



**HAL**  
open science

## Calculation of drug-like molecules solubility using predictive activity coefficient models

Fátima L. Mota, António J. Queimada, Alfonsina E. Andreatta, Simão P. Pinho, Eugénia A. Macedo

► **To cite this version:**

Fátima L. Mota, António J. Queimada, Alfonsina E. Andreatta, Simão P. Pinho, Eugénia A. Macedo. Calculation of drug-like molecules solubility using predictive activity coefficient models. *Fluid Phase Equilibria*, 2012, 322-323, pp.48 - 55. 10.1016/j.fluid.2012.02.003 . hal-01637119

**HAL Id: hal-01637119**

**<https://hal.science/hal-01637119v1>**

Submitted on 26 Oct 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Calculation of Drug-Like Molecules Solubility using Predictive Activity Coefficient Models

**Fátima L. Mota<sup>a</sup>, António J. Queimada<sup>a</sup>, Alfonsina E. Andreatta<sup>b</sup>, Simão P.  
Pinho<sup>c</sup>, and Eugénia A. Macedo<sup>a,\*</sup>**

*<sup>a</sup>LSRE - Laboratory of Separation and Reaction Engineering, - Associate  
Laboratory LSRE/LCM, Faculdade de Engenharia, Universidade do Porto, Rua Dr.  
Roberto Frias, 4200 - 465 Porto, Portugal*

*<sup>b</sup>PLAPIQUI, Universidad Nacional del Sur – CONICET, Camino La Carringanda km 7,  
8000 Bahia Blanca, Argentina*

*<sup>c</sup>LSRE - Laboratory of Separation and Reaction Engineering, - Associate  
Laboratory LSRE/LCM, Escola Superior de Tecnologia e Gestão, Instituto Politécnico  
de Bragança, Campus de Santa Apolónia, 5301-857 Bragança, Portugal*

\*To whom correspondence should be addressed: [eamacedo@fe.up.pt](mailto:eamacedo@fe.up.pt)

## **Abstract**

The A-UNIFAC, UNIFAC, and NRTL-SAC models are used to predict solubility in pure solvents of a set of drug-like molecules. To apply A-UNIFAC, a new set of residual interaction parameters between the -ACOH group and six other groups had to be estimated. The solute model parameters of NRTL-SAC were also estimated for this set of molecules. NRTL-SAC showed better performance at 298.15 K, with an average absolute deviation of 37.6%. Solubility dependence with temperature was also studied: all models presented average deviations around 40%. In general, there is an improvement given by the A-UNIFAC over the UNIFAC in aqueous systems, proving the importance of taking association into account.

The reference solvent approach was also applied improving the results. Solubility in pure solvents can now be predicted with an average deviation around 35.2%. This approach reduces differences previously found between the three models, being a powerful methodology.

**Keywords:** A-UNIFAC, Complex chemicals, Modelling, NRTL-SAC, Solubility, UNIFAC.

## 1. Introduction

Solubility has been recognized as one of the most important properties for designing separation and purification processes of complex molecules, such as biomolecules and active pharmaceutical ingredients [1,2]. In many cases, data are unavailable due to reduced amounts of sample, time limitations, or inherent complexities with experimental measurements. In such cases, the increasing interest and importance of solubility modelling is clear, as can be seen in the very recent works of Ellegaard et al. [3] and Diedrichs and Gmehling [4].

The Non-Random Two-Liquid (NRTL) Segment Activity Coefficient (SAC) is a recent and successful model for solid-liquid equilibrium calculations, and the idea behind is to limit the number of intermolecular interaction parameters [5,6]. This model was successfully employed for complex chemicals where acceptable deviations were obtained between experimental and predicted values: average root mean square error in  $\ln x$  of 0.37, what corresponds to  $\pm 45\%$  of accuracy in solubility predictions [5-8].

The group-contribution methods, like UNIFAC [9,10] and A-UNIFAC [11,12] are adequate techniques to provide reasonably accurate estimates of fluid mixtures non-idealities. Gracin et al. [13] used UNIFAC to predict drug solubilities in pure solvents, but the results were not accurate enough to allow the design of crystallization processes, or the selection of solvents. It is well known that UNIFAC method is not suitable when strong association effects are present. In order to take them into account, an extended version, the A-UNIFAC, was presented by Mengarelli et al. [14], being successfully applied to mixtures containing sugars, alcohols, water, carboxylic acids, esters, aromatic hydrocarbons and alkanes [11,12,15].

Avoiding the knowledge of the solute melting properties, the reference-solvent approach (RSA) [1,16,17] is an interesting alternative for predicting solubilities of

solids. It minimizes the impact of the melting data uncertainties, being so interesting when there is a decomposition reaction, or solid-solid or glass transitions, and allows to fit a small number of unknown parameters from a limited set of well-chosen experimental data. However, it involves the selection of a reference solvent, relative to which all solubility calculations are made. This methodology was used to study the solubilities of drugs in pure solvents [1], and of complex medium-sized chemicals in mixed solvents [16], and the results are promising even in the cases of relatively high solubility compounds where it is not expected.

The aim of this work is to evaluate different activity coefficient models, UNIFAC, A-UNIFAC and NRTL-SAC, for solubility calculations of drug-like molecules, predicting how solubility vary with solvent type and temperature. With the application of RSA, the impact of calculating solubility without explicitly using pure component properties was checked. For that purpose, an evaluation database constituted by compounds with different functional groups and molecular sizes, showing a multiplicity of interactions in different solvents was compiled. They are benzoic acid, salicylic acid, acetylsalicylic acid, ibuprofen, hydroquinone, estriol and estradiol (Figure 1). An extension of the group interaction parameters available for A-UNIFAC was needed, and the solute parameters of the NRTL-SAC model were also estimated. The number of solutes and solvents involved extends considerably previous studies, allowing to understand their usefulness to predict solubility of pharmaceuticals.

## **2. Modelling**

The solubility of a solid solute in a liquid solvent can be calculated by solving the thermodynamic equations of equilibrium [18]. In Appendix A, the most important

equations of the three models used in this work are briefly presented. The required pure solute properties are given in Table 1.

## 2.1 New A-UNIFAC group interaction parameters

For the group-contribution methods, the volume and area parameters were based on the Gmehling et al. [24] revised UNIFAC tables. The residual group interaction ( $a_{m,n}$ ) and association parameters used in A-UNIFAC method were published by Ferreira et al. [12], while those used in UNIFAC method were based on Poling et al. [25] tables. The chemicals have the following associative groups: salicylic acid has carboxylic, hydroxyl and aromatic ring; benzoic acid and ibuprofen have carboxylic and aromatic ring; acetylsalicylic acid has carboxylic, ester and aromatic ring; hydroquinone, estriol and estradiol have hydroxyl and aromatic ring. In the majority of the studied systems, both solute and solvent can associate, like in alcohols and aqueous systems. Benzene derivatives and esters can only associate if the solvent has an electropositive site.

However, to apply the A-UNIFAC method, it was necessary to extend the residual group interaction parameters available [12]: the unknown parameters involving the  $-ACOH$  group were estimated to represent molecules such as salicylic acid, hydroquinone, estriol and estradiol. The association in this group is assumed to be the same as the association in the  $-OH$  group, given by Ferreira et al. [12], but the residual group interaction parameters were estimated using experimental data on low-pressure vapor-liquid equilibria (VLE) of binary mixtures and liquid-liquid equilibria (LLE) of ternary mixtures. Table 2 shows the database used in the fitting procedure. The residual parameters for  $ACOH/ACH$  and  $ACOH/CH_2$  were first calculated using VLE data of phenol-benzene and phenol-alkanes, respectively, while the residual parameters for

ACOH/H<sub>2</sub>O, ACOH/OH, ACOH/COOR, ACOH/COOH interactions were determined simultaneously using all other data expressed in Table 2. Table 3 reports the new estimated group interaction parameters for –ACOH between paraffinic, aromatic, alcohol, water, ester and acid groups.

Table 4 lists the average deviations obtained in the correlation of binary VLE data using the A-UNIFAC method. The results are satisfactory, even if for water-phenol system higher deviations were found. In fact, in order to obtain a good description of both VLE and LLE equilibria a compromise had to be considered when estimating the ACOH/H<sub>2</sub>O parameter. As can be seen in Figure 2, an accurate description of the LLE for the water-phenol-benzene system was achieved using the same set of parameters for both types of equilibria.

## **2.2 New NRTL-SAC solute parameters**

To apply the NRTL-SAC model, the solute parameters had to be estimated using solubility data in pure solvents at 298.15 K [6]. In Table 5, the conceptual segment numbers found for each solute are presented, as well as the number of solvents used for each solute. Using these parameters, the model can be used to obtain the solubility of the same solutes in different solvents, using the solvent parameters reported in the literature [6]. Even if some outliers are found, an average deviation of 41% is obtained in the parameters estimation, with a maximum deviation of 67% for estriol.

## **2.3 Reference Solvent Approach**

To apply the RSA, all the required parameters are known, and no data fitting is required. According to this methodology, the solubility in a solvent *i* can be calculated as far as the solubility in a solvent *j* and an activity coefficient model are known:

$$\ln x_{2i} = \ln x_{2j} + \ln \gamma_{2j} - \ln \gamma_{2i} \quad (1)$$

The selection of the reference solvent ( $j$ ) is based on the minimization:

$$\min_j \left| \ln x_{2j} + \ln \gamma_{2j} - \sum_i \frac{\ln x_{2i} + \ln \gamma_{2i}}{NP} \right| \quad (2)$$

where  $NP$  is the total number of data points for the different  $i$  solvents in the database.

### 3. Results and Discussion

In order to compare the performance of each model, the absolute average deviations (AAD) were calculated. In the following sections, the results are analysed in terms of solubility at 298.15 K, its temperature dependence and, finally, the important features of the RSA methodology.

#### 3.1. Solubility at 298.15 K

Table 6 shows the average AAD's for the solubility calculations in pure solvents at 298.15 K. For comparison, the experimental mole fraction ranges are also presented. The averages presented at the bottom of the table are weighted means, taking into account the number of solvents studied for each solute. It is important to mention that whenever the AAD is higher than 150% in a given solvent, this is excluded from the average presented in all the models. More than 70% of the outliers were found in the calculation of the solubility in water, alkanes or aromatic solvents.

For salicylic acid [47], UNIFAC and NRTL-SAC models perform almost similarly, being acetic acid the only solvent for which the UNIFAC method is significantly better. Comparing group-contribution methods, A-UNIFAC is better than UNIFAC in the case of water solubilities, confirming that UNIFAC work poorly with



aqueous systems. However, the results found with NRTL-SAC in water are still better than those found with the cubic-plus-association equation of state by Mota et al. [48]

The three models are reliable for the calculation of benzoic acid solubilities in pure solvents. The AAD is particularly high in n-pentane with UNIFAC and NRTL-SAC, but A-UNIFAC performs well.

Excluding the esters, for acetylsalicylic acid containing systems, the NRTL-SAC model presents better results than the group-contribution methods, but for many solvents a fair comparison is hampered by the inexistence of NRTL-SAC parameters. A-UNIFAC adds an improvement over UNIFAC for all the studied solvents. The deviations found by Charlton et al. [49], applying the Abraham general solvation model, are in the same order of accuracy of NRTL-SAC.

For ibuprofen [50,51], the results obtained in this work are better than those obtained by Stovall et al. [51] with the Abraham general solvation model, excepting 1-butanol and 1-propanol.

The data reported by Ruelle et al. [23] for estriol were better predicted using the UNIFAC method, but the set of solvents studied is too reduced for a definite conclusion. Concerning the NRTL-SAC model, the number of solvents used for parameter estimation was in accordance with the value suggested by Chen and Crafts [6], presenting a variety of interaction characteristics that would be appropriated to estimate the solute parameter with some physical meaning. Even if the magnitude of the parameters depends on the number and type of solvents, the hydrophobic segment is null, while in the estradiol molecule it is an important segment, although these two molecules are very similar. This is maybe an aspect interfering in the quality of predictions. Bouillot et al. [8] developed a method to choose the best solvents for solute parameters regression.

NRTL-SAC and UNIFAC revealed better performance in the calculation of the data of estradiol solubility in pure solvents at 298.15 K [23]. In Figure 3, the importance of having an activity coefficient model to calculate estradiol solubilities at 298.15 K is shown. As expected, the ideal solubility is in high disagreement with experimental solubilities since the solute-solvent interactions are not taken into account.

Figure 4 presents a comparison between the group-contribution methods for molecules presenting the –ACOH group and the other solutes: the results worsen as long as the number of associative groups increases, but both methods are able to represent solubility in the majority of the solvents studied. The estimation of the –ACOH residual parameter had strong influence, otherwise the results found with A-UNIFAC for hydroquinone, estriol, estradiol and salicylic acid would be much poorer.

### *3.2. Solubility Data as Function of Temperature*

Whenever available, the solubility data as function of temperature were also analyzed. Table 7 shows the AAD's for these calculations taking only into account the solvents for which all the three models were used. The global AAD's are in accordance to those found for pure solvents at 298.15 K.

For salicylic acid [48,52-55], UNIFAC is better for all studied systems, except in water where only NRTL-SAC calculates reasonably these solubilities. A-UNIFAC is still better than UNIFAC for aqueous solubilities.

For benzoic acid [56-58], except in the case of aqueous solubilities where UNIFAC is better, the three models perform almost similarly (A-UNIFAC is slightly better, with average AAD of 15.6%). The solubilities in octanol are presented in Figure 5 where the AAD found with A-UNIFAC is 13.5%. The results obtained by Qingzhu et al. [57] for benzoic acid in 1-octanol with the Wilson and  $\lambda H$  equations correlate the

experimental data as well as the NRTL-SAC predictions in this work. The results found by Wang et al. [58], with the NRTL, for the solubility in acetic acid are not too much different from the calculated by group-contribution methods, what puts into evidence the results calculated here because correlation and prediction are compared.

For the solubility data of acetylsalicylic acid, NRTL-SAC was the best model for the simplest alcohols, ethanol and 2-propanol [21]. The results found for these alcohols by Maia and Giulietti [21] with the Nývlt model, which is a correlative model with three adjustable parameters, are as good as the obtained with NRTL-SAC model. For the aqueous solubilities, NRTL-SAC is the best, but A-UNIFAC leads to a great improvement over UNIFAC: AAD from 87.0 to 14.8%.

NRTL-SAC gave good estimates of ibuprofen aqueous solubilities where the others fail. Except ethyl acetate and propylene glycol, for all the other solvents A-UNIFAC originates better predictions. The results with A-UNIFAC were better than those reported by Hanhnenkamp et al. [59] with UNIFAC, modified UNIFAC or COSMO-RS models.

For hydroquinone containing systems [60], the UNIFAC and NRTL-SAC results are better, except in water where A-UNIFAC is the best. For hydroquinone, the estimation of  $-ACOH$  group interaction parameters for A-UNIFAC method had a great importance: if the group assignment was considered to be  $-AC + -OH$  instead, the method shows special difficulties to deal with alcohols and esters, presenting deviations around 90%; with the proper molecule division, with an  $-ACOH$ , the method is perfectly able to represent those solubilities.

Analyzing the estradiol data as a function of temperature [61], A-UNIFAC gave better results in cyclohexane and benzene, while NRTL-SAC is better in ethanol.

Even if for some systems the results achieved were not satisfactory, the improvement obtained by considering an association term in the A-UNIFAC comparing with the original UNIFAC is shown in Figure 6 for all the aqueous solubilities studied in this work. In fact, preliminary studies developed for solubilities in mixed solvents containing water have shown the same trend, which is important for this type of solutes.

### *3.3. Reference Solvent Approach*

An evident problem arising from the use of solubility equation is the need of solute liquid and solid heat capacities, and melting properties, which usually are not available present large discrepancies in the open literature. In this way, the solubility data at 298.15 K were analyzed using the same activity coefficient models together with the RSA. Whenever the RSA was applied, no further model parameters are estimated. Table 6 reports the AAD's for the solubility calculations in pure solvents at 298.15 K, where it is also presented the weighted average for each model and the reference solvent. It must be mentioned that in a small number of cases (benzoic acid and ibuprofen with UNIFAC, estradiol with A-UNIFAC and NRTL-SAC, and acetylsalicylic acid with NRTL-SAC) the application of the models without the RSA gives better results, but the differences are minor.

In general the results improve significantly after applying RSA, and this approach also eliminates the AAD differences previously observed between models. This is a clear advantage of this methodology because the melting properties, which are a source of inconsistencies, are eliminated.

It should be referred that with NRTL-SAC model, the reference solvent is always an alcohol, while with UNIFAC an alcohol or an ester is found. In fact, excepting the acetylsalicylic acid with UNIFAC, and ibuprofen with the NRTL-SAC,

the solvents found to be the reference already presented good performances without RSA. In the selection of the reference solvent type for the A-UNIFAC method, establish a definite pattern is more difficult.

#### **4. Conclusions**

The compounds under study are complex medium-sized chemicals with different structures and functional groups, with different affinities to associate. Group-contribution methods were compared with the NRTL-SAC model to predict their solubility data in pure and mixed solvent systems. In order to extend the A-UNIFAC method to the associating molecules under study, VLE and LLE data were used to extend the residual parameter table: new residual interaction parameters between the – ACOH group and several other were estimated. For NRTL-SAC model, four parameters had also to be firstly correlated for each solute.

In general, NRTL-SAC is able to correlate and predict solubility data for the studied systems. For aqueous systems A-UNIFAC was in general superior to UNIFAC, proving the need of taking into account the association effects. The results obtained with the reference solvent approach were almost similar for all the models; the improvement relatively to the results without this methodology is in general significant, proving its usefulness.

Therefore, if drug solubility estimates are needed, for systems in which data in a set of selected solvents are readily available, it is suggested to apply the NRTL-SAC model. For solutes where experimental solubility data are not available or difficult to obtain, and if all the group parameters are known, the application of UNIFAC method with the reference solvent approach is a valid option.

## Acknowledgements

The authors are grateful for the support provided by *Fundação para a Ciência e a Tecnologia* (FCT, Lisboa, Portugal) and by LSRE financing by FEDER/POCI/2010.

Fátima L. Mota acknowledges her FCT Ph.D. scholarship SFRH/BD/32372/2006.

## Nomenclature

$a_{m,n}$  = residual group interaction parameter (g/mol)

$M$  = molar mass (g/mol)

$NC$  = number of components in the mixture

$NP$  = number of data points

$NS$  = number of solvents

$P$  = pressure (Pa)

$T$  = absolute temperature (K)

$x$  = mole fraction

$X$  = hydrophobic segment in NRTL-SAC model

$Y+$  = polar repulsive segment in NRTL-SAC model

$Y-$  = polar attractive segment in NRTL-SAC model

$Z$  = hydrophilic segment in NRTL-SAC model

## Greek Symbols

$\gamma$  = activity coefficient

$\Delta C_p$  = difference between the heat capacity of the pure liquid and solid (J/mol.K)

$\Delta_{fus}H$  = fusion enthalpy (J/mol)

## Subscripts

$m$  = melting

## List of Abbreviations

AAD = absolute average deviation  $AAD (\%) = \frac{1}{NP} \sum_i \frac{|x_2^{\text{exp}} - x_2^{\text{calc}}|}{x_2^{\text{exp}}} \times 100$

A-UNIFAC = modified UNIFAC method which takes account association

LLE = Liquid liquid equilibria

NRTL-SAC = Nonrandom two-liquid segment activity coefficient model

RSA = reference solvent approach

UNIFAC = UNiversal quasi chemical Functional-group Activity Coefficients method

VLE = Vapor liquid equilibria

## Appendix A

In the UNIFAC method, the activity coefficient is obtained from a combinatorial ( $\gamma_i^C$ ) and a residual ( $\gamma_i^R$ ) contributions [9,62]:

$$\ln \gamma_i = \ln \gamma_i^C + \ln \gamma_i^R \quad (\text{A1})$$

The combinatorial term, is:

$$\ln \gamma_i^C = \ln \frac{\phi_i}{x_i} + \frac{z}{2} q_i \ln \frac{\theta_i}{\phi_i} + l_i - \frac{\phi_i}{x_i} \sum_J x_J l_J \quad (\text{A2})$$

where  $\theta_i$  and  $\phi_i$  are the area and volume fractions of component  $I$ , and similarly to the variable  $l_i$  are defined by equations found elsewhere [62],  $z$  is the coordination number (set equal to 10) and  $q_i$  is the area parameter of component  $I$ . The residual term, is:

$$\ln \gamma_i^R = \sum_k v_k^{(i)} (\ln \Gamma_k - \ln \Gamma_k^{(i)}) \quad (\text{A3})$$

$$\ln \Gamma_k = q_k \left[ 1 - \ln \left( \sum_m \theta_m \psi_{mk} \right) - \sum_m \frac{\theta_m \psi_{km}}{\sum_n \theta_n \psi_{nm}} \right] \quad (\text{A4})$$

where  $\Gamma_k$  is the residual activity coefficient of group  $k$ ,  $\Gamma_k^{(I)}$  is the residual activity coefficient of group  $k$  in a reference solution containing only molecules of type  $I$ ,  $v_k^{(I)}$  is the number of group  $k$  present in molecule  $I$  and  $\psi_{mk}$  is a binary interaction parameter [9,62].

In order to take the association effects into account, an extended version of UNIFAC, the A-UNIFAC, was presented by Mengarelli et al. [14] The association activity coefficient term is a function of the fraction of non-bonded sites in the solution ( $X^{A_k}$ ) and in pure-component  $I$  ( $X_I^{A_k}$ ):

$$\ln \gamma_I^{assoc} = \sum_{k=1}^{NGA} \left\{ \vartheta_k^I \sum_{A_k} \left[ \ln \left( \frac{X^{A_k}}{X_I^{A_k}} \right) + \frac{X_I^{A_k} - X^{A_k}}{2} \right] + \sum_{A_k} \left( \frac{1}{X^{A_k}} - \frac{1}{2} \right) N_k \left( \frac{\partial X^{A_k}}{\partial n_I} \right)_{T,P,n_j} \right\} \quad (\text{A5})$$

where  $N_k$  is the number of moles of associating group  $k$ ,  $\vartheta_k^I$  represents the number of groups of type  $k$  in a molecule of component  $I$ ,  $n_I$  is the number of moles of component  $I$ ,  $NGA$  is the number of associating groups and  $A_k$  is the associating site of group  $k$ . The fraction of non-bonded sites is a function of the associating strength ( $\Delta^{A_k B_j}$ ) between site  $A$  of group  $k$  and site  $B$  of group  $j$  and the associating group density in the mixture  $\rho_j$  [12,14]:

$$\rho_j = \frac{\sum_{I=1}^{NC} \vartheta_j^I x_I}{\sum_{I=1}^{NC} r_I x_I} \quad (\text{A6})$$

where  $NC$  is the number of components in the mixture and  $r_I$  represents the UNIQUAC molecular volume of species  $I$  (evaluated from UNIFAC group volume parameters  $R_k$ ). The associative compounds are represented by specific associating groups: the one-site carboxylic group, which can self- and cross-associate; the two-sites hydroxyl group, which can also self- and cross-associate; the one-site ester group and one-site aromatic ring, which can only cross-associate with electropositive associating sites.



The Non-Random Two-Liquid Segment Activity Coefficient model [5,6] (NRTL-SAC) is an activity coefficient model where the liquid non-idealities are described in terms of four types of segments: hydrophobic ( $X$ ), polar attractive ( $Y^-$ ) and repulsive ( $Y^+$ ), and hydrophilic ( $Z$ ), to account for several interactions. The activity coefficient is obtained from the sum of a combinatorial ( $\gamma_I^C$ ) and a residual ( $\gamma_I^R$ ) contributions. The basic equations are:

$$\ln \gamma_I = \ln \gamma_I^C + \ln \gamma_I^R \quad (\text{A7})$$

$$\ln \gamma_I^C = \ln \frac{\phi_I}{x_I} + 1 - r_I \sum_J \frac{\phi_J}{r_J} \quad (\text{A8})$$

$$\ln \gamma_I^R = \ln \gamma_I^{lc} = \sum_m r_{m,I} (\ln \Gamma_m^{lc} - \ln \Gamma_m^{lc,I}) \quad (\text{A9})$$

where  $I$  and  $J$  are component indexes,  $m$  is the segment based species indexes,  $r_{m,I}$  is the number of segment species  $m$  contained in component  $I$ ,  $r_I$  is the total segment number of component  $I$ ,  $\phi_I$  is the segment mole fraction of component  $I$ ,  $\Gamma_m^{lc}$  is the activity coefficient of segment species  $m$ ,  $\Gamma_m^{lc,I}$  is the activity coefficient of segment species  $m$  contained only in component  $I$ .

## References

- [1] J. Abildskov and J.P. O'Connell, *Ind. Eng. Chem. Res.* 42 (2003) 5622-5634.
- [2] P. Kolář, J.W. Shen, A. Tsuboi and T. Ishikawa, *Fluid Phase Equilib.* 194 (2002) 771-782.
- [3] M.D. Ellegaard, J. Abildskov and J.P. O'Connell, *Ind. Eng. Chem. Res.* 49 (2010) 11620-11632.
- [4] A. Diedrichs and J. Gmehling, *Ind. Eng. Chem. Res.* 50 (2011) 1757-1769.
- [5] C.C. Chen and Y.H. Song, *Ind. Eng. Chem. Res.* 43 (2004) 8354-8362.
- [6] C.C. Chen and P.A. Crafts, *Ind. Eng. Chem. Res.* 45 (2006) 4816-4824.
- [7] F.L. Mota, A.P. Carneiro, A.J. Queimada, S.P. Pinho and E.A. Macedo, *Eur. J. Pharm. Sci.* 37 (2009) 499.
- [8] B. Bouillot, S. Teychene and B. Biscans, *Fluid Phase Equilib.* 309 (2011) 36-52.
- [9] A. Fredenslund, R.L. Jones and J.M. Prausnitz, *AIChE J.* 21 (1975) 1086-1099.
- [10] H.K. Hansen, P. Rasmussen, A. Fredenslund, M. Schiller and J. Gmehling, *Ind. Eng. Chem. Res.* 30 (1991) 2352-2355.

- [11] O. Ferreira, E.A. Brignole and E.A. Macedo, *Ind. Eng. Chem. Res.* 42 (2003) 6212-6222.
- [12] O. Ferreira, E.A. Macedo and S.B. Bottini, *Fluid Phase Equilib.* 227 (2005) 165-176.
- [13] S. Gracin, T. Brinck and A.C. Rasmuson, *Ind. Eng. Chem. Res.* 41 (2002) 5114-5124.
- [14] A.C. Mengarelli, E.A. Brignole and S.B. Bottini, *Fluid Phase Equilib.* 163 (1999) 195-207.
- [15] F. Montañés, A. Olano, E. Ibáñez and T. Fornari, *AIChE J.* 53 (2007) 2411-2418.
- [16] J. Abildskov and J.P. O'Connell, *Mol. Sim.* 30 (2004) 367-378.
- [17] J. Abildskov and J.P. O'Connell, *Fluid Phase Equilib.* 228 (2005) 395-400.
- [18] J.M. Prausnitz, R.N. Lichtenthaler and E.G. Azevedo, *Molecular Thermodynamics of Fluid-Phase Equilibria*, Third ed., Prentice Hall PTR, New Jersey, 1999.
- [19] BYU DIPPR 801 Thermophysical Properties Database, Brigham Young University: Provo, UT, 1998.
- [20] D.Q. Li, J.C. Liu, D.Z. Liu and F.A. Wang, *Fluid Phase Equilib.* 200 (2002) 69-74.
- [21] G.D. Maia and M. Giuliatti, *J. Chem. Eng. Data* 53 (2008) 256-258.
- [22] S. Gracin and A.C. Rasmuson, *J. Chem. Eng. Data* 47 (2002) 1379-1383.
- [23] P. Ruelle, A. Farina-Cuendet and U.W. Kesselring, *Perspect. Drug Discov. Des.* 18 (2000) 61-112.
- [24] J. Gmehling, P. Rasmussen and A. Fredenslund, *Ind. Eng. Chem. Proc. Des. Dev.* 21 (1982) 118-127.
- [25] B. Poling, J.M. Prausnitz and J.P. O'Connell, *The Properties of Gases and Liquids*, Fifth ed., New York, 2001.
- [26] J. Gmehling, *J. Chem. Eng. Data* 27 (1982) 371-373.
- [27] A.R. Martin and C.M. George, *J. Chem. Soc.* (1933) 1413-1416.
- [28] G. Kortüm, H.J. Freier and F. Woerner, *Chem. Ing. Tech.* 25 (1953) 125-133.
- [29] G. Kortüm and H.J. Freier, *Chem. Ing. Tech.* 26 (1954) 670-673.
- [30] S.M. Hosseini and G. Schneider, *Z. Phys. Chem.* 36 (1963) 137.
- [31] G. Kortüm and H.J. Freier, *Monatshefte für Chemie / Chemical Monthly* 85 (1954) 693-702.

- [32] M. Klauck, A. Grenner, K. Taubert, A. Martin, R. Meinhardt and J. Schmelzer, *Ind. Eng. Chem. Res.* 47 (2008) 5119-5126.
- [33] F.A.H. Schreinemakers, *Zeitschrift Fur Physikalische Chemie--Stoichiometrie Und Verwandtschaftslehre* 35 (1900) 459-479.
- [34] V. Kliment, V. Fried and J. Pick, *Collect. Czech. Chem. Commun.* 29 (1964) 2008-&.
- [35] A.P. Karavaeva, G.N. Koshel and L.M. Galiaskarova, *Osnov. Organ. Sintez i Neftekhimiya, Yaroslavl* 3 (1975) 98.
- [36] H. Brusset and J. Gaynes, *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences* 236 (1953) 1563-1565.
- [37] N.P. Markuzin, *Zh. Prikl. Khim. (Leningrad)* 34 (1961) 1175.
- [38] G.Y. Kolyuchkina, V.S. Timofeev and L.A. Serafimov, *Uch. Zap. Mosk. Inst. Tonkoi Khim. Tekhnol.* 1 (1972) 78.
- [39] J.R.A. Gonzalez, E.A. Macedo, M.E. Soares and A.G. Medina, *Fluid Phase Equilib.* 26 (1986) 289-302.
- [40] D.R. Cova, *J. Chem. Eng. Data* 5 (1960) 282-284.
- [41] R.A. Murogova, G.L. Tudorovskaya, I.D. Gridin, A.P. Yurchuk, N.I. Pleskach, V.D. Kozlova and T.V. Bazyleva, *Zh. Prikl. Khim. (Leningrad)* 45 (1972) 824.
- [42] L.H. de Oliveira and M. Aznar, *J. Chem. Thermo.* 42 (2010) 1379-1385.
- [43] S. Takahashi, K. Otake, T. Takahashi and A. Iguchi, *Kagaku Kogaku Ronbunshu* 14 (1988) 531-535.
- [44] A. Aarna and T. Kaps, *Eesti NSV Tead. Akad. Toim., Keem., Geol.* 23 (1974) 16.
- [45] K.S. Narasimhan, C.C. Reddy and K.S. Chari, *J. Chem. Eng. Data* 7 (1962) 340-343.
- [46] *AIChE Symp. Ser.*, 86 (1990) 38.
- [47] K.M. de Fina, T.L. Sharp, L.E. Roy and W.E. Acree, *J. Chem. Eng. Data* 44 (1999) 1262-1264.
- [48] F.L. Mota, A.J. Queimada, S.P. Pinho and E.A. Macedo, *Ind. Eng. Chem. Res.* 47 (2008) 5182-5189.
- [49] A.K. Charlton, C.R. Daniels, W.E. Acree and M.H. Abraham, *J. Solution Chem.* 32 (2003) 1087-1102.
- [50] P. Bustamante, M.A. Pena and J. Barra, *Int. J. Pharm.* 194 (2000) 117-124.

- [51] D.M. Stovall, C. Givens, S. Keown, K.R. Hoover, E. Rodriguez, W.E. Acree and M.H. Abraham, *Phys. Chem. Liq.* 43 (2005) 261-268.
- [52] A. Apelblat and E. Manzurola, *J. Chem. Thermodyn.* 21 (1989) 1005-1008.
- [53] F.L. Nordström and A.C. Rasmuson, *J. Chem. Eng. Data* 51 (2006) 1668-1671.
- [54] M.A. Peña, P. Bustamante, B. Escaler, A. Reíllo and J.M. Bosque-Sendra, *J. Pharm. Biomed. Anal.* 36 (2004) 571-578.
- [55] A. Shalmashi and A. Eliassi, *J. Chem. Eng. Data* 53 (2008) 199-200.
- [56] J. Chipman, *J. Am. Chem. Soc.* 46 (1924) 2443-2448.
- [57] J. Qingzhu, M. Peisheng, Y. Shouzhi, W. Qiang, W. Chang and L. Guiju, *J. Chem. Eng. Data* 53 (2008) 1278-1282.
- [58] Q.B. Wang, L.X. Hou, Y.W. Cheng and X. Li, *J. Chem. Eng. Data* 52 (2007) 936-940.
- [59] I. Hahnenkamp, G. Graubner and J. Gmehling, *Int. J. Pharm.* 388 (2010) 73-81.
- [60] X.N. Li, Q.X. Yin, W. Chen and J.K. Wang, *J. Chem. Eng. Data* 51 (2006) 127-129.
- [61] E.G. Salole, *J. Pharm. Biomed. Anal.* 5 (1987) 635-648.
- [62] A. Fredenslund, J. Gmehling, M.L. Michelsen, P. Rasmussen and J.M. Prausnitz, *Ind. Eng. Chem. Proc. Des. Dev.* 16 (1977) 450-462.
- [63] G.L. Perlovich and A. Bauer-Brandl, *Pharm. Res.* 20 (2003) 471-478.

**Table 1. Pure solute properties.**

| Solute               | $T_m$ (K)          | $\Delta_{\text{fus}}H$ (kJ/mol) | $\Delta C_p$ (J/mol.K) | $M$ (g/mol) |
|----------------------|--------------------|---------------------------------|------------------------|-------------|
| Salicylic acid       | 432.0 <sup>a</sup> | 24.6 <sup>a</sup>               | 37.0 <sup>b</sup>      | 138.1       |
| Benzoic acid         | 395.5 <sup>c</sup> | 17.5 <sup>c</sup>               | 40.9 <sup>b</sup>      | 122.1       |
| Acetylsalicylic acid | 416.2 <sup>d</sup> | 29.8 <sup>d</sup>               | 26.6 <sup>b</sup>      | 180.2       |
| Ibuprofen            | 347.2 <sup>e</sup> | 25.5 <sup>e</sup>               | 50.3 <sup>e</sup>      | 206.3       |
| Hydroquinone         | 445.5 <sup>b</sup> | 27.1 <sup>b</sup>               | 26.7 <sup>b</sup>      | 110.1       |
| Estriol              | 555.0 <sup>f</sup> | 42.7 <sup>f</sup>               | ----                   | 288.4       |
| Estradiol            | 445.0 <sup>f</sup> | 40.6 <sup>f</sup>               | ----                   | 272.4       |

<sup>a</sup>[1]; <sup>b</sup>[19]; <sup>c</sup>[20]; <sup>d</sup>[21]; <sup>e</sup>[22]; <sup>f</sup>[23].

**Table 2. Database used for the estimation of A-UNIFAC residual group interaction parameters.**

| <i>i</i> | <i>j</i>         | System                       | C atoms<br>in groups                             | NP  | Type of<br>data | T range<br>(K) | P range<br>(bar) | Ref.       |
|----------|------------------|------------------------------|--|-----|-----------------|----------------|------------------|------------|
| ACOH     | ACH              | Phenol + benzene             | -  | 133 | VLE             | 323-449        | 0.005-2.6        | [26-31]    |
|          | CH <sub>2</sub>  | Phenol + alkane              | C <sub>8</sub> , C <sub>10</sub>                 | 31  | VLE             | 383-393        | 0.09-0.64        | [26]       |
|          | H <sub>2</sub> O | Phenol + water               | -  | 103 | VLE             | 288-455        | 0.02-1           | [32-38]    |
|          |                  | Phenol + water +<br>benzene  | -  | 10  | LLE             | 298            | 1                | [39]       |
| OH       |                  | Phenol +<br>cyclohexanol     | -  | 20  | VLE             | 376-392        | 0.12             | [40,41]    |
|          |                  | Phenol + water +<br>alcohols | C <sub>4</sub> -C <sub>5</sub>                   | 40  | LLE             | 298-313        | 1                | [42,43]    |
| COOR     |                  | Phenol + esters              | C <sub>4</sub> , C <sub>6</sub> -C <sub>7</sub>  | 67  | VLE             | 362-465        | 0.26-1           | [34,44]    |
|          |                  | Phenol + water + ester       | C <sub>2</sub> , C <sub>4</sub> , C <sub>6</sub> | 99  | LLE             | 298-318        | 1                | [39,43,45] |
| COOH     |                  | Phenol + acids               | C <sub>3</sub>                                   | 26  | VLE             | 383-403        | 0.08-0.70        | [46]       |
|          |                  | Phenol + water + acid        | C <sub>9</sub>                                   | 9   | LLE             | 298            | 1                | [39]       |

NP: number of experimental points.

**Table 3. New A-UNIFAC residual group interaction parameters  $a_{m,n}$  (K).**

| $M$  | $n$              | $a_{m,n}$ | $a_{n,m}$ |
|------|------------------|-----------|-----------|
| ACOH | CH <sub>2</sub>  | 343.4     | -55.40    |
|      | ACH              | 74.03     | -42.87    |
|      | OH               | -722.6    | -410.1    |
|      | H <sub>2</sub> O | -355.0    | -179.2    |
|      | COOR             | -756.7    | 12.03     |
|      | COOH             | 402.9     | 1028      |

**Table 4. VLE calculations: average deviations in compositions ( $\delta y$ ) and pressure ( $\delta P/P$ ) using the A-UNIFAC method.**

| Systems               | $\delta y$ (%) <sup>a</sup> | $\delta P/P$ (%) <sup>b</sup> |
|-----------------------|-----------------------------|-------------------------------|
| Phenol + benzene      | 1.43                        | 5.37                          |
| Phenol + alkane       | 2.11                        | 1.93                          |
| Phenol + water        | 20.6                        | 9.12                          |
| Phenol + cyclohexanol | 11.0                        | 5.53                          |
| Phenol + esters       | 13.8                        | 2.53                          |
| Phenol + acids        | 4.13                        | 1.16                          |

$${}^a \delta y (\%) = \sqrt{\frac{\sum_i (y^{\text{exp}} - y^{\text{calc}})^2}{NP}} \times 100 \quad ; \quad {}^b \frac{\delta P}{P} (\%) = \sqrt{\frac{1}{NP} \sum_i \left( \frac{P^{\text{exp}} - P^{\text{calc}}}{P^{\text{exp}}} \right)^2} \times 100$$



**Table 5. NRTL-SAC molecular parameters ( $r_{m,l}$ ) for the studied compounds and number of solvents used (NS).**

| Solute               | NS | X     | Y-    | Y+    | Z     |
|----------------------|----|-------|-------|-------|-------|
| Salicylic acid       | 26 | 0.886 | 1.841 | 0.372 | 0.444 |
| Benzoic acid         | 26 | 0.316 | 0.046 | 1.876 | 0.580 |
| Acetylsalicylic acid | 16 | 0.508 | 0.219 | 0.009 | 1.000 |
| Ibuprofen            | 28 | 0.789 | 0.104 | 1.837 | 0.107 |
| Hydroquinone         | 7  | 0.243 | 1.915 | 1.749 | 0.020 |
| Estriol              | 10 | 0.000 | 0.003 | 0.037 | 1.216 |
| Estradiol            | 24 | 1.087 | 0.001 | 1.317 | 1.463 |

**Table 6. AAD's<sup>a,b</sup> (%) for solubility modelling in pure solvents at 298.15 K, experimental solubility range, and respective reference solvent when RSA is applied.**

|                       | NS | Xexp range                      | AAD (%)  |                     |        |                     |          |                |
|-----------------------|----|---------------------------------|----------|---------------------|--------|---------------------|----------|----------------|
|                       |    |                                 | A-UNIFAC | A-UNIFAC + RSA      | UNIFAC | UNIFAC + RSA        | NRTL-SAC | NRTL-SAC + RSA |
| Salicylic acid        | 12 | [1.5 x 10 <sup>-4</sup> ; 0.21] | 36.5     | 29.8                | 18.9   | 19.0                | 18.4     | 15.7           |
|                       |    |                                 |          | 2-methyl-1-propanol |        | Ethyl acetate       |          | 1-butanol      |
| Benzoic acid          | 14 | [5.9 x 10 <sup>-3</sup> ; 0.20] | 24.0     | 23.0                | 29.7   | 30.5                | 27.9     | 14.1           |
|                       |    |                                 |          | Benzene             |        | Ethyl acetate       |          | 1-butanol      |
| Acetyl salicylic acid | 26 | [1.0 x 10 <sup>-3</sup> ; 0.09] | 74.1     | 30.8                | 79.6   | 31.2                | 31.3     | 36.3           |
|                       |    |                                 |          | 2-methyl-1-butanol  |        | 2-methyl-1-propanol |          | 1-octanol      |
| Ibuprofen             | 19 | [9.9 x 10 <sup>-7</sup> ; 0.34] | 40.0     | 40.4                | 31.5   | 40.3                | 59.6     | 58.9           |
|                       |    |                                 |          | cyclohexane         |        | Isopropyl myristate |          | 1-pentanol     |
| Estradiol             | 6  | [5.9 x 10 <sup>-7</sup> ; 0.02] | 80.7     | 83.9                | 58.7   | 47.1                | 55.6     | 68.4           |
|                       |    |                                 |          | Ethanol             |        | ethanol             |          | ethanol        |
|                       |    |                                 | 51.2     | 35.7                | 47.6   | 32.7                | 37.6     | 37.1           |

<sup>a</sup> Average calculated without considering the outliers (AAD > 150%)

<sup>b</sup> Outliers – salicylic acid: acetic acid, water; benzoic acid: n-pentane; acetylsalicylic acid: benzene, methyl acetate; ibuprofen: benzene, acetic acid; estradiol: cyclohexane.

**Table 7. AAD's<sup>a,b</sup> (%) for solubility modelling in pure solvents as a function of temperature.**

|                      | AAD (%)  |        |          |
|----------------------|----------|--------|----------|
|                      | A-UNIFAC | UNIFAC | NRTL-SAC |
| Salicylic acid       | 63.3     | 8.0    | 10.7     |
| Benzoic acid         | 15.6     | 25.7   | 19.3     |
| Acetylsalicylic acid | 77.5     | 84.1   | 14.3     |
| Ibuprofen            | 40.0     | 45.1   | 72.3     |
| Hydroquinone         | 58.0     | 40.5   | 35.3     |
| Estradiol            | 55.7     | 23.1   | 21.8     |
|                      | 51.7     | 37.8   | 29.0     |

<sup>a</sup> Average calculated without considering the outliers (AAD > 150%)

<sup>b</sup> Outliers – salicylic acid: water; benzoic acid: water; hydroquinone: acetic acid, water; estradiol: cyclohexane, ethanol.

**Figure 1.** Chemical structures of the compounds under study.

**Figure 2.** LLE of water-phenol-benzene at 298.15 K: ●, experimental[39]; —, A-UNIFAC method.

**Figure 3.** Estradiol mole fraction solubility at 298.15 K [23] calculated by different models (●, A-UNIFAC; □, UNIFAC; ▲, NRTL-SAC) and ideal (—).

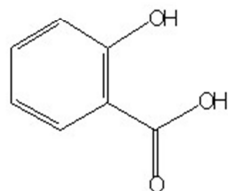
**Figure 4.** Comparison between group-contribution methods at 298.15 K: closed symbols, A-UNIFAC; open symbols, UNIFAC. (a) ●, acetylsalicylic acid [49,63]; ◆, benzoic acid [63]; ■, ibuprofen [50,51]; (b) ●, salicylic acid [47]; ◆, hydroquinone [60]; ▲, estriol [23]; ■, estradiol [23].

**Figure 5.** Benzoic acid mole fraction solubility in 1-octanol (◆) as a function of temperature: experimental [57] and calculated (—, A-UNIFAC; - -, UNIFAC; -·-, NRTL-SAC; ····, ideal).

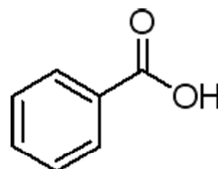
**Figure 6.** Comparison between group-contribution methods for aqueous solubilities at different temperatures: ●, A-UNIFAC; ○, UNIFAC.

**Figure 1:**

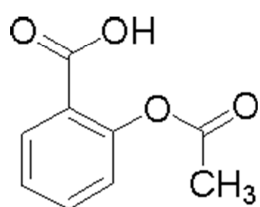
**Salicylic acid**



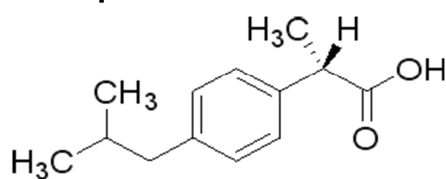
**Benzoic acid**



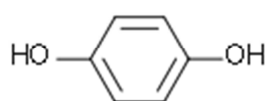
**Acetylsalicylic acid**



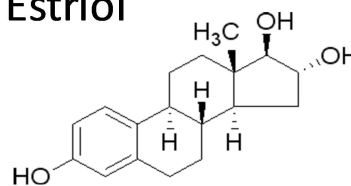
**Ibuprofen**



**Hydroquinone**



**Estriol**



**Estradiol**

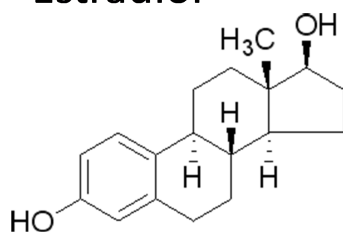
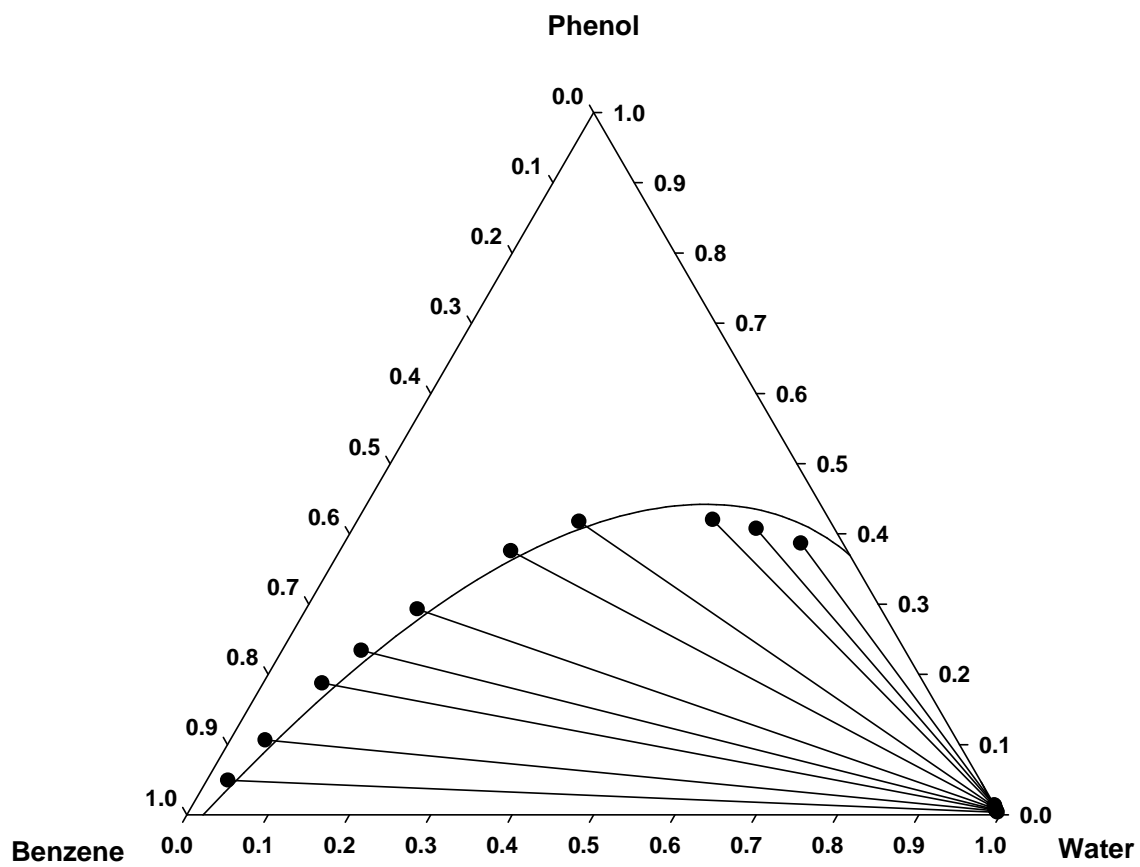


Figure 2:



**Figure 3:**

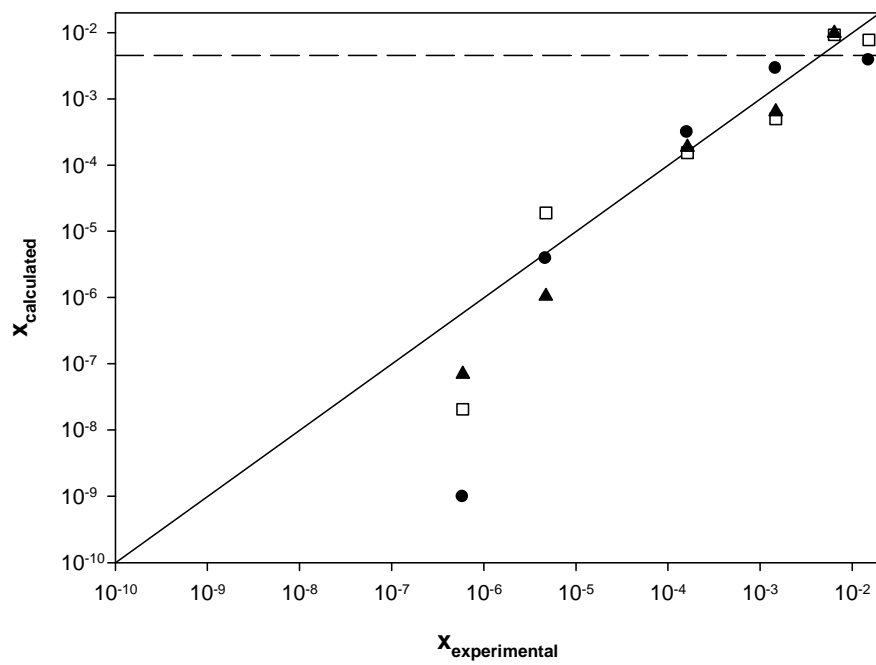


Figure 4:

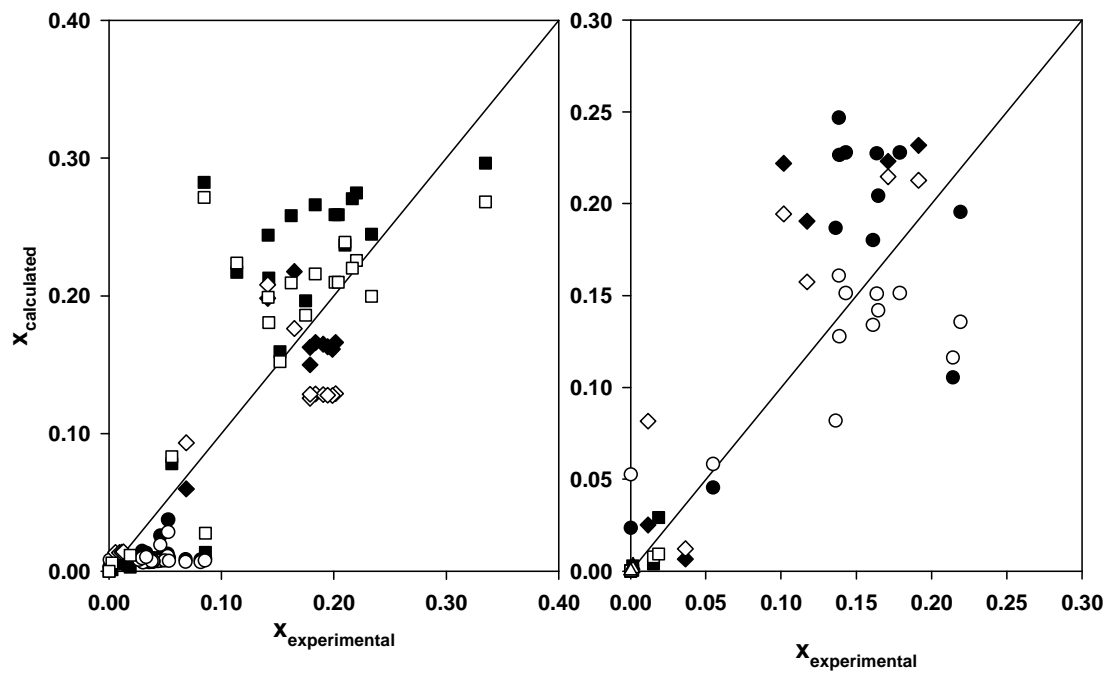
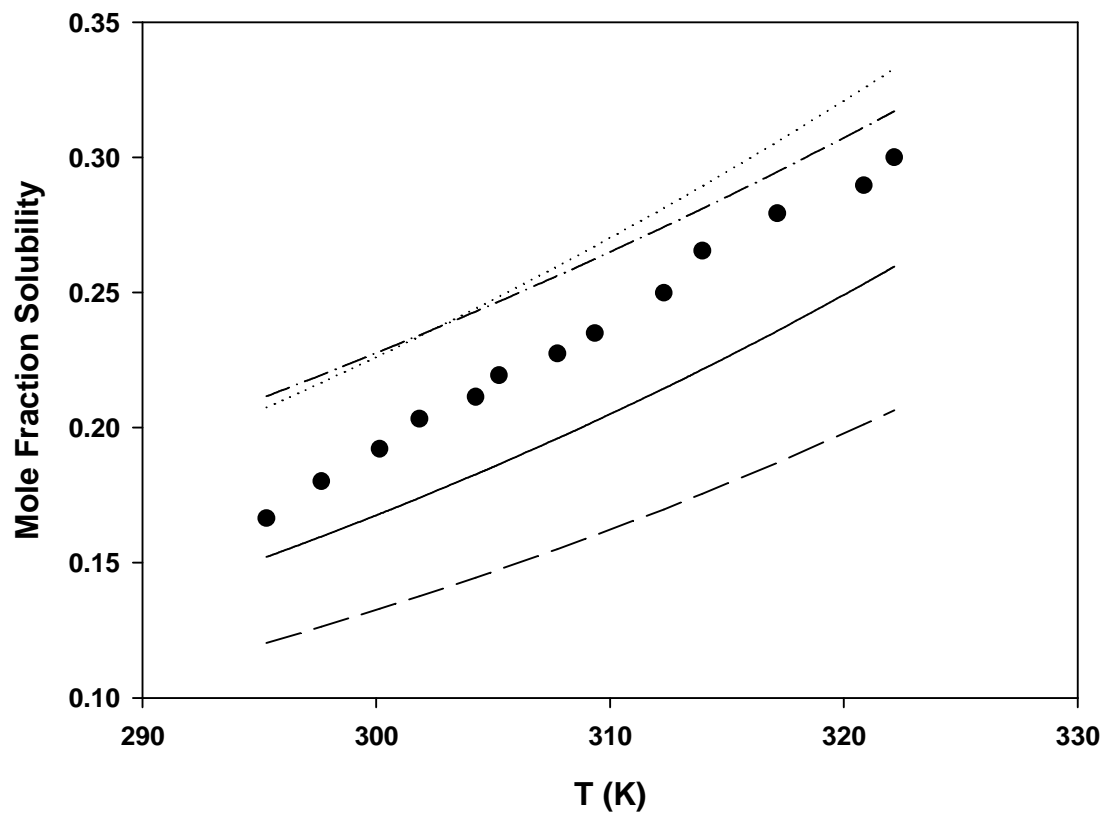




Figure 5:



**Figure 6:**

