavior of mixtures containing carboxylic acids.

**Keywords:** COSMO-SAC, Solubility prediction, Carboxylic acid, Solid-liquid equilibrium

## 1. Introduction

Carboxylic acids are widely used as raw materials or final products in the fine chemical and pharmaceutical industry [1-3]. The synthesis of carboxylic acids normally involves the formation of by-products, and purification steps are essential in many applications. In some cases, the processes require high purity products for the subsequent reactions or the quality of the final products [4, 5]. The crystallization process is an appropriate method for the production of high-purity products [6-8], in which reliable phase equilibrium information is essential.

State-of-the-art thermodynamic models, such as UNIFAC [9], NRTL-SAC [10], and PC-SAFT [11], are commonly used in the industry. Such models require parameters optimized on the basis of reliable experimental data. Compared with the vapor-liquid equilibrium (VLE) and the liquid-liquid equilibrium (LLE), experimental data and parameters on the solid-liquid equilibrium (SLE) received less attention, and theoretical studies are not frequently reported in the literature. It is rather common to use semi-empirical models for practical process calculations. However, empirical models are limited for binary cases and cannot be conveniently extended to multicomponent mixtures. Carboxylic acids contain both a hydrogen donor and acceptor in a single molecule, for which special considerations are required [12, 13]. Even for data-dependent models, minimal theoretical considerations are necessary [14].

Methods based on quantum mechanical calculations, such as COSMO-RS [15] and COSMO-SAC [16], have the advantage of predicting phase equilibria without experimental information. They only require structural information and correctly generated sigma-profile data. The COSMO-SAC model, initially developed by Lin and Sandler [16], has been

modified and improved over the years [17-22]. Most of the improvements were focused on the VLE and LLE by introducing special terms for hydrogen bonding [17, 18] and dispersive molecular interactions [19-21], while Shu and Lin [23] introduced a correction term for the prediction of the SLE. Carboxylic acids have strong interactions, which is often described by dimerization in the gas phase, and we expect that the specific consideration of such interactions may help to improve the activity prediction in SLE and VLE/LLE calculations.

In this study, we propose an extended version of COSMO-SAC (2017) for the solubility prediction of carboxylic acids. Additional parameters for the consideration of carboxylic acids were suggested, while preserving the original hydrogen bonding interaction parameters for alcohols. Experimental binary solubility data (559 points) were used for the optimization of the new parameters. The capability of the extended model was tested by comparison with UNIFAC and COSMO-SAC (2017) in the prediction of ternary solubility, binary solubility of complex components, and binary VLE calculations.

## 2. Theory

### 2.1 Solid-liquid equilibrium

The solid-liquid equilibrium requires the equality of fugacities of solute  $i$  in solid and liquid phases.

$$f_i^S = x_i \gamma_i f_i^L, \quad (1)$$

where  $f_i^S$  and  $f_i^L$  are the fugacities in the solid and liquid phase, respectively;  $x_i$  is the mole fraction, and  $\gamma_i$  is the activity coefficient in the liquid phase. From the equality of fugacities,

the solubility can be calculated [26]. The following equation was derived assuming that heat capacity differences are constant with respect to temperature.

$$\ln \frac{f_i^L}{f_i^S} = \ln x_i \gamma_i = \frac{\Delta H_{fus,i}}{RT_{fus,i}} \left( \frac{T_{f,i}}{T} - 1 \right) - \frac{\Delta Cp_i}{R} \left( \frac{T_{fus,i}}{T} - 1 \right) + \frac{\Delta Cp_i}{R} \ln \frac{T_{fus,i}}{T}, \quad (2)$$

where  $\Delta H_{fus,i}$  is the molar enthalpy of fusion (melting),  $T_{fus,i}$  is the fusion (melting) point temperature,  $T$  is the equilibrium temperature, and  $\Delta Cp_i$  is the difference between the molar heat capacities in the solid and liquid state. The three terms in the right-hand side of above equation is not of equal importance; the first term is the dominant one and the remaining two, of opposite sign, tend to cancel out each other in many cases. Furthermore, uncertainties in the heat of fusion data are quite large. We can conveniently simplify the form of solubility equation by reducing heat capacity term as follows;

$$\ln x_i = \frac{\Delta H_{fus,i}}{RT_{fus,i}} \left( 1 - \frac{T_{f,i}}{T} \right) - \ln \gamma_i. \quad (3)$$

Excluding the solute-specific properties, only the activity coefficient is the solvent and temperature-dependent variable. Thus, the non-ideal behavior of the mixture can be expressed by adjusting the activity coefficient.

## 2.2 COSMO-SAC model

The original COSMO-SAC model was developed by Lin and Sandler [16] and focused on the prediction of the VLE and LLE. Since then, various versions of COSMO-SAC-based methods have been developed. The models based on COSMO-SAC reported in 2007 [17] and 2010 [18] separated the molecular surface into two and three parts, respectively, to represent the non-ideal behavior of hydrogen bonds. Molecular dispersive interactions were considered

in COSMO-SAC (2013) [19], COSMO-SAC (dsp) [20], and COSMO-SAC (2017) [21]. Detailed descriptions of COSMO-SAC can be found elsewhere [16-21].

The COSMO-SAC model used in this study is based on the COSMO-SAC 2017 version. We introduced a probability function and empirical parameters for the self-association of carboxylic acids, while preserving other terms in the 2017 version.

The probability function in COSMO-SAC (2017) is expressed as

$$p_i^{hb}(\sigma) = p_i^{OH}(\sigma) + p_i^{OT}(\sigma), \quad (4)$$

where  $p$  is the sigma profile of component  $i$  at segment charge  $\sigma$ , and the superscripts  $hb$ ,  $OH$ , and  $OT$  are hydrogen bonding, a hydroxyl group, and other hydrogen bonding groups, respectively. We extended the probability function as follows:

$$p_i^{hb}(\sigma) = p_i^{OH}(\sigma) + p_i^{COOH}(\sigma) + p_i^{OT}(\sigma), \quad (5)$$

where the superscript  $COOH$  represents the term for the carboxylic acid group. A similar approach has been reported by Hsieh and Lin [27] for the consideration of amino groups.

The segment exchange energy calculation was also modified. The segment exchange energy is calculated as follows:

$$\Delta W(\sigma_n^s, \sigma_m^t) = C_{ES}(\sigma_m^t + \sigma_n^s)^2 - C_{hb}(\sigma_m^t - \sigma_n^s)^2, \quad (6)$$

where  $C_{ES}$  and  $C_{hb}$  are empirical parameters:  $C_{ES}$  is an electrostatic constant and  $C_{hb}$  is the hydrogen bonding interaction parameter.  $C_{hb}$  was also refined according to the modification of the probability function and is given by



$$\begin{aligned}
1 \quad C_{hb}(\sigma_m^t, \sigma_n^s) = & \quad C_{OH-OH} & \quad \text{if } s = t = OH \text{ and } \sigma_m^t, \sigma_n^s < 0 \\
& C_{OH-OT} & \quad \text{if } s = OH, t = OT \text{ and } \sigma_m^t, \sigma_n^s < 0 \\
& C_{OH-COOH} & \quad \text{if } s = OH, t = COOH \text{ and } \sigma_m^t, \sigma_n^s < 0 \\
& C_{COOH-COOH} & \quad \text{if } s = t = COOH \text{ and } \sigma_m^t, \sigma_n^s < 0 \\
& C_{COOH-OT} & \quad \text{if } s = COOH, t = OT \text{ and } \sigma_m^t, \sigma_n^s < 0 \\
& C_{OT-OT} & \quad \text{if } s = t = OT \text{ and } \sigma_m^t, \sigma_n^s < 0 \\
2 \quad & 0 & \quad \text{otherwise.} \tag{7}
\end{aligned}$$

3 Only three additional parameters ( $C_{OH-COOH}$ ,  $C_{COOH-COOH}$ , and  $C_{COOH-OT}$ ) are required in  
4 addition to the COSMO-SAC (2017) model.

### 5 3. Computational details

6 The procedure for calculating the activity coefficient is identical to that described in the  
7 COSMO-SAC (2017) model. The COSMO files stored in the VT database established by  
8 Mullins et al. [28, 29] were used. COSMO files not included in the VT database were  
9 generated by DFT/COSMO calculations using the DMol<sup>3</sup> software [30], following the  
10 procedure suggested by Mullins et al. [28, 29]. OH and COOH hydrogen bonds were  
11 identified by comparing the distance between two atoms and the sum of their atomic radii in a  
12 single molecule. If the distance between the two atoms in a single molecule was smaller than  
13 the sum of their atomic radii, they were regarded as bonded.

14 The enthalpy of fusion and fusion temperature data were obtained from the NIST webbook  
15 [31], and the values are listed in Table 1. Binary SLE data (559 points) were used for  
16 regression, and they are listed in Table 2. Benzoic, succinic, and adipic acids which have  
17 solid phase near ambient condition were chosen as solutes, and 20 common organic

components were selected as solvents. Water was excluded in selected solvents because of poor prediction of UNIFAC and COSMO based model for water containing system [21]. We expect that similar parameterization for hydrogen bond of water as separate segment will improve the overall result. The solute species can be further extended if more reliable data are collected. Within the collected data sets, some deviated significantly from others. Such data sets were marked as inconsistent data points.

**Table 1.** Data for enthalpy of fusion and fusion temperature used in this study and obtained from NIST webbook [31]

Solute	$\Delta H_f$ ( $kJ \cdot mol^{-1}$ )	$T_f$ (K)
Benzoic acid	17.99	395.5
Succinic acid	32.95	457.0
Adipic acid	34.85	426.4

**Table 2.** Experimental datasets used to regress the carboxylic acid interaction parameter of the extended COSMO-SAC model.

Solute	Solvent	Temperature range (K)	Data points	Number of references
Benzoic acid	<i>n</i> -Hexane	298.15–362.60	41	5
Benzoic acid	Cyclohexane	283.15–389.85	51	4
Benzoic acid	Benzene	277.50–383.20	25	6
Benzoic acid	Toluene	273.20–383.20	21	4
Benzoic acid	Acetone	277.59–322.72	14	4
Benzoic acid	<i>N,N</i> -Dimethylformamide	294.65–369.75	46	4
Benzoic acid	Methanol	298.15–301.20	4	4
Benzoic acid	Ethanol*	278.15–323.15	10	5
Benzoic acid	1-Butanol	298.15–363.20	10	4
Benzoic acid	1-Octanol*	295.30–322.15	15	3
Benzoic acid	Acetic acid	291.40–346.65	31	3
Adipic acid	Methanol	273.15–313.15	5	1
Adipic acid	Ethanol	273.15–340.95	29	4
Adipic acid	1-Propanol	273.15–333.15	6	1
Adipic acid	2-Propanol	273.15–333.15	6	1
Adipic acid	1-Butanol	273.15–458.15	40	3
Adipic acid	<i>t</i> -Butanol	273.15–333.15	4	1
Adipic acid	Acetone*	283.15–333.15	37	4
Adipic acid	1,4-Dioxane	273.15–333.15	5	1
Adipic acid	Acetic acid	288.15–347.65	26	4
Adipic acid	Cyclohexanone	292.75–355.15	12	1
Adipic acid	Cyclohexanol	298.80–352.47	12	1
Adipic acid	<i>N,N</i> -Dimethylformamide	290.8–335.25	10	1
Adipic acid	Dimethylacetamide	294.55–336.73	11	1
Adipic acid	Ethyl acetate	288.15–323.15	8	1
Succinic acid	Ethanol*	283.15–331.95	34	4
Succinic acid	1-Propanol	283.55–333.75	11	1
Succinic acid	Isopropanol	282.75–334.25	10	1
Succinic acid	Acetone	282.85–317.75	7	1
Succinic acid	Acetic acid	290.75–334.15	18	2

\*Mixtures with inconsistent data sources

The parameters were optimized using a simplex algorithm to minimize the following objective function:

$$obj = \frac{1}{m} \sum_i^m \frac{1}{n_i} (\sum_j^n (x_j^{calc} - x_j^{exp})^2)^{1/2}, \quad (8)$$

where  $m$  is the number of datasets,  $n_i$  is the number of data points in a dataset  $i$ , and the superscripts *calc* and *exp* indicate the calculated solubility and experimental data, respectively. The initial estimates of the parameters were set from the hydrogen bonding interaction parameters for the OH group. The regression results were analyzed by the least mean square error (*lmse*) and relative mean square error (*rmse*) between the predicted and experimental solubility.

$$lmse = \frac{1}{N} \sum_i^n (x_i^{calc} - x_i^{exp})^2, \quad (9)$$

$$rmse = \frac{1}{N} \sum_i^n \left( \frac{(x_i^{calc} - x_i^{exp})}{x_i^{calc}} \right)^2, \quad (10)$$

where  $N$  is the number of total data points.

## 4. Results and discussion

### 4.1 Summary of regression results

The results of the regression for the extended COSMO-SAC parameters are presented in Fig. 1 along with the comparison between the calculation results of the original COSMO-SAC, the original UNIFAC [9], and KT-UNIFAC [32]. The three panels of Fig. 1 (a, b, and c) show the calculation results for the solubility of adipic acid, benzoic acid, and succinic acid in organic solvents, depicting the correlation between the predicted and experimental solubility. If the points are closer to the diagonal crossing line, the model describes the solubility more

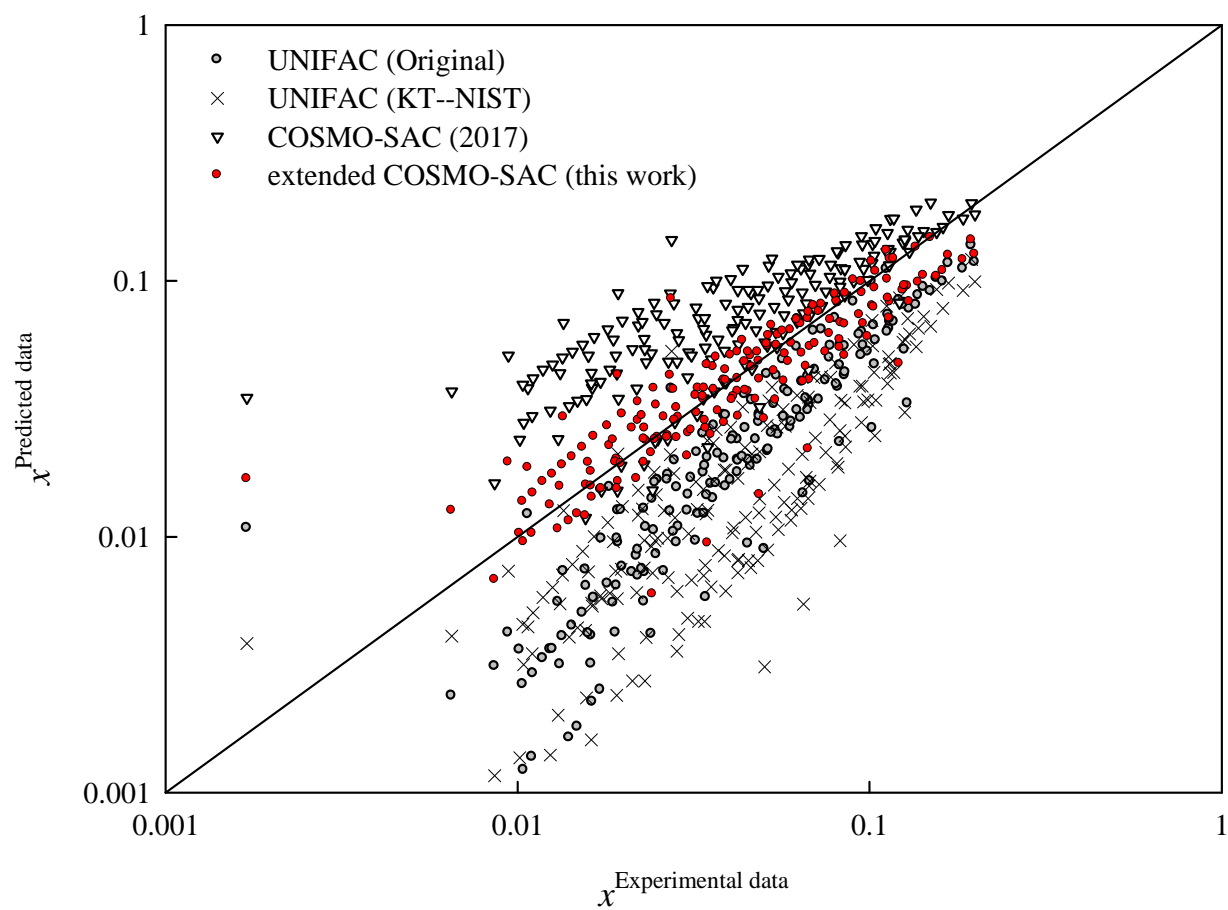
1 accurately. Due to the use of extended parameters, the extended model works better than the  
2 original one. Compared with the two versions of the UNIFAC model, the COSMO-SAC  
3 models generally work better. The UNIFAC models were originally fitted with VLE and LLE  
4 data, and it seems natural that they perform poorly for SLE prediction. The modified  
5 UNIFAC model suggested by Gmehling [33] maintains separate parameters for SLE. Such a  
6 model works better than the original one, but the parameters are not open to the public  
7 domain, which is why we cannot compare that model with ours.

8 Fig. 2 shows a typical prediction result for the solubility of succinic acid in ethanol for four  
9 models and experimental data from four different sources. As shown in the figure, the  
10 extended COSMO-SAC model better describes the temperature dependence of the solubility  
11 compared with the other models. Among the four datasets, the solubility data points from Hu  
12 et al. [36] do not agree with the data from the other references. Such disagreements were  
13 detected in five data sets (20 points) out of the total 559 data points. The inconsistent data  
14 sources are marked in Table 2, and they were excluded from the regression analysis.

15 The hydrogen bonding interaction parameters of the original COSMO-SAC (2017) and the  
16 extended COSMO-SAC are listed in Table 3. The *lmse* and *rmse* of the prediction models are  
17 summarized in Table 4.

1

2 (a)

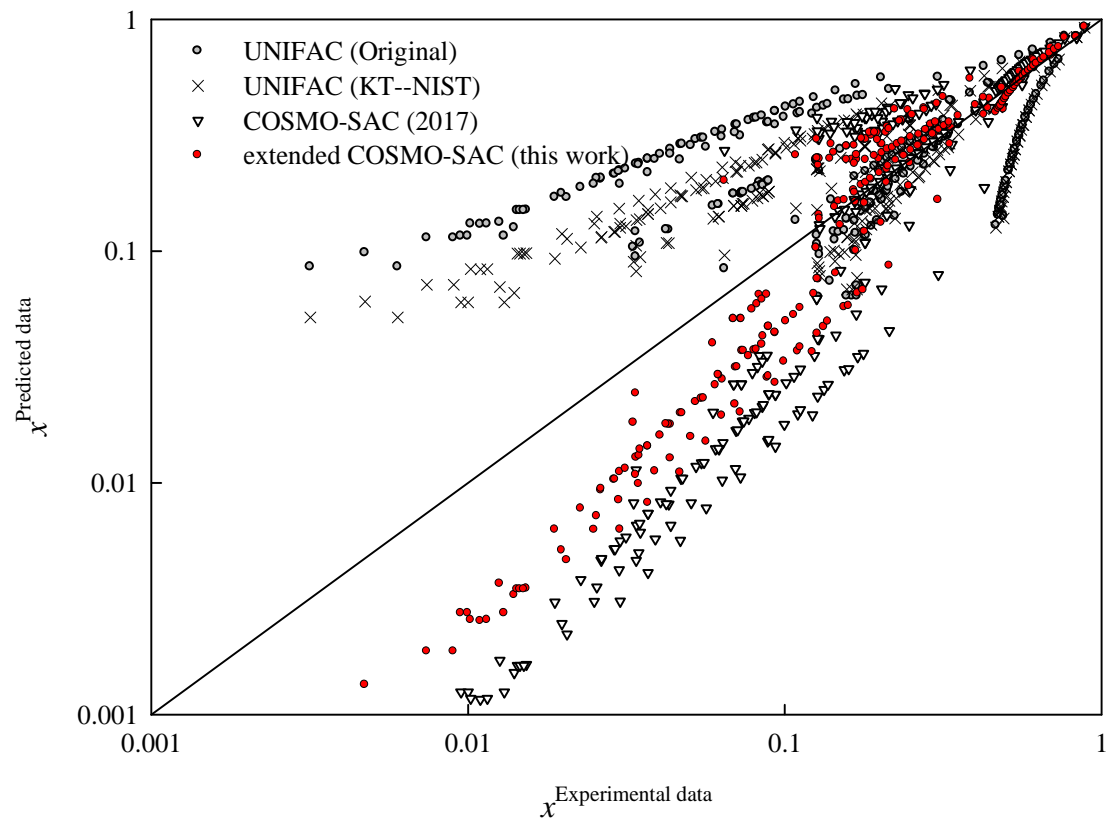


3

4

5

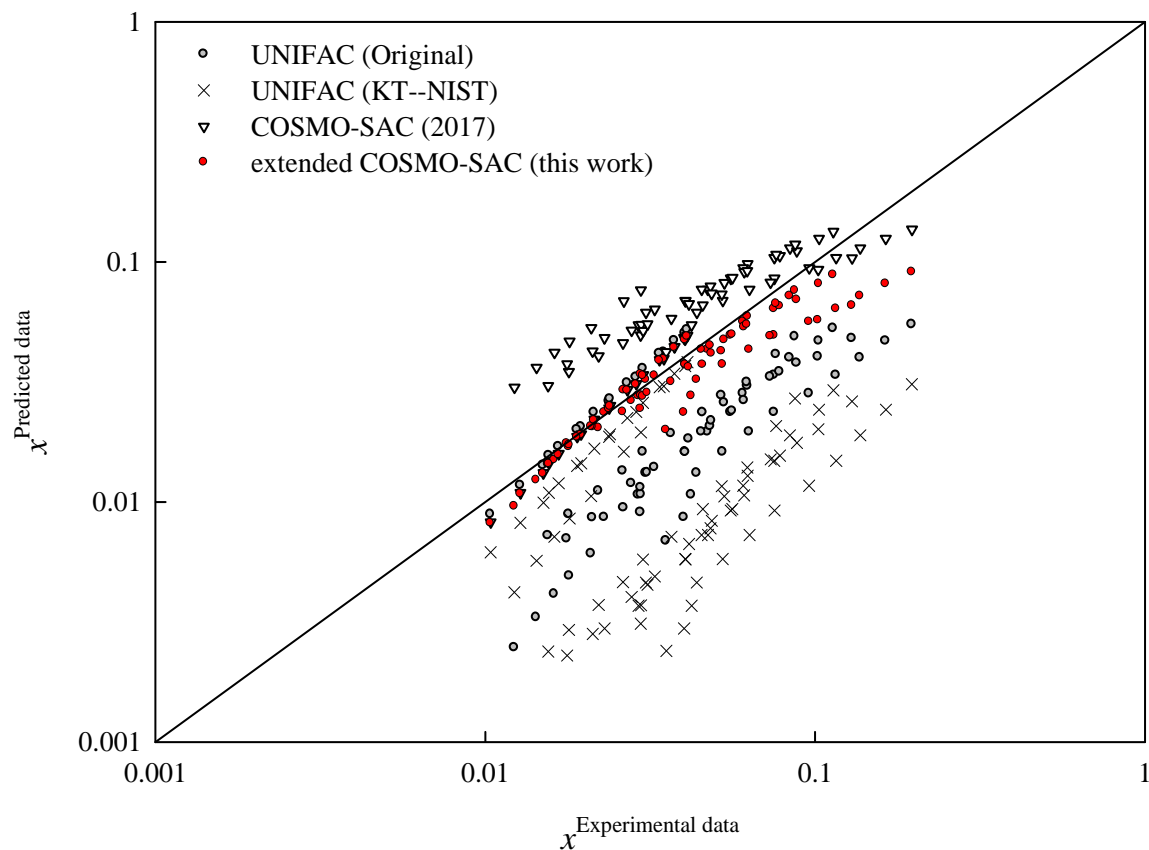
1 (b)



2

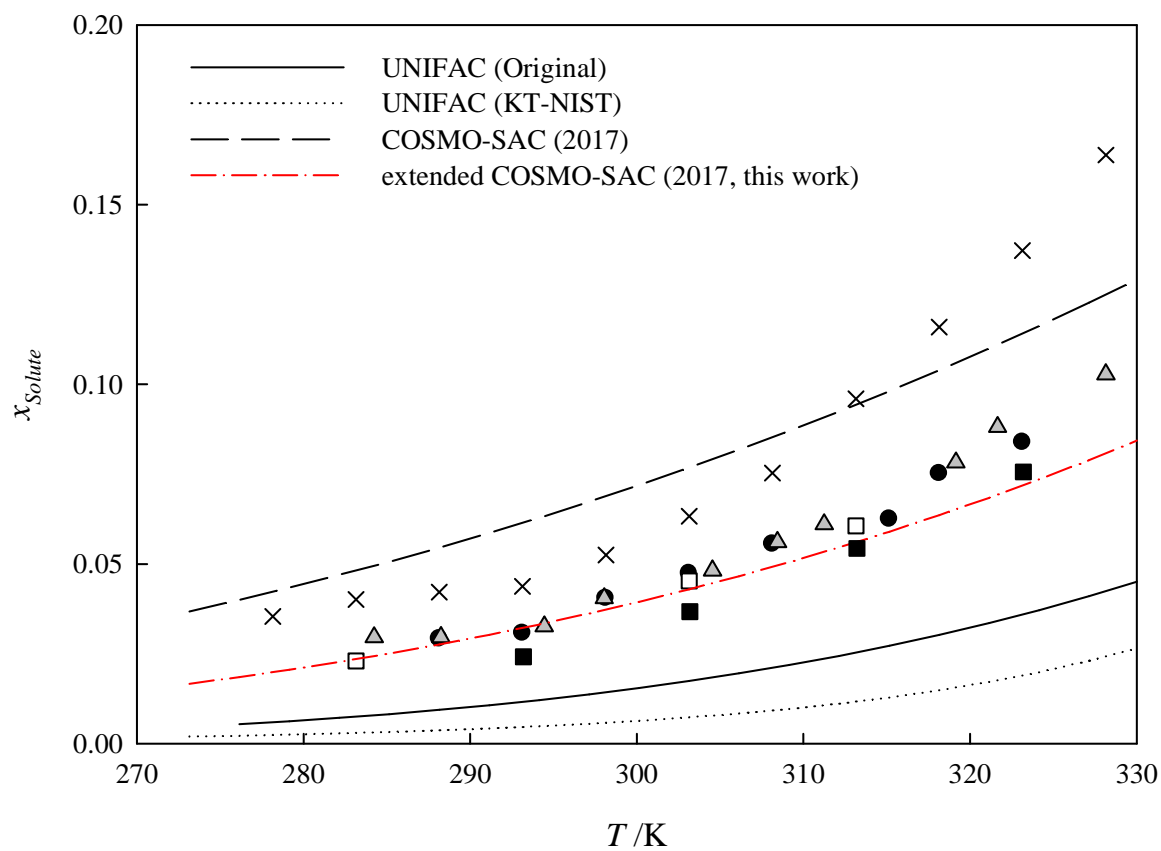
3

1 (c)

2  
3

4 **Fig. 1.** Comparison of experimental data with prediction results for the binary solid solubility  
 5 of (a) adipic acid, (b) benzoic acid, and (c) succinic acid; ■, UNIFAC (original); X, UNIFAC  
 6 (KT-NIST); ▽, COSMO-SAC (2017); ●, modified COSMO-SAC (this work).





**Fig. 2.** Comparison of experimental data with prediction results for the binary solid solubility of succinic acid in ethanol. The symbols are experimental data: ● [34]; □ [35]; × [36]; ▲ [37], ■ [38].

**Table 3.** Hydrogen bonding interaction parameters of the original COSMO-SAC (2017) and the extended COSMO-SAC model.

Parameter [ $J \cdot mol^{-1} \cdot nm^4 \cdot e^{-2}$ ]	COSMO-SAC (2017)	Extended COSMO-SAC (this work)
$C_{OH-OH}$	4013.78	4013.78
$C_{OH-OT}$	3016.43	3016.43
$C_{OT-ot}$	932.31	932.31
$C_{COOH-COOH}$		2225.67
$C_{COOH-OH}$		3020.18
$C_{COOH-ot}$		1872.84

**Table 4.** Comparison of lmse and rmse for SLE prediction from different prediction methods

Model	lmse	rmse ( $\times 10^3$ )
UNIFAC (original)	8.217	16.41
UNIFAC (KT-NIST)	2.955	11.15
Original COSMO-SAC (2017)	1.597	5.56
Extended COSMO-SAC (this work)	0.324	3.53

#### 4.2 Regression results for binary SLE

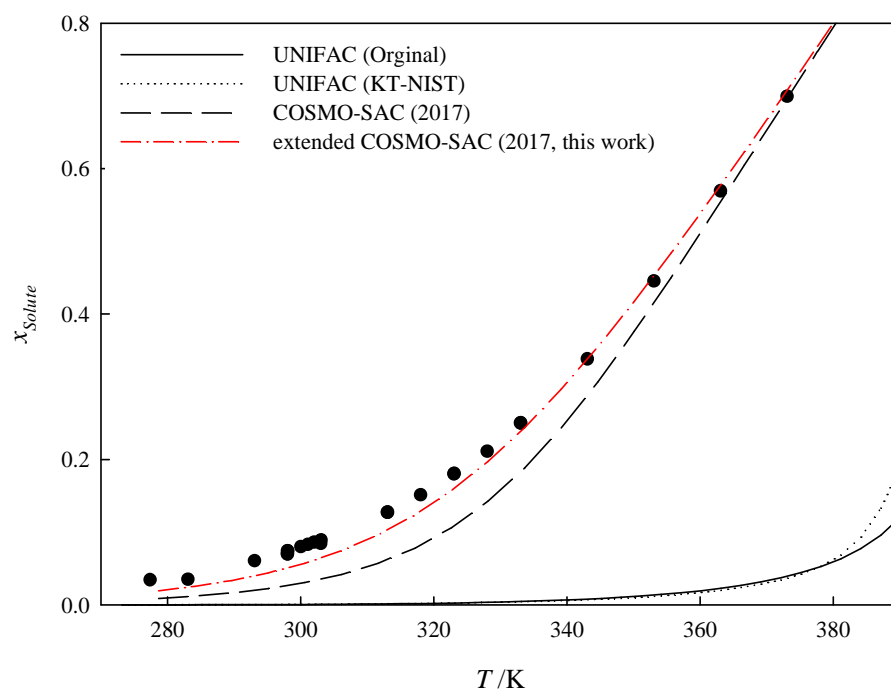
Fig. 3 shows typical binary solubility calculations for carboxylic acids in three types of solvents - (a) non-hydrogen bonding, (b) hydroxyl, (c) carboxylic, and (d) weakly polar solvent. Due to the introduction of the extended parameters, the extended COSMO-SAC shows the best results.

Benzoic acid in a benzene system is illustrated in Fig. 3 (a). The improvement with extended parameters is relatively small compared with the other cases. The only difference with the original COSMO-SAC model is the consideration of self-hydrogen bond effects among solutes. When acid components are dissolved in hydroxyl hydrogen-bonding solvents (adipic acid in ethanol, Fig. 3 (b)), the extended model works much better due to the consideration of a different hydrogen-bond effect. When carboxylic acids are dissolved in similar solvents (benzoic acid in acetic acid, Fig. 3(c)), the effect of hydrogen bonds canceled out, and the resulting solubility is almost the same as that in the original COSMO-SAC model. The results for the weakly polar solvent (succinic acid in acetone, Fig. 3 (d)), also show an improvement over the original model due to the introduction of the new parameters for the other hydrogen bonding group (OT) and the carboxylic acid group (COOH) interaction. The original UNIFAC and KT-UNIFAC calculations are also presented for comparison.

1

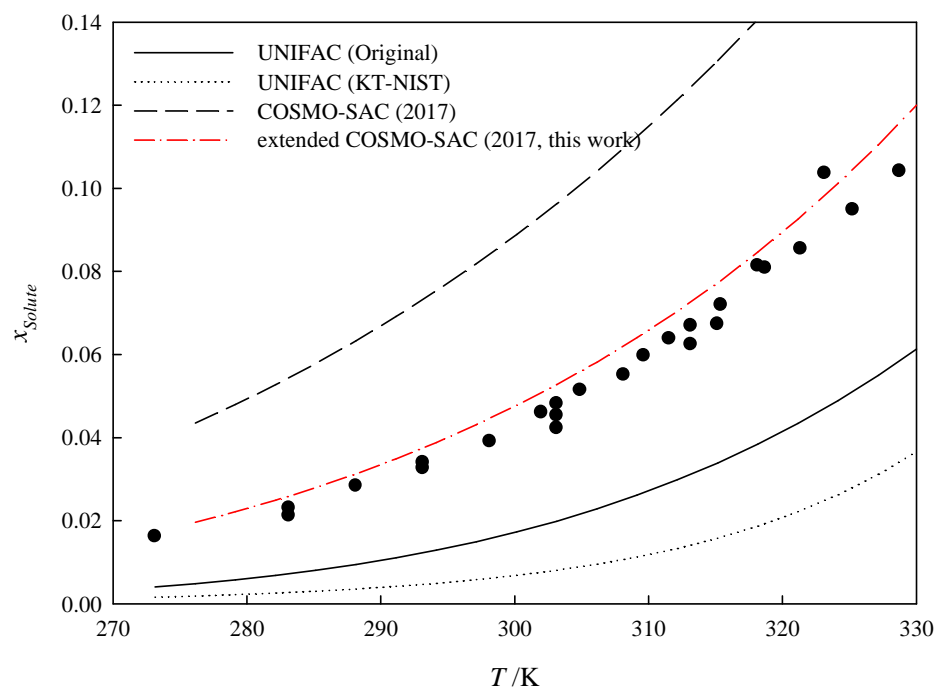
2

3 (a)



4

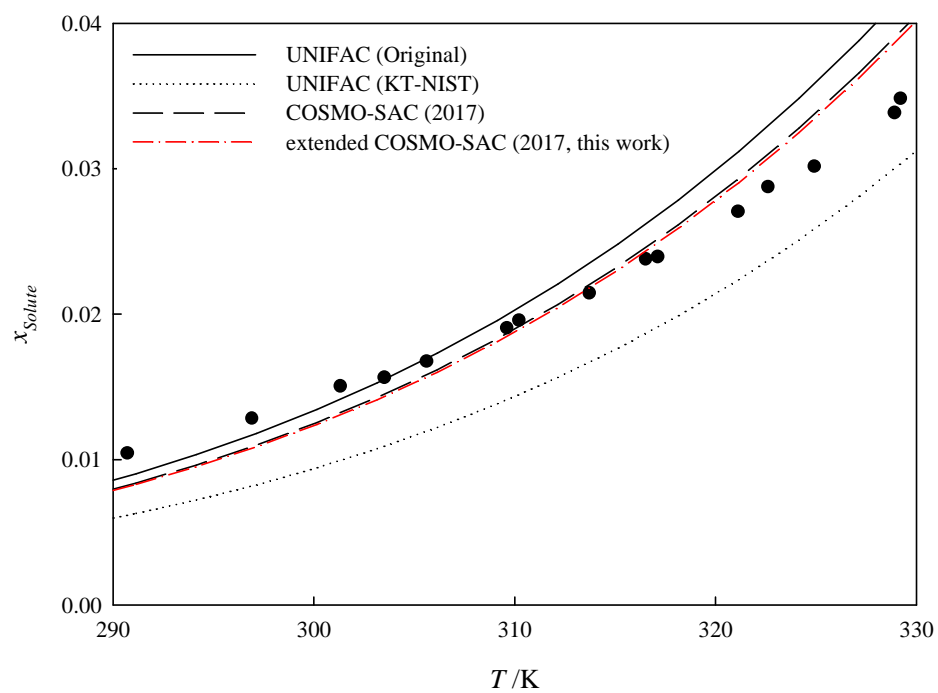
5 (b)



1

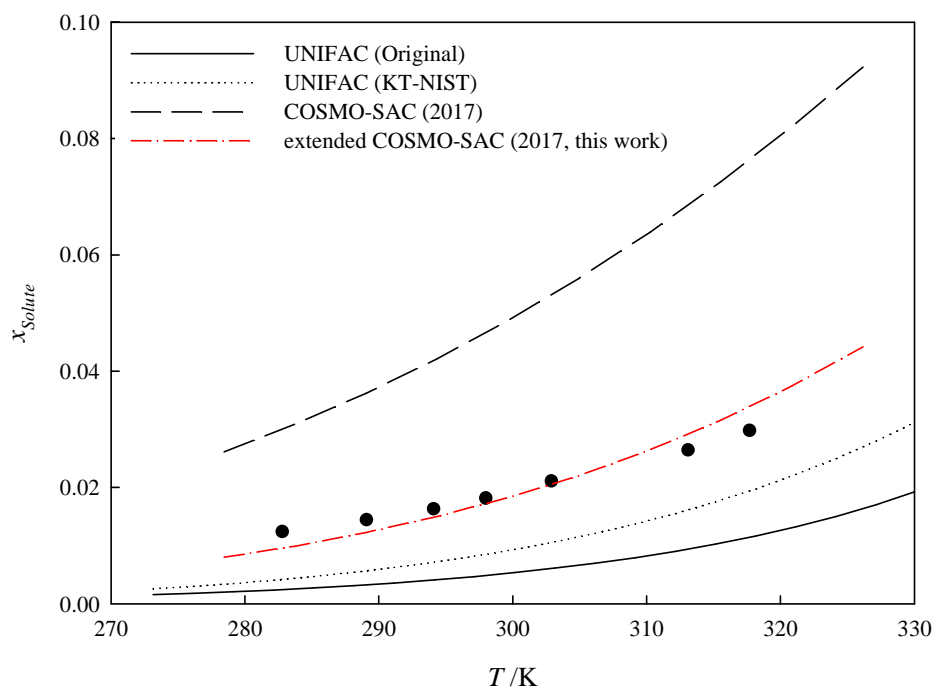
2

1 (c)



2

3 (d)



**Fig. 3.** Comparison of binary solubility prediction results with experimental data for mixtures containing carboxylic acid and (a) non-hydrogen bonding solvent (benzoic acid in benzene [39-42]), (b) hydroxyl group solvent (adipic acid in ethanol [2, 35, 43]), (c) carboxylic acid group solvent (benzoic acid in acetic acid [39, 44]), and (d) weakly polar solvent (succinic acid in acetone [37]). The black circles are experimental data.

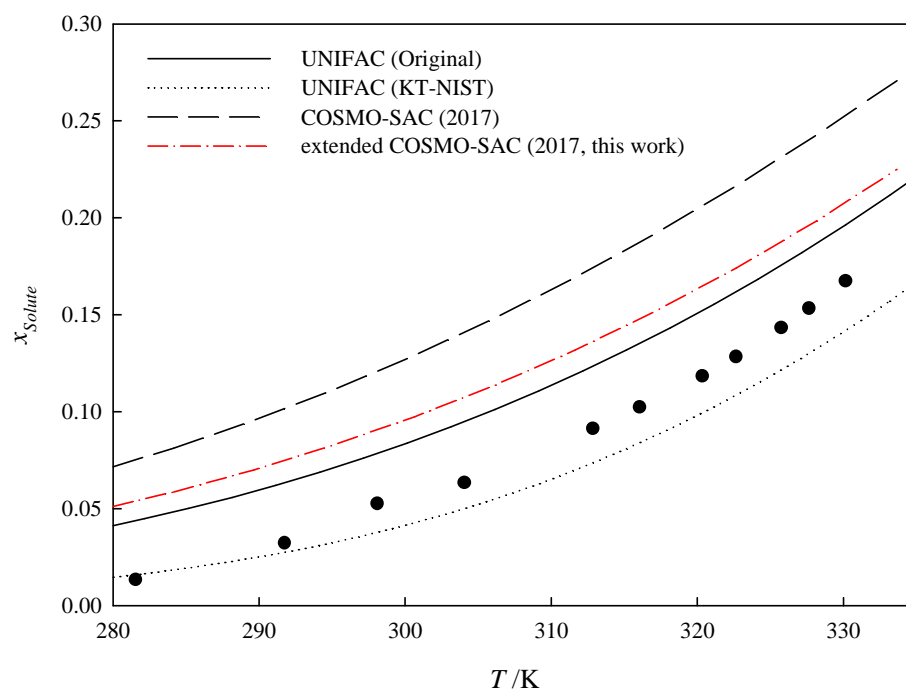
#### 4.3 Prediction results of binary SLE for complex molecules

Fig. 4 shows the predicted solubility of complex molecules (aspirin and ibuprofen) in alcohol. The solute species contain carboxylic acid groups but were not used in the parameter regression. The calculation framework of original COSMO-SAC do not consider the formation of intramolecular hydrogen bond within a given molecule. For such reason, the calculation result for components with more than two H-bond forming functional groups is relatively poor compared with result for components with a single group. In this case, the UNIFAC-based models work better than the COSMO-based models. Interestingly, the original UNIFAC model shows the best prediction in this case. When we restrict our attention to the two COSMO-based models, the extended COSMO-SAC model best predicts the experimental data. We expect that the proposed parameters can be further improved consideration of intramolecular hydrogen bonding and regression with extended data sets for pharmaceutical components.



1

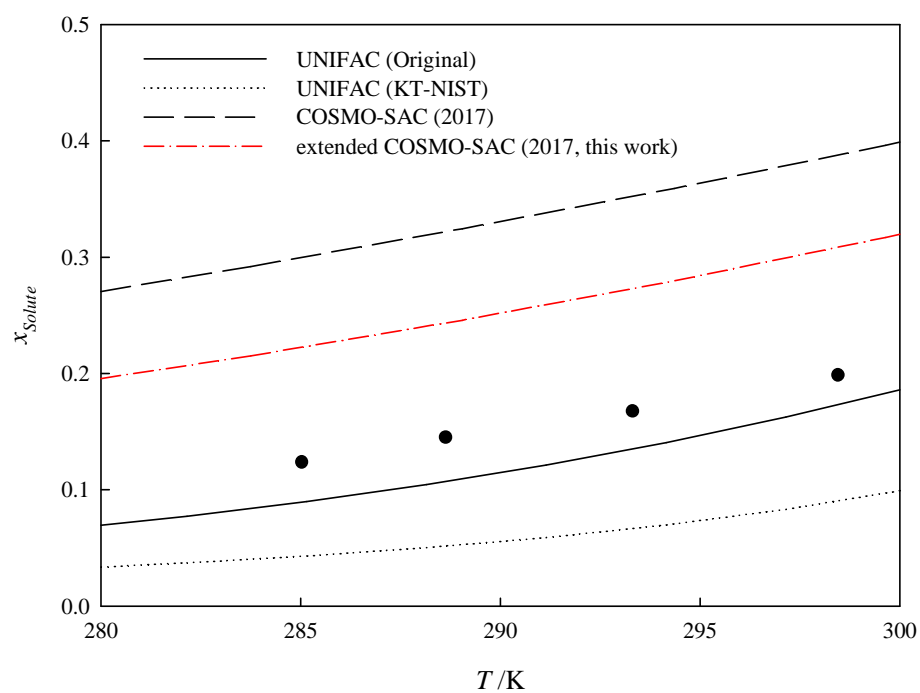
2 (a)



3

4 (b)

5

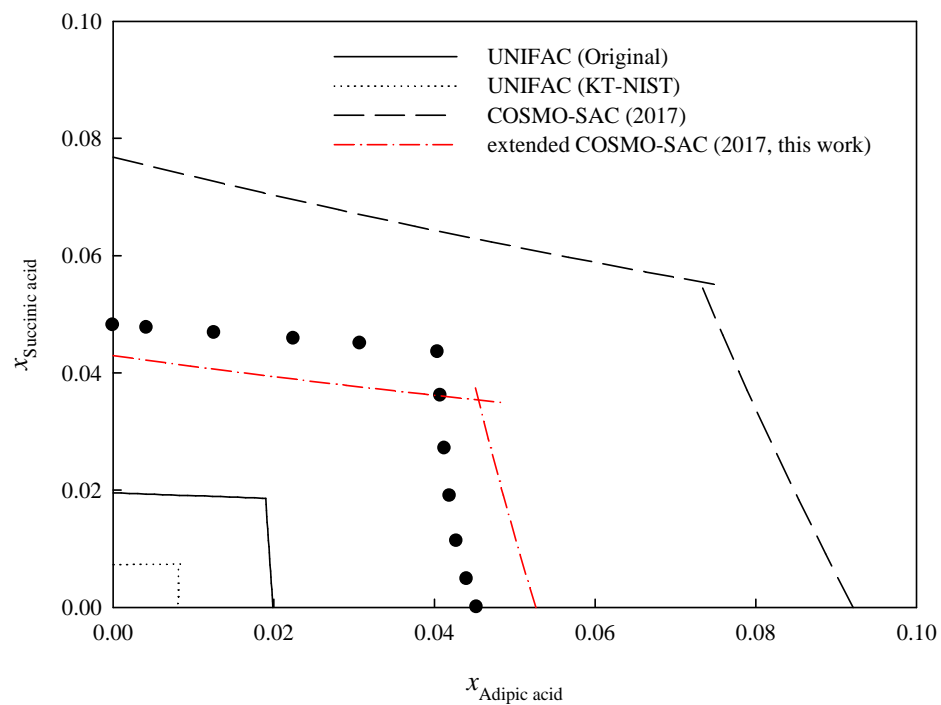


**Fig. 4.** Comparison of the binary solubility prediction results with experimental data of (a) aspirin/2-propanol [45] and (b) ibuprofen/ethanol [46]. The black circles are experimental data.

#### 4.4 Prediction of ternary SLE

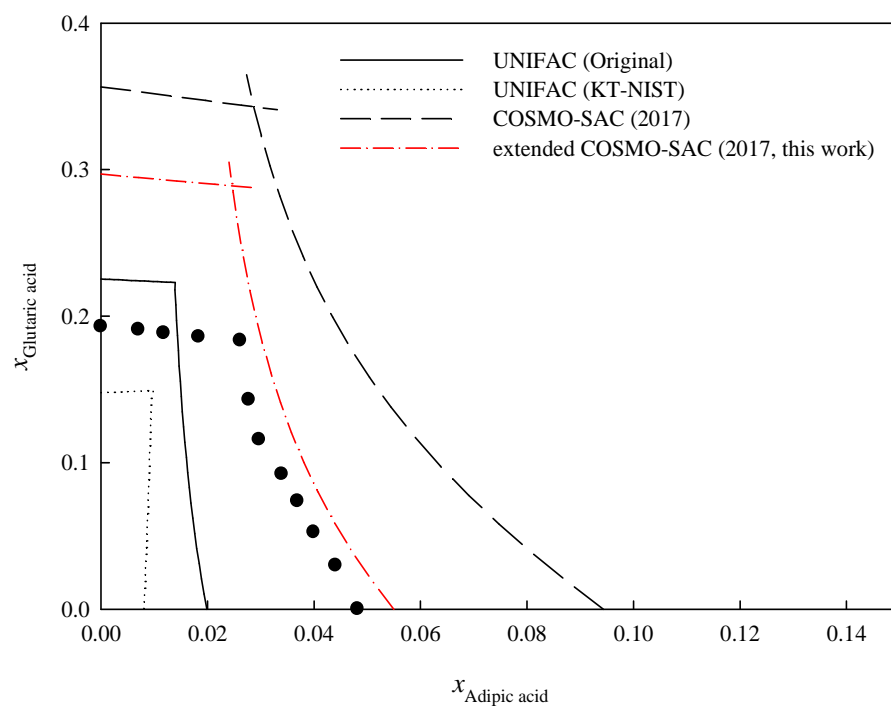
Ternary SLE prediction was performed to evaluate the capability for extension of the proposed model. Fig. 5 shows the ternary solubility of the multi-solute systems (a) adipic acid/succinic acid in ethanol and (b) adipic acid/glutaric acid in ethanol. For the adipic acid/succinic acid in ethanol system, the extended COSMO-SAC model shows the best representation of the ternary phase boundaries. In the panel for the adipic acid + glutaric acid in ethanol system (Fig. 5 (b)), the binary solubility of glutaric acid is not accurate for the models, except for the original UNIFAC model. However, the extended COSMO-SAC model shows an appropriate representation of the ternary phase boundary, except for the region close to the binary limit of glutaric acid.

1 (a)



2

3 (b)

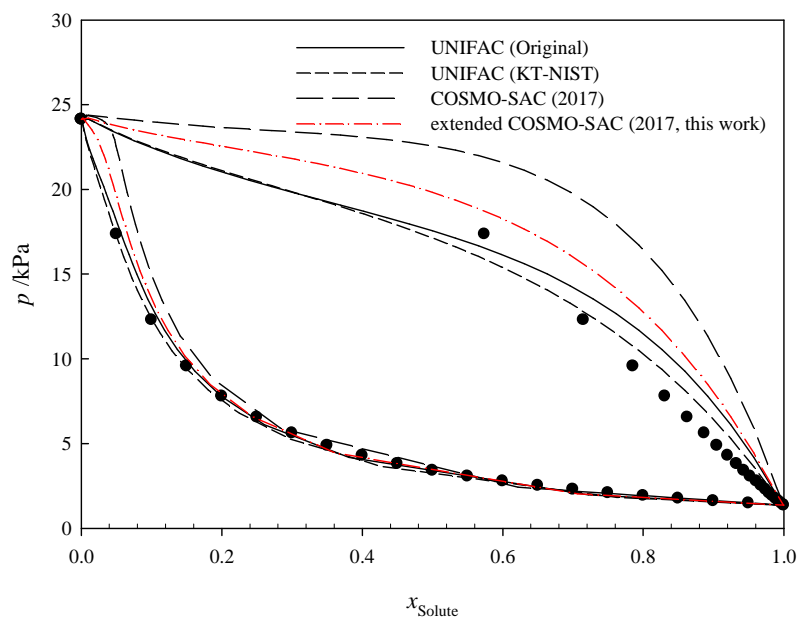


**Fig. 5.** Comparison of the ternary solubility prediction results with experimental data: (a) adipic acid/succinic acid/ethanol [35] and (b) adipic acid/glutaric acid/ethanol [43] at 303.15 K. The black circles are experimental data.

#### 4.5 Prediction of binary VLE

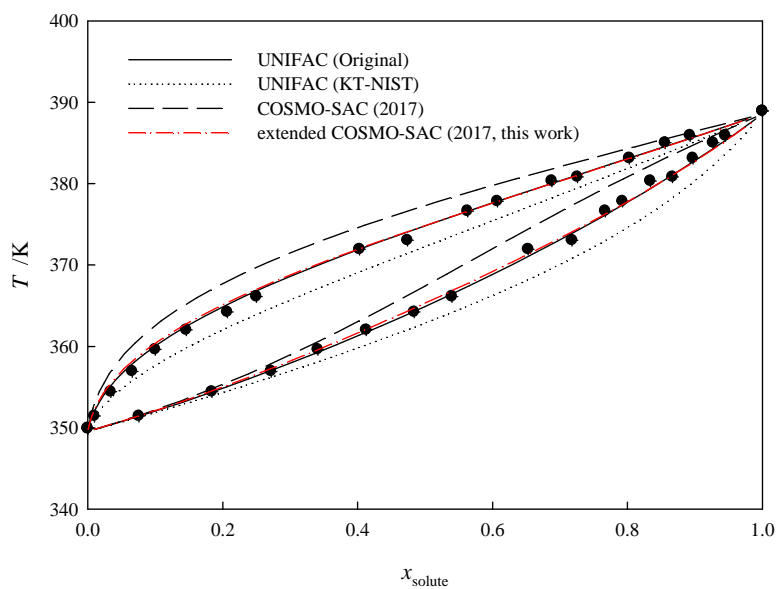
The purpose of COSMO-SAC model is to predict phase equilibrium with minimal structural information of constituent molecules. In principle, the COSMO-SAC model is a predictive model independent of condition and phase of the mixture, and do not require additional parameters. However, we introduced a small set of additional parameters to improve SLE calculation of acidic molecules. In this section effect of introduction of parameters were investigated by extension of parameters to VLE calculations. The results for the prediction of the VLE phase behavior are shown in Fig. 6. We compared four kinds of mixtures – acid + nonpolar, acid + alcohol, acid + acid, and acid + weakly polar components. As shown in Fig. 6, the extended COSMO-SAC model works generally better than the original COSMO-SAC model. Compared with the UNIFAC models, which are optimized with experimental VLE and LLE data sets, the extended COSMO-SAC model shows comparable prediction results. By comparing the experimental VLE with predicted values, we can conclude that the suggested modification reflect the molecular interactions in appropriate manner.

1 (a)



2

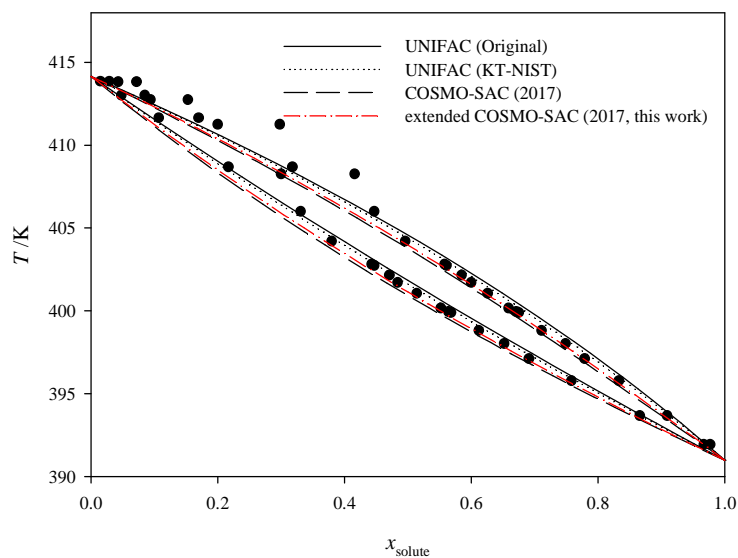
3 (b)



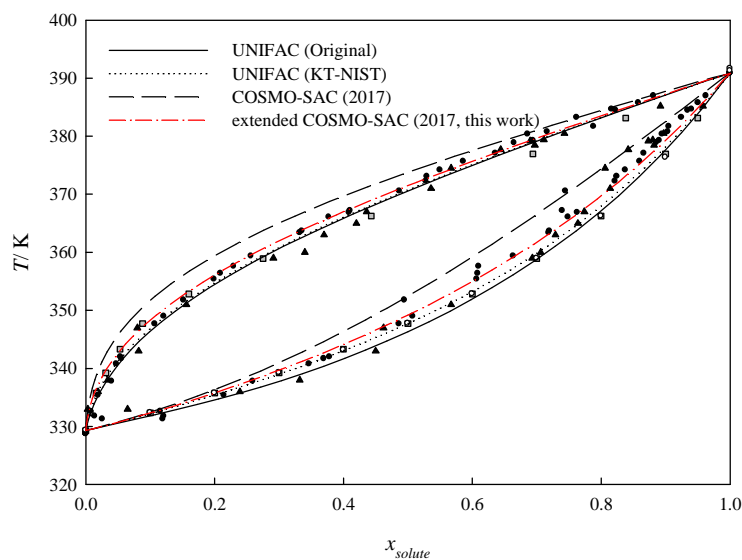
4

5

(c)



(d)



**Fig. 6.** Binary VLE prediction results compared with the experimental data of (a) propionic acid and benzene at 313.14 K [47], (b) acetic acid and ethanol at 93.99 kPa [48], (c) propionic acid and acetic acid at 101.325 kPa [49], and (d) acetic acid in acetone at 101.325 kPa [50-53]. The symbols represent experimental data.



## 5. Conclusions

In this study, we suggested an extended COSMO-SAC model with hydrogen bonding interactions of carboxylic acids. We divided the probability function into four parts and introduced three hydrogen interaction parameters for carboxylic groups. The specific interaction parameters were obtained from the regression of binary SLE data (559 points) in 80 data sets. The extended COSMO-SAC model showed a better representation of SLE behavior compared with the original COSMO-SAC (2017) and UNIFAC-based methods. The proposed model also showed better prediction results as compared to the original COSMO-SAC model when applied to complex mixtures, ternary behavior, and VLE predictions. The proposed scheme can be further refined with the aid of extensive solubility databases.

## Acknowledgements

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2018R1A2B6007002, NRF-2014R1A5A1009799, and NRF-2019M3E6A1064876).

## Declaration of interest

The authors declare that there are no conflicts of interests.

## Nomenclature

$C_{ES}$	electrostatic constant [ $\text{kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^4\cdot\text{e}^{-2}$ ]
$C_{hb}$	hydrogen bonding interaction parameter [ $\text{kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^4\cdot\text{e}^{-2}$ ]
$\Delta C p_i$	difference in solid and liquid heat capacities of component $i$ [ $\text{kJ}\cdot\text{mol}^{-1}\cdot\text{K}^{-2}$ ]
$f_i^L$	fugacity of component $i$ in liquid phase [-]
$f_i^S$	fugacity of component $i$ in solid phase
$\Delta H_f$	enthalpy of fusion [ $\text{kJ mol}^{-1}$ ]
$lmse$	least mean square error [-]
$m$	number of datasets [-]
$N$	number of total data points [-]
$n_i$	number of data points in dataset $i$ [-]
$p_i(\sigma)$	sigma profile of component $i$ at segment charge $\sigma$ [-]
$rmse$	relatively mean square error [-]

- 1  $T$  equilibrium temperature [K]
- 2  $T_f$  fusion temperature [K]
- 3  $\Delta W$  segment exchange energy [kcal·mol<sup>-1</sup>]
- 4  $x_i$  mole fraction of component  $i$  in liquid phase [-]
- 5  $\gamma_i$  activity coefficient of component  $i$  in liquid phase [-]

6

7

## 1    **References**

- 2    [1] A. Castellan, J.C.J. Bart, S. Cavallaro, Industrial production and use of adipic acid, *Catal.*  
3    *Today* 9 (1991) 237-254. [https://doi.org/10.1016/0920-5861\(91\)80049-F](https://doi.org/10.1016/0920-5861(91)80049-F).
- 4    [2] L. Fan, P. Ma, Z. Xiang, Measurement and correlation for solubility of adipic acid in  
5    several solvents, *Chin. J. Chem. Eng.* 15 (2007) 110-114. [https://doi.org/10.1016/S1004-](https://doi.org/10.1016/S1004-9541(07)60042-1)  
6    9541(07)60042-1.
- 7    [3] D.P. McNamara, S.L. Childs, J. Giordano, A. Iarriccio, J. Cassidy, M.S. Shet, R.  
8    Mannion, E. O'Donnell, A. Park, Use of a glutaric acid cocrystal to improve oral  
9    bioavailability of a low solubility API, *Pharm. Res.* 23 (2006) 1888-1897.  
10    <http://doi.org/10.1007/s11095-006-9032-3>.
- 11    [4] Y. Sun, L. Yan, H. Fu, Z. Xiu, Salting-out extraction and crystallization of succinic acid  
12    from fermentation broths, *Process Biochem.* 49 (2014) 506-511.  
13    <https://doi.org/10.1016/j.procbio.2013.12.016>.
- 14    [5] I. Bechthold, K. Bretz, S. Kabasci, R. Kopitzky, A. Springer, Succinic acid: A new  
15    platform chemical for biobased polymers from renewable resources, *Chem. Eng. Technol.* 31  
16    (2008) 647-654. <https://doi.org/10.1002/ceat.200800063>.
- 17    [6] E.H. Kim, S.-S. You, J.W. Kang, Effective separation of succinic acid by combined  
18    crystallization, *Korean J. Chem. Eng.* 35 (2018) 204-209. [https://doi.org/10.1007/s11814-](https://doi.org/10.1007/s11814-017-0232-9)  
19    017-0232-9.
- 20    [7] B. Shen, Q. Wang, Y. Wang, X. Ye, F. Lei, X. Gong, Solubilities of Adipic Acid in  
21    Acetic Acid + Water Mixtures and Acetic Acid + Cyclohexane Mixtures, *J. Chem. Eng. Data.*  
22    58 (2013) 938-942. <https://doi.org/10.1021/je301202v>.
- 23    [8] X. Sheng, W. Luo, Q. Wang, Determination and Correlation for the Solubilities of  
24    Succinic Acid in Cyclohexanol + Cyclohexanone + Cyclohexane Solvent Mixtures, *J. Chem.*  
25    *Eng. Data.* 63 (2018) 801-811. <https://doi.org/10.1021/acs.jced.7b00956>.
- 26    [9] A. Fredenslund, R.L. Jones, J.M. Prausnitz, Group-contribution estimation of activity  
27    coefficients in nonideal liquid mixtures, *AIChE J.* 21 (1975) 1086-1099.  
28    <https://doi.org/10.1002/aic.690210607>.
- 29    [10] C.-C. Chen, P.A. Crafts, Correlation and prediction of drug molecule solubility in mixed  
30    solvent systems with the nonrandom two-liquid segment activity coefficient (NRTL-SAC)  
31    model, *Ind. Eng. Chem. Res.* 45 (2006) 4816-4824. <https://doi.org/10.1021/ie051326p>.
- 32    [11] J. Gross, G. Sadowski, Perturbed-chain SAFT: An equation of state based on a  
33    perturbation theory for chain molecules, *Ind. Eng. Chem. Res.* 40 (2001) 1244-1260.  
34    <https://doi.org/10.1021/ie0003887>.
- 35    [12] O. Ivasenko, D.F. Perepichka, Mastering fundamentals of supramolecular design with  
36    carboxylic acids. Common lessons from X-ray crystallography and scanning tunneling  
37    microscopy, *Chem. Soc. Rev.* 40 (2011) 191-206. <https://doi.org/10.1039/C0CS00022A>.
- 38    [13] A. Vahid, J.R. Elliott, Transferable intermolecular potentials for carboxylic acids and  
39    their phase behavior, *AIChE J.* 56 (2010) 485-505. <https://doi.org/10.1002/aic.11966>.
- 40    [14] C. Yushu, A. Afef, M. Fabrice, S. Roland, M.R. Jeday, Thermodynamic modeling of  
41    mixtures containing carboxylic acids using the PC-SAFT equation of state, *Ind. Eng. Chem.*  
42    *Res.* 51 (2012) 13846-13852. <https://doi.org/10.1021/ie301930q>.
- 43    [15] A. Klamt, G. Schüürmann, COSMO: a new approach to dielectric screening in solvents  
44    with explicit expressions for the screening energy and its gradient, *J. Chem. Soc., Perkin*  
45    *Trans.* 2 (1993) 799-805. <https://doi.org/10.1039/P29930000799>.

- [16] S.-T. Lin, S.I. Sandler, A priori phase equilibrium prediction from a segment contribution solvation model, *Ind. Eng. Chem. Res.* 41 (2002) 899-913. <https://doi.org/10.1021/ie001047w>.
- [17] S. Wang, S.I. Sandler, C.-C. Chen, Refinement of COSMO-SAC and the applications, *Ind. Eng. Chem. Res.* 46 (2007) 7275-7288. <https://doi.org/10.1021/ie070465z>.
- [18] C.-M. Hsieh, S.I. Sandler, S.-T. Lin, Improvements of COSMO-SAC for vapor-liquid and liquid-liquid equilibrium predictions, *Fluid Phase Equilib.* 297 (2010) 90-97. <https://doi.org/10.1016/j.fluid.2010.06.011>.
- [19] R. Xiong, S.I. Sandler, R.I. Burnett, An improvement to COSMO-SAC for predicting thermodynamic properties, *Ind. Eng. Chem. Res.* 53 (2014) 8265-8278. <https://doi.org/10.1021/ie404410v>.
- [20] C.-M. Hsieh, S.-T. Lin, J. Vrabec, Considering the dispersive interactions in the COSMO-SAC model for more accurate predictions of fluid phase behavior, *Fluid Phase Equilib.* 367 (2014) 109-116. <https://doi.org/10.1016/j.fluid.2014.01.032>.
- [21] R. Fingerhut, W.-L. Chen, A. Schedemann, W. Cordes, J. Rarey, C.M. Hsieh, J. Vrabec, S.T. Lin, Comprehensive assessment of COSMO-SAC models for predictions of fluid-phase equilibria, *Ind. Eng. Chem. Res.* 56 (2017) 9868-9884. <https://doi.org/10.1021/acs.iecr.7b01360>.
- [22] B. Bouillot, S. Teychené, B. Biscans, An evaluation of COSMO-SAC model and its evolutions for the prediction of drug-like molecule solubility: Part 1, *Ind. Eng. Chem. Res.* 52 (2013) 9276-9284. <https://doi.org/10.1021/ie3015318>.
- [23] C.-C. Shu, S.-T. Lin, Prediction of drug solubility in mixed solvent systems using the COSMO-SAC activity coefficient model, *Ind. Eng. Chem. Res.* 50 (2011) 142-147. <https://doi.org/10.1021/ie100409y>.
- [24] B. Bouillot, S. Teychené, B. Biscans, Discussion and improvement of the refined COSMO-SAC parameters for solubility predictions: Part 2, *Ind. Eng. Chem. Res.* 52 (2013) 9285-9294. <https://doi.org/10.1021/ie301905k>.
- [25] E. Paulechka, V. Diky, A. Kazakov, K. Kroenlein, M. Frenkel, Reparameterization of COSMO-SAC for phase equilibrium properties based on critically evaluated data, *J. Chem. Eng. Data* 60 (2015) 3554-3561. <https://doi.org/10.1021/acs.jced.5b00483>.
- [26] J.M. Prausnitz, R.N. Lichtenthaler, E.G. de Azevedo, *Molecular thermodynamics of fluid-phase equilibria*, Pearson Education, 1998.
- [27] C.-M. Hsieh, S.-T. Lin, Prediction of liquid-liquid equilibrium from the Peng-Robinson+COSMO-SAC equation of state, *Chem. Eng. Sci.* 65 (2010) 1955-1963. <https://doi.org/10.1016/j.ces.2009.11.036>.
- [28] E. Mullins, R. Oldland, Y.A. Liu, S. Wang, S.I. Sandler, C.-C. Chen, M. Zwolak, K.C. Seavey, Sigma-profile database for using COSMO-based thermodynamic methods, *Ind. Eng. Chem. Res.* 45 (2006) 4389-4415. <https://doi.org/10.1021/ie060370h>.
- [29] E. Mullins, Y.A. Liu, A. Ghaderi, S.D. Fast, Sigma profile database for predicting solid solubility in pure and mixed solvent mixtures for organic pharmacological compounds with COSMO-based thermodynamic methods, *Ind. Eng. Chem. Res.* 47 (2008) 1707-1725. <https://doi.org/10.1021/ie0711022>.
- [30] A.D. Studio, 1.7, Accelrys Software Inc., San Diego, CA, USA. 2006, in.
- [31] W. Acree Jr., J.S. Chickos, Phase transition enthalpy measurements of organic and organometallic compounds, NIST Chemistry WebBook, NIST Standard Reference Database Number 69, 2016. <https://doi.org/10.18434/T4D303>.

- [32] J.W. Kang, V. Diky, R.D. Chirico, J.W. Magee, C.D. Muzny, I. Abdulagatov, A.F. Kazakov, M. Frenkel, A new method for evaluation of UNIFAC interaction parameters, *Fluid Phase Equilib.* 309 (2011) 68-75. <https://doi.org/10.1016/j.fluid.2011.07.001>.
- [33] Dortmund Data Bank Software & Separation Technology, The UNIFAC Consortium. [http://unifac.ddbst.de/unifac\\_.html](http://unifac.ddbst.de/unifac_.html), 2020 (accessed 21 February 2020).
- [34] H. Zhang, Q. Yin, Z. Liu, J. Gong, Y. Bao, M. Zhang, H. Hao, B. Hou, C. Xie, An odd-even effect on solubility of dicarboxylic acids in organic solvents, *J. Chem. Thermodyn.* 77 (2014) 91-97. <https://doi.org/10.1016/j.jct.2014.05.009>.
- [35] R. Li, L. Meng, S. Han, C. Du, J. Xu, R. Xu, J. Wang, H. Zhao, Determination and correlation of solid-liquid phase equilibrium and phase diagram for a multicomponent system of mixed dibasic acids. Ternary system of succinic acid + adipic acid + ethanol, *J. Chem. Eng. Data* 61 (2016) 2105-2113. <https://doi.org/10.1021/acs.jced.6b00031>.
- [36] Y.-H. Hu, Z.-G. Chen, W.-G. Yang, Y. Shi, H.-L. Sun, Y.-L. Li, Solubility of succinic acid in ethanol plus water systems from 278.15 K to 333.15 K, *J. Solution Chem.* 42 (2013) 102-110. <http://doi.org/10.1007/s10953-012-9951-z>.
- [37] Q. Yu, S. Black, H. Wei, Solubility of butanedioic acid in different solvents at temperatures between 283 K and 333 K, *J. Chem. Eng. Data* 54 (2009) 2123-2125. <https://doi.org/10.1021/je900021g>.
- [38] M.L. Oliveira, M.R. Franco Jr, Solubility of 1, 4-butanedioic acid in aqueous solutions of ethanol or 1-propanol, *Fluid phase equilibria*, 326 (2012) 50-53.
- [39] A. Beerbower, P.L. Wu, A. Martin, Expanded solubility parameter approach I: Naphthalene and benzoic acid in individual solvents, *J. Pharm. Sci.* 73 (1984) 179-188. <https://doi.org/10.1002/jps.2600730210>.
- [40] C.K. Hancock, J.N. Pawloski, J.P. Idoux, Quantitative solubility—structure relationships for some meta- and para-substituted benzoic acids in benzene and in cyclohexane, *J. Org. Chem.* 31 (1966) 3801-3804. <https://doi.org/10.1021/jo01349a073>.
- [41] J. Chipman, The solubility of benzoic acid in benzene and in toluene, *J. Am. Chem. Soc.* 46 (1924) 2445-2448. <https://doi.org/10.1021/ja01676a012>.
- [42] H. Buchowski, Excess functions of mixing along the liquidus and the effect of aprotic solvents upon the solubility of solid non-electrolytes, *J. Solution Chem.* 20 (1991) 139-151. <http://doi.org/10.1007/BF00651646>.
- [43] L. Meng, J. Xu, C.B. Du, S. Han, R.J. Xu, J. Wang, H.K. Zhao, Determination and correlation of solid-liquid phase equilibrium and phase diagram for multicomponent system of mixed dibasic acids. (I) Ternary system of adipic acid + glutaric acid + ethanol system, *Fluid Phase Equilib.* 409 (2016) 256-263. <https://doi.org/10.1016/j.fluid.2015.09.036>.
- [44] B. Long, J. Li, R. Zhang, L. Wan, Solubility of benzoic acid in acetone, 2-propanol, acetic acid and cyclohexane: Experimental measurement and thermodynamic modeling, *Fluid Phase Equilib.* 297 (2010) 113-120. <https://doi.org/10.1016/j.fluid.2010.06.021>.
- [45] W.E. Acree Jr., IUPAC-NIST Solubility Data Series. 99. Solubility of benzoic acid and substituted benzoic acids in both neat organic solvents and organic solvent mixtures, *J. Phys. Chem. Ref. Data* 42 (2013) 033103. <https://doi.org/10.1063/1.4816161>.
- [46] S. Wang, Z. Song, J. Wang, Y. Dong, M. Wu, Solubilities of ibuprofen in different pure solvents, *J. Chem. Eng. Data* 55 (2010) 5283-5285. <https://doi.org/10.1021/je100255z>.
- [47] I. Malijevská, J. Pick, Liquid-vapour equilibrium in strongly associating binary systems aliphatic acid-benzene, *Collect. Czech. Chem. Commun.* 43 (1978) 2096-2104. <https://doi.org/10.1135/cccc19782096>.

- [48] A. Rius, J.L. Otero, A. Macarron, Equilibres liquide—vapeur de mélanges binaires donnant une réaction chimique: systèmes méthanol—acide acétique; éthanol—acide acétique; n-propanol—acide acétique; n-butanol—acide acétique, *Chem. Eng. Sci.* 10 (1959) 105-111. [https://doi.org/10.1016/0009-2509\(59\)80029-4](https://doi.org/10.1016/0009-2509(59)80029-4).
- [49] I. Maliževská, M. Sýsová, D. Vlčková, Vapour-liquid equilibrium in strongly associated systems. The systems acetic acid-propionic acid and acetic acid-trifluoroacetic acid, *Collect. Czech. Chem. Commun.* 51 (1986) 194-205. <https://doi.org/10.1135/cccc19860194>.
- [50] R. York Jr., R.C. Holmes, Vapor-liquid equilibria of the system acetone—acetic acid—water, *Ind. Eng. Chem.* 34 (1942) 345-350. <https://doi.org/10.1021/ie50387a020>.
- [51] R. Francesconi, F. Comelli, B. Canepa, Liquid-vapour equilibrium with association of components. The acetic acid-acetone and water-pyridine mixtures, *Chim. Ind. Milano* 56 (1974) 485-491.
- [52] K.S. Howard, F.P. Pike, Viscosities and densities of acetone-benzene and acetone-acetic acid systems up to their normal boiling points, *J. Chem. Eng. Data* 4 (1959) 331-333. <https://doi.org/10.1021/je60004a013>.
- [53] D.F. Othmer, Composition of vapors from boiling binary solutions, *Ind. Eng. Chem.* 35 (1943) 614-620. <https://doi.org/10.1021/ie50401a018>.

**Declaration of interests**

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: